An Introduction to Logic

Madhavan Mukund
Chennai Mathematical Institute
E-mail: madhavan@cmi.ac.in

Abstract

These are lecture notes for an introductory course on logic aimed at graduate students in Computer Science. The notes cover techniques and results from propositional logic, modal logic, propositional dynamic logic and first-order logic. The notes are based on a course taught to first year PhD students at SPIC Mathematical Institute, Madras, during August–December, 1997.

Contents

1 Propositional Logic **1**
- 1.1 Syntax . . . 1
- 1.2 Semantics . . . 1
- 1.3 Axiomatisations . . . 3
- 1.4 Maximal Consistent Sets and Completeness . . . 6
- 1.5 Compactness and Strong Completeness . . . 8

2 Modal Logic **13**
- 2.1 Syntax . . . 13
- 2.2 Semantics . . . 13
- 2.3 Correspondence Theory . . . 15
- 2.4 Axiomatising valid formulas . . . 19
- 2.5 Bisimulations and expressiveness . . . 24
- 2.6 Decidability: Filtrations and the finite model property . . . 28
- 2.7 Labelled transition systems and multi-modal logic . . . 33

3 Dynamic Logic **36**
- 3.1 Syntax . . . 36
- 3.2 Semantics . . . 36
- 3.3 Axiomatising valid formulas . . . 38

4 First-Order Logic **47**
- 4.1 Syntax . . . 48
- 4.2 Semantics . . . 49
- 4.3 Formalisations in first-order logic . . . 52
- 4.4 Satisfiability: Henkin's reduction to propositional logic . . . 57
- 4.5 Compactness and the Löwenheim-Skolem Theorem . . . 62
- 4.6 A Complete Axiomatisation . . . 63
- 4.7 Variants of the Löwenheim-Skolem Theorem . . . 67
- 4.8 Elementary Classes . . . 68
- 4.9 Elementarily Equivalent Structures . . . 70
- 4.10 An Algebraic Characterisation of Elementary Equivalence . . . 74
- 4.11 Decidability . . . 80

1 Propositional Logic

1.1 Syntax

We begin with a countably infinite set of atomic propositions $\mathscr{P} = \{p_0, p_1, \ldots\}$ and two logical connectives $\neg$ (read as *not*) and $\vee$ (read as *or*).

The set Φ of formulas of propositional logic is the smallest set satisfying the following conditions:

- Every atomic proposition p is a member of Φ.
- If α is a member of Φ, so is $(\neg\alpha)$.
- If α and β are members of Φ, so is $(\alpha \vee \beta)$.

We shall normally omit parentheses unless we need to explicitly clarify the structure of a formula. We follow the convention that $\neg$ binds more tightly than $\vee$. For instance, $\neg\alpha \vee \beta$ stands for $((\neg\alpha) \vee \beta)$.

***Exercise** 1.1* Show that Φ is a countably infinite set. ⊣

The fact that Φ is the *smallest* set satisfying this inductive definition provides us with the principle of *structural induction*.

Structural induction principle Let S be a set such that:

- Every atomic proposition p is a member of S.
- If α is a member of S, so is $(\neg\alpha)$.
- If α and β are members of S, so is $(\alpha \vee \beta)$.

Then, $\Phi \subseteq S$.

1.2 Semantics

To assign meaning to formulas, we begin by assigning meaning to the atomic propositions. Let $\top$ denote the truth value *true* and $\bot$ the truth value *false*.

- A *valuation* v is a function $v : \mathscr{P} \to \{\top, \bot\}$.

We can also think of a valuation as a subset of $\mathscr{P}$—if $v : \mathscr{P} \to \{\top, \bot\}$, then $v \subseteq \mathscr{P} = \{p \mid v(p) = \top\}$. Thus, the set of all valuations is $2^{\mathscr{P}}$, the set of all subsets of $\mathscr{P}$.

We extend each valuation $v : \mathscr{P} \to \{\top, \bot\}$ to a map $\widehat{v} : \Phi \to \{\top, \bot\}$ as follows:

- For $p \in \mathscr{P}$, $\widehat{v}(p) = v(p)$.

- For α of the form $\neg\beta$, $\widehat{v}(\alpha) = \begin{cases} \top & \text{if } \widehat{v}(\beta) = \bot \\ \bot & \text{otherwise} \end{cases}$
- For α of the form $\beta \vee \gamma$, $\widehat{v}(\alpha) = \begin{cases} \bot & \text{if } \widehat{v}(\beta) = \widehat{v}(\gamma) = \bot \\ \top & \text{otherwise} \end{cases}$

The principle of structural induction can be used to formally argue that $\widehat{v}$ is well-defined (that is, $\widehat{v}$ is indeed a function and is defined for all formulas).

Just as v can be defined as a subset of $\mathscr{P}$, $\widehat{v}$ can be defined as a subset of Φ—namely, $\widehat{v} = \{\alpha \mid \widehat{v}(\alpha) = \top\}$.

***Exercise** 1.2* We saw that every subset of $\mathscr{P}$ defines a valuation v. Does every subset of Φ define an extended valuation $\widehat{V}$? ⊣

Since every valuation v gives rise to a unique extension $\widehat{v}$, we shall always denote $\widehat{v}$ as just v.

Derived connectives It will be convenient to introduce some additional connectives when discussing propositional logic.

$$\begin{aligned} \alpha \wedge \beta &\stackrel{\text{def}}{=} \neg(\neg\alpha \vee \neg\beta) \\ \alpha \supset \beta &\stackrel{\text{def}}{=} \neg\alpha \vee \beta \\ \alpha \equiv \beta &\stackrel{\text{def}}{=} (\alpha \supset \beta) \wedge (\beta \supset \alpha) \end{aligned}$$

The connective $\wedge$ is read as *and*, $\supset$ as *implies* and $\equiv$ as *if and only if*.

***Exercise** 1.3* Express $v(\alpha \wedge \beta)$, $v(\alpha \supset \beta)$ and $v(\alpha \equiv \beta)$ in terms of $v(\alpha)$ and $v(\beta)$. ⊣

***Exercise** 1.4* According to the Pigeonhole Principle, if we try to place $n+1$ pigeons in n pigeonholes, then at least one pigeonhole must have two or more pigeons. For $i \in \{1, 2, \ldots, n+1\}$ and $j \in \{1, 2, \ldots, n\}$, let the atomic proposition p_{ij} denote that the i^{th} pigeon is placed in the the j^{th} pigeonhole. Write down a formula expressing the Pigeonhole Principle. What is the length of your formula as a function of n? ⊣

Satisfiability and validity A formula α is said to be *satisfiable* if there is a valuation v such that $v(\alpha) = \top$. We write $v \vDash \alpha$ to indicate that $v(\alpha) = \top$.

The formula α is said to be *valid* if $v \vDash \alpha$ for every valuation v. We write $\vDash \alpha$ to indicate that α is valid. We also refer to valid formulas of propositional logic as *tautologies*.

***Example** 1.5* Let p be an atomic proposition. The formula p is satisfiable. The formula $p \vee \neg p$ is valid. The formula $p \wedge \neg p$ is not satisfiable.

The following observation connects the notions of satisfiability and validity.

Proposition 1.6 *Let α be a formula. α is valid iff $\neg\alpha$ is not satisfiable.*

In applications of logic to computer science, a central concern is to develop algorithms to check for satisfiability and validity of formulas. The preceding remark shows that the two notions are dual: an algorithm which tests validity of formulas can be converted into one for testing satisfiability and vice versa.

In principle, testing the validity of a formula α involves checking its truth value across an uncountable number of valuations. However, it is sufficient to look at the effect of valuations on the atomic propositions mentioned in α.

Let us define $\mathrm{Voc}(\alpha)$, the *vocabulary* of α, as follows:

- For $p \in \mathscr{P}$, $\mathrm{Voc}(p) = \{p\}$.
- If $\alpha = \neg\beta$, then $\mathrm{Voc}(\alpha) = \mathrm{Voc}(\beta)$.
- If $\alpha = \beta \vee \gamma$, then $\mathrm{Voc}(\alpha) = \mathrm{Voc}(\beta) \cup \mathrm{Voc}(\gamma)$.

Proposition 1.7 *Let α be a formula and v_1, v_2 be valuations. If v_1 and v_2 agree on $\mathrm{Voc}(\alpha)$ then $v_1(\alpha) = v_2(\alpha)$.*

This justifies the familiar algorithm for testing validity: build a truth-table for the propositions mentioned in α and check if all rows yield the value $\top$.

1.3 Axiomatisations

Though we have a straightforward algorithm for testing validity of formulas in propositional logic, such algorithms do not exist for more complicated logical systems. In particular, there is no such algorithm for first-order logic.

However, it is still possible to effectively enumerate all the valid formulas of first-order logic. One way of presenting such an enumeration is through an axiomatisation of the logic. To prepare the ground for studying axiomatisations of more complex logics, we begin with an axiomatisation for propositional logic.

Axiom System AX The axiom system *AX* consists of three axioms and one inference rule.

$$\begin{aligned}
&(\mathrm{A1}) && \alpha \supset (\beta \supset \alpha)\\
&(\mathrm{A2}) && (\alpha \supset (\beta \supset \gamma)) \supset ((\alpha \supset \beta) \supset (\alpha \supset \gamma))\\
&(\mathrm{A3}) && (\neg\beta \supset \neg\alpha) \supset ((\neg\beta \supset \alpha) \supset \beta)\\
&\text{(Modus Ponens, or MP)} && \frac{\alpha,\ \alpha \supset \beta}{\beta}
\end{aligned}$$

The rule MP is read as follows—from α and $\alpha \supset \beta$, infer β. It is important to note that these are *axiom schemes*—that is, they are not actual formulas but templates which can be instantiated into real formulas by consistently substituting concrete formulas for α, β and γ. For instance, if $p, q \in \mathscr{P}$, $p \supset (q \supset p)$ is an instance of axiom (A1). An alternate way to present such an axiomatisation is to list the axioms as concrete formulas and have an additional inference rule to permit uniform substitution of new formulas into an existing formula.

Derivations A *derivation* of α using the axiom system *AX* is a finite sequence of formulas $\beta_1, \beta_2, \ldots \beta_n$ such that:

- $\beta_n = \alpha$
- For each $i \in \{1, 2, \ldots, n\}$, β_i is either an instance of one of the axioms (A1)–(A3), or is obtained by applying the rule (MP) to formulas β_j, β_k, where $j, k < i$—that is, β_k is of the form $\beta_j \supset \beta_i$.

We write $\vdash_{AX} \alpha$ to denote that α is derivable using the axiom system *AX* and say that α is a *thesis* of the system. We will normally omit the subscript *AX*.

Here is an example of a derivation using our axiom system.

1.	$(p \supset ((p \supset p) \supset p)) \supset ((p \supset (p \supset p)) \supset (p \supset p))$	Instance of (A2)
2.	$p \supset ((p \supset p) \supset p)$	Instance of (A1)
3.	$(p \supset (p \supset p)) \supset (p \supset p))$	From 1 and 2 by MP
4.	$p \supset (p \supset p)$	Instance of (A1)
5.	$p \supset p$	From 3 and 4 by MP

***Exercise** 1.8* Show that $(\neg\beta \supset \neg\alpha) \equiv (\alpha \supset \beta)$ is a thesis of *AX*. ⊣

The axiom system we have presented is called a Hilbert-style axiomatisation. There are several other ways of presenting axiomatisations. One common alternative to Hilbert-style systems is the *sequent calculus* notation due to Gentzen. Typically, Hilbert-style axiomatisations have a large number of axioms and very few inference rules, while sequent calculi have very few axioms and a large number of inference rules. Sequent calculi are often easier to work with when searching for derivations, but are also more complicated from a technical point of view. We shall look at sequent calculi later, when we come to first-order logic.

Another fact worth remembering is that the axiom system *AX* defined here is just one of many possible Hilbert-style axiom systems for propositional logic.

The main technical result we would like to establish is that the set of formulas derivable using *AX* is precisely the set of valid formulas of propositional logic.

Theorem 1.9 *For all formulas α, $\vdash \alpha$ iff $\models \alpha$.*

We break up the proof of this theorem into two parts. The first half is to show that every thesis of AX is valid. This establishes the *soundness* of the axiom system,

Lemma 1.10 (Soundness) *For all formulas α, if $\vdash \alpha$ then $\models \alpha$.*

Proof: If $\vdash \alpha$, then we can exhibit a derivation $\beta_1, \beta_2, \ldots, \beta_n$ of α. Formally, the proof of the lemma is by induction on the length of this derivation. Since every formula in the sequence $\beta_1, \beta_2, \ldots, \beta_n$ is either an instance of one of the axioms or is obtained by applying the rule (MP), it suffices to show that all the axioms define valid formulas and that (MP) preserves validity—in other words, if α is valid and $\alpha \supset \beta$ is valid, then β is valid. This is straightforward and we omit the details. ⊣

The other half of Theorem 1.9 is more difficult to establish. We have to argue that every valid formula is derivable. Formally, this would show that our axiomatisation is *complete*.

We follow the approach of the logician Leon Henkin and attack the problem indirectly. Consider the contrapositive of the statement we want to prove—that is, if a formula α is *not* a thesis, then it is *not* valid.

Consistency We write $\nvdash \alpha$ to denote that α is not a thesis. We say that α is *consistent* (with respect to *AX*) if $\nvdash \neg\alpha$.

***Exercise** 1.11*

(i) Show that $\alpha \vee \beta$ is consistent iff either α is consistent or β is consistent.

(ii) Show that if $\alpha \wedge \beta$ is consistent then both α and β are consistent. Is the converse true?

(iii) Suppose that $\vdash \alpha \supset \beta$. Which of the following is true?

  (a) If α is consistent then β is consistent.

  (b) If β is consistent then α is consistent. ⊣

By Proposition 1.6 we know that α is not valid iff $\neg\alpha$ is satisfiable. Suppose we can show the following.

Lemma 1.12 (Henkin) *For all formulas β, if β is consistent then β is satisfiable.*

We can then argue that our axiomatisation is complete. Consider a formula β which is not derivable. It can be shown that $\neg\neg\beta \supset \beta$ is a thesis. If β is not derivable, neither is $\neg\neg\beta$—otherwise, we can use the rule MP to derive β from $\neg\neg\beta \supset \beta$. Since $\nvdash \neg(\neg\beta)$, $\neg\beta$ is consistent. By Lemma 1.12, $\neg\beta$ is satisfiable. Hence, by Proposition 1.6, β is not valid.

To prove Lemma 1.12, we extend the notion of consistency from a single formula to sets of formulas. A finite set of formulas $X = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is consistent if the formula $\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n$ is consistent—that is, $\nvdash \neg(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n)$. An arbitrary set of formulas $X \subseteq \Phi$ is consistent if every finite subset of X is consistent. (Henceforth, $Y \subseteq_{\text{fin}} X$ denotes that Y is a finite subset of X.)

A *maximal consistent set (MCS)* is a consistent set which cannot be extended by adding any formulas. In other words, $X \subseteq \Phi$ is an MCS iff X is consistent and for each formula $\alpha \notin X$, $X \cup \{\alpha\}$ is inconsistent.

Lemma 1.13 (Lindenbaum) *Every consistent set can be extended to an MCS.*

Proof: Let X be an arbitrary consistent set. Let $\alpha_0, \alpha_1, \alpha_2, \ldots$ be an enumeration of Φ.

We define an infinite sequence of sets $X_0, X_1, X_2, \ldots$ as follows.

- $X_0 = X$
- For $i \geq 0$, $X_{i+1} = \begin{cases} X_i \cup \{\alpha_i\} & \text{if } X_i \cup \{\alpha_i\} \text{ is consistent} \\ X_i & \text{otherwise} \end{cases}$

Each set in this sequence is consistent, by construction, and $X_0 \subseteq X_1 \subseteq X_2 \subseteq \cdots$. Let $Y = \bigcup_{i \geq 0} X_i$. We claim that Y is an MCS extending X. To establish this, we have to show that Y is consistent and that it maximal.

If Y is not consistent, then there is a subset $Z \subseteq_{\text{fin}} Y$ which is inconsistent. Let $Z = \{\beta_1, \beta_2, \ldots, \beta_n\}$. We can write Z as $\{\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_n}\}$ where the indices correspond to our enumeration of Φ. Let $j = \max(i_1, i_2, \ldots, i_n)$. Then it is clear that $Z \subseteq_{\text{fin}} X_{j+1}$ in the sequence $X_0 \subseteq X_1 \subseteq X_2 \subseteq \cdots \subseteq Y$. This implies that X_{j+1} is inconsistent, which is a contradiction.

Having established that Y is consistent, we show that it is maximal. Suppose that $Y \cup \{\beta\}$ is consistent for some formula $\beta \notin Y$. Let $\beta = \alpha_j$ in our enumeration of Φ. Since $\alpha_j \notin Y$, α_j was not added at step $j+1$ in our construction. This means that $X_j \cup \{\alpha_j\}$ is inconsistent. In other words, there exists $Z \subseteq_{\text{fin}} X_j$ such that $Z \cup \{\alpha_j\}$ is inconsistent. Since $X_j \subseteq Y$, we must have $Z \subseteq_{\text{fin}} Y$ as well, which contradicts the assumption that $Y \cup \{\alpha_j\}$ is consistent. ⊣

Maximal consistent sets have a rich structure which we shall exploit to prove completeness.

Lemma 1.14 *Let X be a maximal consistent set. Then:*

(i) For all formulas α, $\alpha \in X$ iff $\neg\alpha \notin X$.

(ii) For all formulas α, β, $\alpha \vee \beta \in X$ iff $\alpha \in X$ or $\beta \in X$.

We postpone the proof of these properties and first show how they lead to completeness.

Maximal consistent sets and valuations Let X be an MCS. Define the valuation v_X to be the set $\{p \in \mathscr{P} \mid p \in X\}$—in other words, $v_X(p) = \top$ iff $p \in X$.

Proposition 1.15 *Let X be an MCS. For all formulas α, $v_X \vDash \alpha$ iff $\alpha \in X$.*

Proof: The proof is by induction on the structure of α.

Basis: $\alpha = p$, where $p \in \mathscr{P}$. Then, $v_X \vDash p$ iff (by the definition of v_X) $p \in X$.

Induction step: There are two cases to consider—when α is of the form $\neg\beta$ and when α is of the form $\beta \vee \gamma$.

($\alpha = \neg\beta$) $v_X \vDash \neg\beta$ iff (by the definition of valuations) $v_X \nvDash \beta$ iff (by the induction hypothesis) $\beta \notin X$ iff (by the properties satisfied by MCSs) $\neg\beta \in X$.

($\alpha = \beta \vee \gamma$) $v_X \vDash \beta \vee \gamma$ iff (by the definition of valuations) $v_X \vDash \beta$ or $v_X \vDash \gamma$ iff (by the induction hypothesis) $\beta \in X$ or $\gamma \in X$ iff (by the properties satisfied by MCSs) $\beta \vee \gamma \in X$. ⊣

Thus, every MCS X defines a canonical valuation v_X which satisfies precisely those formulas that belong to X. (Conversely, every valuation also defines an MCS in a canonical way: given a valuation v, $X_v = \{\alpha \mid v \vDash \alpha\}$. It is not difficult to establish that the valuation v_{X_v} generated by X_v is exactly the same as v.)

Proposition 1.15 immediately yields a proof of Henkin's lemma.

Proof of Lemma 1.12: Let α be a consistent formula. By Lindenbaum's Lemma, $\{\alpha\}$ can be extended to an MCS X. By Proposition 1.15, $v_X \vDash \alpha$ since $\alpha \in X$. Thus, α is satisfiable. ⊣

To complete our argument, we have to prove Lemma 1.14.

Proof Sketch of Lemma 1.14: Let X be an MCS.

(i) For every formula α, we have to show that $\alpha \in X$ iff $\neg\alpha \notin X$.

We first show that $\{\alpha, \neg\alpha\} \not\subseteq X$. For this, we need the fact that $\alpha \supset \neg\neg\alpha$ and $\neg\neg\alpha \supset \alpha$ are both derivable using AX. We omit these derivations.

We know that $\alpha \supset \alpha$, or, equivalently, $\neg\alpha \vee \alpha$ is a thesis. From this, we can derive $\neg\neg(\neg\alpha \vee \alpha)$. But $\neg(\neg\alpha \vee \alpha)$ is just $\alpha \wedge \neg\alpha$, so we have $\neg(\alpha \wedge \neg\alpha)$ as a thesis. This means that $\{\alpha, \neg\alpha\}$ is inconsistent, whence it cannot be a subset of X (recalling that X is consistent).

Next we show that at least one of α and $\neg\alpha$ is in X. Suppose neither formula belongs to X. Since X is an MCS, there must be sets $B \subseteq_{\text{fin}} X$ and $C \subseteq_{\text{fin}} X$ such that $B \cup \{\alpha\}$ is inconsistent and $C \cup \{\neg\alpha\}$ are inconsistent. Let $B = \{\beta_1, \beta_2, \ldots, \beta_n\}$ and $C = \{\gamma_1, \gamma_2, \ldots, \gamma_m\}$. Let $\widehat{\beta}$ abbreviate the formula $\beta_1 \wedge \beta_2 \wedge \ldots \wedge \beta_n$ and $\widehat{\gamma}$ abbreviate the formula $\gamma_1 \wedge \gamma_2 \wedge \ldots \wedge \gamma_m$. Then, we have $\vdash \neg(\alpha \wedge \widehat{\beta})$ and $\vdash \neg(\neg\alpha \wedge \widehat{\gamma})$. Rewriting $\wedge$ in terms of $\vee$, this is equivalent to $\vdash \neg\alpha \vee \neg\widehat{\beta}$ and $\vdash \neg\neg\alpha \vee \neg\widehat{\gamma}$. From this, we can conclude that $\vdash \alpha \supset \neg\widehat{\beta}$ and $\vdash \neg\alpha \supset \neg\widehat{\gamma}$.

We now use that fact that $(\alpha \supset \beta) \supset ((\delta \supset \gamma) \supset ((\alpha \vee \delta) \supset (\beta \vee \gamma)))$ is a thesis. (Once again, we omit the derivation). Instantiating this with $\alpha = \alpha$, $\delta = \neg\alpha$, $\beta = \neg\widehat{\beta}$ and $\gamma = \neg\widehat{\gamma}$ we can derive $(\alpha \vee \neg\alpha) \supset (\neg\widehat{\beta} \vee \neg\widehat{\gamma})$. Since $\vdash \alpha \vee \neg\alpha$, we get $\vdash \neg\widehat{\beta} \vee \neg\widehat{\gamma}$. By rewriting $\vee$ in terms of $\wedge$, we can derive $\neg(\widehat{\beta} \wedge \widehat{\gamma})$. But this implies that $(B \cup C) \subseteq_{\text{fin}} X$ is inconsistent, which is a contradiction.

(ii) The proof of the second part follows in a similar manner, assuming the derivability of appropriate formulas. We omit the details. ⊣

1.5 Compactness and Strong Completeness

Often, we are not interested in absolute validity, but in restricted validity. Rather than asking whether a formula α is *always* true, we ask whether α is true in all valuations which satisfy certain properties. One way of restricting the class of valuations under consideration is to specify a set of formulas X and only look at those valuations where X is true. If α is true wherever the formulas from X are true, then α is a logical consequence of X.

Logical consequence Let X be a set of formulas and v a valuation. We write $v \vDash X$ to denote that $v \vDash \beta$ for every formula $\beta \in X$. A formula α is a *logical consequence of* X, written $X \vDash \alpha$, if for every valuation v such that $v \vDash X$ it is also the case that $v \vDash \alpha$.

The notion of logical consequence is central to the way we formalise mathematics. For instance, when we study algebraic structures such as groups, we first formulate axioms which characterise groups. Any theorem we prove about groups can be rephrased as a statement which is a logical consequence of these axioms: in other words, the theorem is true whenever the group axioms are also true.

As with validity, we now look at a syntactic approach to logical consequence.

Derivability Let X be a set of formulas. We say that a formula α is *derivable from* X, written $X \vdash \alpha$ if there exists a sequence $\alpha_1, \alpha_2, \ldots, \alpha_n$ of formulas such that $\alpha_n = \alpha$ and for $i \in \{1, 2, \ldots, n\}$, α_i is either a member of X, or an instance of one of the axioms (A1)–(A3) of AX, or is derived from α_j, α_k, $j, k < i$, using the inference rule MP. (Notice that unlike axioms, we cannot use the formulas in X as templates to generate new formulas for use in a derivation. The formulas in X are concrete formulas and must be used "as is".)

The theorem we would like to prove is the following.

Theorem 1.16 (Strong Completeness) *Let $X \subseteq \Phi$ and $\alpha \in \Phi$. Then, $X \vDash \alpha$ iff $X \vdash \alpha$.*

It is possible to prove this directly using a technique similar to the one used to prove the soundness and completeness of AX (see Exercise 1.22). However, we will prove it indirectly using two auxiliary results which are of independent interest—the Deduction Theorem and the Compactness Theorem.

We begin with the Deduction Theorem, which is a statement about derivability.

Theorem 1.17 (Deduction) *Let $X \subseteq \Phi$ and $\alpha, \beta \in \Phi$. Then, $X \cup \{\alpha\} \vdash \beta$ iff $X \vdash \alpha \supset \beta$.*

Proof: ($\Leftarrow$) Suppose that $X \vdash \alpha \supset \beta$. Then, by the definition of derivability, $X \cup \{\alpha\} \vdash \alpha \supset \beta$ as well. Since $\alpha \in X \cup \{\alpha\}$, $X \cup \{\alpha\} \vdash \alpha$. Applying MP, we get $X \cup \{\alpha\} \vdash \beta$.

($\Rightarrow$) Suppose that $X \cup \{\alpha\} \vdash \beta$. Then, there is a derivation $\beta_1, \beta_2, \ldots, \beta_n$ of β. The proof is by induction on n.

If $n = 1$, then β is either an instance of an axiom or a member of $X \cup \{\alpha\}$. If β is an instance of an axiom, then $X \vdash \beta$ as well. Further, from axiom (A1), $X \vdash \beta \supset (\alpha \supset \beta)$. Applying MP, we get $X \vdash \alpha \supset \beta$.

If $\beta \in X$, there are two cases to consider. If $\beta \in X \setminus \{\alpha\}$, then $X \vdash \beta$. Once again we have $X \vdash \beta \supset (\alpha \supset \beta)$ and hence $X \vdash \alpha \supset \beta$. On the other hand, if $\beta = \alpha$, we have $X \vdash \alpha \supset \alpha$ from the fact that $\alpha \supset \alpha$ is derivable in AX.

If $n > 1$, we look the justification for adding $\beta_n = \beta$ to the derivation. If β_n is an instance of an axiom or a member of $X \cup \{\alpha\}$, we can use the same argument as in the base case to show $X \vdash \alpha \supset \beta$.

On the other hand, if β_n was derived using MP, there exist β_i and β_j, with $i, j < n$ such that β_j is of the form $\beta_i \supset \beta_n$. By axiom (A2), $X \vdash (\alpha \supset (\beta_i \supset \beta_n)) \supset ((\alpha \supset \beta_i) \supset (\alpha \supset \beta_n))$. By the induction hypothesis, we know that $X \vdash \alpha \supset (\beta_i \supset \beta_n)$ and $X \vdash \alpha \supset \beta_i$. Applying MP twice, we get $X \vdash \alpha \supset \beta_n$. ⊣

The Deduction Theorem reflects a method of proof which is common in mathematics—proving that property x implies property y is equivalent to assuming x and inferring y.

The second step in proving Strong Completeness is the Compactness Theorem, which is a statement about logical consequence. To prove this we need the following lemma about trees, due to König.

Lemma 1.18 (König) *Let T be a finitely branching tree—that is, every node has a finite number of children (though this number may be unbounded). If T has infinitely many nodes, then T has an infinite path.*

Proof: Let T be a finitely branching tree with infinitely many nodes. Call a node x in T *bad* if the subtree rooted at x has infinitely many nodes. Clearly, if a node x is bad, at least one of its children must be bad: x has only finitely many children and if all of them were good, the subtree rooted at x would be finite.

We now construct an infinite path $x_0 x_1 x_2 \ldots$ in T. Since T has an infinite number of nodes, the root of T is a bad node. Let x_0 be the root of T. It has at least one bad successor. Pick one of the bad successors of x_0 and designate it x_1. Pick one of the bad successors of x_1 and designate it x_2, and so on. ⊣

Theorem 1.19 (Compactness) *Let $X \subseteq \Phi$ and $\alpha \in \Phi$. Then $X \models \alpha$ iff there exists $Y \subseteq_{\text{fin}} X$, $Y \models \alpha$.*

We shall first prove the following related result. Let X be a set of formulas. We say that X is satisfiable if there exists a valuation v such that $v \models X$.

Lemma 1.20 (Finite satisfiability) *Let $X \subseteq \Phi$. Then, X is satisfiable iff every $Y \subseteq_{\text{fin}} X$ is satisfiable.*

Proof: ($\Rightarrow$) Suppose X is satisfiable. Then, there is a valuation v such that $v \models X$. Clearly, $v \models Y$ for each $Y \subseteq_{\text{fin}} X$ as well.

($\Leftarrow$) Suppose X is *not* satisfiable. We have to show that there exists $Y \subseteq_{\text{fin}} X$ which is not satisfiable.

Assume that our set of atomic propositions $\mathscr{P}$ is enumerated $\{p_1, p_2, \ldots\}$. Let $\mathscr{P}_0 = \emptyset$ and for $i \in \{1, 2, \ldots\}$, let $\mathscr{P}_i = \{p_1, p_2, \ldots, p_i\}$. For $i \in \{1, 2, \ldots\}$, let Φ_i be the set of formulas generated using only atomic propositions from $\mathscr{P}_i$ and let $X_i = X \cap \Phi_i$.

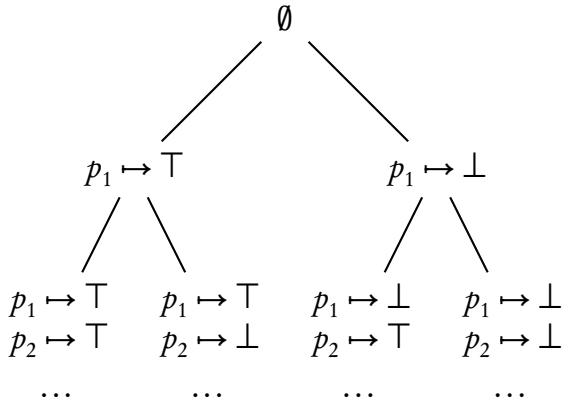


Figure 1: The tree T in the proof of Lemma 1.20

We construct a tree T whose nodes are valuations over the sets $\mathscr{P}_i$, $i \in \{0, 1, 2, \ldots\}$. More formally, the set of nodes is given by $\{v \mid \exists i \in \{0, 1, 2, \ldots\}.\ v : \mathscr{P}_i \to \{\top, \bot\}\}$. The root of T is the unique function $\emptyset \to \{\top, \bot\}$.

The relation between nodes is given as follows. Let $v : \mathscr{P}_i \to \{\top, \bot\}$. Then v has two children $v', v'' : \mathscr{P}_{i+1} \to \{\top, \bot\}$, where v' extends v to $\mathscr{P}_{i+1}$ by setting p_{i+1} to $\top$ and v'' extends v to $\mathscr{P}_{i+1}$ by setting p_{i+1} to $\bot$. More formally, for each $p \in \mathscr{P}_i$, $v'(p) = v''(p) = v(p)$ and $v'(p_{i+1}) = \top$ and $v''(p_{i+1}) = \bot$. (See Figure 1).

Observe that T is a complete infinite binary tree. The nodes at level i of the tree consist of all possible valuations over $\mathscr{P}_i$—there are precisely 2^i such valuations for each i. Notice that if v_j at level j is an ancestor of v_i at level i then v_i agrees with v_j on the atomic propositions in $\mathscr{P}_j$.

The infinite paths in T are in 1-1 correspondence with valuations over $\mathscr{P}$. Let $\pi = v_0 v_1 v_2 \ldots$ be an infinite path in the tree. The valuation $v_\pi : \mathscr{P} \to \{\top, \bot\}$ is given by $p_i \mapsto v_i(p_i)$ for $i \in \{1, 2, \ldots\}$. Conversely, given a valuation $v : \mathscr{P} \to \{\top, \bot\}$, we can define a unique path $\pi_v = v_0 v_1 v_2 \ldots$ by setting v_0 to be the root of T and $v_i : \mathscr{P}_i \to \{\top, \bot\}$ to be the restriction of v to $\mathscr{P}_i$—that is, for all $p \in \mathscr{P}_i$, $v_i(p) = v(p)$. It is easy to verify that these two maps are inverses of each other.

Let us call a node v in T bad if $v(\beta) = \bot$ for some $\beta \in X$. Clearly, if v is bad, then so is every valuation in the subtree rooted at v. We prune T by deleting all bad nodes which also have bad ancestors. (Equivalently, along any path in T, we retain only those nodes upto and including the first bad node along the path.) It is not difficult to verify that the set of nodes which remains forms a subtree T' of T all of whose leaf nodes are bad and all of whose non-leaf nodes are not bad.

We claim that T' has only a finite number of nodes. Assuming that this is true, let the set of leaf nodes of T' be $\{v_1, v_2, \ldots, v_m\}$. Since each v_i is bad, there is a corresponding formula $\beta_i \in X$ such that $v_i(\beta_i) = \bot$. We claim that $\{\beta_1, \beta_2, \ldots, \beta_m\} \subseteq_{\text{fin}} X$ is not satisfiable. Consider any valuation v. The corresponding path π_v must pass through one of the nodes in $\{v_1, v_2, \ldots, v_m\}$, say v_j. But then, $v_{\pi_v}(\beta_j) = v_j(\beta_j) = \bot$. Thus, $v \not\models \{\beta_1, \beta_2, \ldots, \beta_m\}$.

To see why T' must be finite, suppose instead that it has an infinite set of nodes. Then, by König's Lemma, it contains an infinite path $\pi = v_0 v_1 v_2 \ldots$ such that none of the nodes along this path is

bad. The path π is also an infinite path in T. We know that π defines a valuation v_π. Consider any formula $\beta \in X$. Then $\beta \in X_j$ for some $j \in \{1,2,\ldots\}$, so $v_\pi(\beta) = v_j(\beta) = \top$. Thus, $v_\pi \vDash X$, which contradicts our assumption that X is not satisfiable. ⊣

We can now complete our proof of compactness.

Proof of Theorem 1.19 (Compactness):

($\Leftarrow$) If $Y \subseteq_{\text{fin}} X$ and $Y \vDash \alpha$ then it is clear that $X \vDash \alpha$. For, if $v \vDash X$, then $v \vDash Y$ as well and, by the assumption that $Y \vDash \alpha$, $v \vDash \alpha$ as required.

($\Rightarrow$) For all $Z \subseteq \Phi$ and all $\beta \in \Phi$, it is clear that $Z \vDash \beta$ iff $Z \cup \{\neg\beta\}$ is not satisfiable.

Suppose $X \vDash \alpha$. Then, $X \cup \{\neg\alpha\}$ is not satisfiable. By Lemma 1.20, there is a subset $Y \subseteq_{\text{fin}} X \cup \{\neg\alpha\}$ such that Y is not satisfiable. Thus, $(Y \setminus \{\neg\alpha\}) \cup \{\neg\alpha\}$ is not satisfiable either, where $(Y \setminus \{\neg\alpha\}) \subseteq_{\text{fin}} X$. This implies that $Y \setminus \{\neg\alpha\} \vDash \alpha$. ⊣

With the Deduction Theorem and the Compactness Theorem behind us, we can prove Strong Completeness.

Proof of Theorem 1.16 (Strong Completeness):

To show that $X \vdash \alpha$ implies $X \vDash \alpha$ is routine. Conversely, suppose that $X \vDash \alpha$. By compactness, there is a finite subset $Y \subseteq_{\text{fin}} X$ such that $Y \vDash \alpha$. Let $Y = \{\beta_1, \beta_2, \ldots, \beta_m\}$. It is then easy to see that $\beta_1 \supset (\beta_2 \supset \cdots (\beta_m \supset \alpha)\cdots)$ is valid. Hence, by the completeness theorem for propositional logic, $\vdash \beta_1 \supset (\beta_2 \supset \cdots (\beta_m \supset \alpha)\cdots)$. Applying the Deduction Theorem m times we get $\{\beta_1, \beta_2, \ldots, \beta_m\} \vdash \alpha$. Since $\{\beta_1, \beta_2, \ldots, \beta_m\} \subseteq X$, it follows that $X \vdash \alpha$. ⊣

Observe that we could alternatively derive compactness from strong completeness. If $X \vDash \alpha$ then, by strong completeness, $X \vdash \alpha$. We let $Y \subseteq_{\text{fin}} X$ be the subset of formulas actually used in the derivation of α. Thus, $Y \vdash \alpha$ as well. By the other half of strong completeness, $Y \vDash \alpha$.

We conclude our discussion of propositional logic with two exercises. The first leads to an alternative proof of compactness which is more along the lines of the completeness proof for propositional logic. The second exercise leads to a direct proof of strong completeness.

Exercise 1.21 **(Compactness)**

Let X be a set of formulas. X is said to be a *finitely satisfiable set (FSS)* if every $Y \subseteq_{\text{fin}} X$ is satisfiable.

Equivalently, X is an FSS if there is no finite subset $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ of X such that $\neg(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n)$ is valid.

(Note that if X is an FSS we are not promised a single valuation v which satisfies every finite subset of X. Each finite subset could be satisfied by a *different* valuation).

Show that:

(i) Every FSS can be extended to a maximal FSS.

(ii) If X is a maximal FSS then:

  (a) For every formula α, $\alpha \in X$ iff $\neg\alpha \notin X$.

  (b) For all formulas α, β, $(\alpha \vee \beta) \in X$ iff ($\alpha \in X$ or $\beta \in X$).

(iii) Every maximal FSS X generates a valuation v_X such that for every formula α, $v_X \models \alpha$ iff $\alpha \in X$.

From these facts conclude that:

(iv) Any FSS X is simultaneously satisfiable (that is, for any FSS X, there exists v_X such that $v_X \models X$).

(v) For all X and all α, $X \models \alpha$ iff there exists $Y \subseteq_{\text{fin}} X$ such that $Y \models \alpha$. ⊣

***Exercise** 1.22* **(Strong Completeness)**
We define a new notion of consistency. A set X is said to be *consistent* if there is no formula α such that $X \vdash \alpha$ and $X \vdash \neg\alpha$.

Show that:

(i) X is consistent iff every finite subset of X is consistent.

(ii) Every consistent set X can be extended to a maximal consistent set (MCS).

(iii) Every MCS X generates a valuation v_X such that for all formulas α, $v_X \models \alpha$ iff $\alpha \in X$.

(iv) Every consistent set X is satisfiable: that is, there exists a valuation v_X such that $v_X \models X$.

(v) If $X \models \alpha$ then $X \cup \{\neg\alpha\}$ is not consistent.

(vi) Use the Deduction Theorem to show that that if $X \models \alpha$ then $X \vdash \neg\alpha \supset (\beta \wedge \neg\beta)$ for some formula β.

Conclude that if $X \models \alpha$ then $X \vdash \alpha$. ⊣

2 Modal Logic

In propositional logic, a valuation is a static assignment of truth values to atomic propositions. In computer science applications, atomic propositions describe properties of the current state of a program. It is natural to expect that the truth of an atomic proposition varies as the state changes. Modal logic is a framework to describe such a situation.[1]

The basic idea in modal logic is to look at a collection of possible valuations simultaneously. Each valuation represents a possible state of the world. Separately, we specify how these "possible worlds" are connected to each other. We then enrich our logical language with a way of referring to truth across possible worlds.

2.1 Syntax

As in propositional logic, we begin with a countably infinite set of atomic propositions $\mathscr{P} = \{p_0, p_1, \ldots\}$ and two logical connectives $\neg$ (read as *not*) and $\vee$ (read as *or*). We add a unary *modality* $\Box$ (read as *box*).

The set Φ of formulas of modal logic is the smallest set satisfying the following:

- Every atomic proposition p is a member of Φ.
- If α is a member of Φ, so is $(\neg\alpha)$.
- If α and β are members of Φ, so is $(\alpha \vee \beta)$.
- If α is a member of Φ, so is $(\Box\alpha)$.

As before, we omit parentheses if there is no ambiguity. The derived propositional connectives $\wedge$, $\supset$ and $\equiv$ are defined as before. In addition, we have a derived modality $\Diamond$ (read *diamond*) which is *dual* to the modality $\Box$, defined as follows: $\Diamond\alpha \stackrel{\text{def}}{=} \neg\Box\neg\alpha$.

2.2 Semantics

Frames A *frame* is a structure $F = (W, R)$, where W is a set of *possible worlds* and $R \subseteq W \times W$ is the *accessibility relation*. If $w\ R\ w'$, we say that w' is an R-neighbour of w.

In more familiar terms, a frame is just a directed graph over the set of nodes W. We do not make any assumptions about the set W—not even the fact that it is countable.

[1] Traditional modal logic arose out of philosophical enquiries into the nature of necessary and conditional truth. We shall concentrate on the technical aspects of the subject and avoid all discussion of the philosophical foundations of modal logic.

Models A model is a pair $M = (F, V)$ where $F = (W, R)$ is a frame and $V : W \to 2^{\mathscr{P}}$ is a *valuation*.[2]

Recall that a propositional valuation $v : \mathscr{P} \to \{\top, \bot\}$ can also be viewed as a set $v \subseteq \mathscr{P}$ consisting of those atomic propositions p such that $v(p) = \top$. We have implicitly used this when defining valuations in modal logic. Formally, V is a function which assigns a propositional valuation to each world in W—in other words, for each $w \in W$, $V(w) : \mathscr{P} \to \{\top, \bot\}$. Thus, V is actually a function of the form $W \to (\mathscr{P} \to \{\top, \bot\})$, which we abbreviate as $V : W \to 2^{\mathscr{P}}$.

Satisfaction The notion of truth is localised to each world in a model. We write $M, w \vDash \alpha$ to denote that α is true at the world w in the model M. The satisfaction relation is defined inductively as follows.

$$
\begin{array}{ll}
M, w \vDash p & \text{iff } p \in V(w) \text{ for } p \in \mathscr{P} \\
M, w \vDash \neg\alpha & \text{iff } M, w \nvDash \alpha \\
M, w \vDash \alpha \vee \beta & \text{iff } M, w \vDash \alpha \text{ or } M, w \vDash \beta \\
M, w \vDash \Box\alpha & \text{iff for each } w' \in W, \text{ if } w \; R \; w' \text{ then } M, w' \vDash \alpha
\end{array}
$$

Thus, $M, w \vDash \Box\alpha$ if every world accessible from w satisfies α. Notice that if w is isolated—that is, there is no world w' such that $w \; R \; w'$—then $M, w \vDash \Box\alpha$ for *every* formula α.

***Exercise** 2.1* Verify that $M, w \vDash \Diamond\alpha$ iff there exists w', $w \; R \; w'$ and $M, w' \vDash \alpha$. ⊣

Satisfiability and validity As usual, we say that α is *satisfiable* if there exists a frame $F = (W, R)$ and a model $M = (F, V)$ such that $M, w \vDash \alpha$ for some $w \in W$. The formula α is *valid*, written $\vDash \alpha$, if for every frame $F = (W, R)$, for every model $M = (F, V)$ and for every $w \in W$, $M, w \vDash \alpha$.

***Example** 2.2* Here are some examples of valid formulas in modal logic.

(i) Every tautology of propositional logic is valid. Consider a tautology α and a world w in a model $M = ((W, R), V)$. Since the truth of α depends only on $V(w)$, and α is true under all propositional valuations, $M, w \vDash \alpha$.

(ii) The formula $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$ is valid. Consider a model $M = ((W, R), V)$ and a world $w \in W$. Suppose that $M, w \vDash \Box(\alpha \supset \beta)$. We must argue that $M, w \vDash \Box\alpha \supset \Box\beta$. Let $M, w \vDash \Box\alpha$. Then we must show that $M, w \vDash \Box\beta$. In other words, we must show that every R-neighbour w' of w satisfies β. Since we assumed $M, w \vDash \Box(\alpha \supset \beta)$, we know that $M, w' \vDash \alpha \supset \beta$. Moreover, since $M, w \vDash \Box\alpha$, $M, w' \vDash \alpha$. By the semantics of the connective $\supset$, it follows that $M, w' \vDash \beta$, as required.

(iii) Suppose that α is valid. Then, $\Box\alpha$ must also be valid. Consider any model $M = ((W, R), V)$ and any $w \in W$. To check that $M, w \vDash \Box\alpha$ we have to verify that every R-neighbour of w satisfies α. Since α is valid, $M, w' \vDash \alpha$ for all $w' \in W$. So, every R-neighbour of w does satisfy α and $M, w \vDash \Box\alpha$.

[2] The semantics we describe here was first formalised by Saul Kripke, so these models are often called Kripke models in the literature.

***Exercise** 2.3* The argument given in part (i) of Exercise 2.2 applies only to non-modal instances of propositional tautologies—for instance, the explanation does not justify the validity of the formula $\Box\alpha \vee \neg\Box\alpha$. Show that *all* substitution instances of propositional tautologies are valid formulas in modal logic. ⊣

As in propositional logic, one of our central concerns in modal logic is to be able to decide when formulas are satisfiable (or, dually, valid). Notice that unlike the truth-table based algorithm for propositional logic, there is no obvious decision procedure for satisfiability in modal logic. To check satisfiability of a formula α, though it suffices to look at valuations over the vocabulary of α, we also have to specify an underlying frame. There is no a priori bound on the size of this frame.

Later in this section we will describe a sound and complete axiomatisation for modal logic. This will give us an effective way of enumerating all valid formulas. After that, we will encounter a technique by which we can bound the size of the underlying frame required to satisfy a formula α. But, we first examine an aspect of modal logic which does not have any counterpart in propositional logic.

2.3 Correspondence Theory

The modalities $\Box$ and $\Diamond$ can be used to describe interesting properties of the accessibility relation R of a frame $F = (W, R)$. This area of modal logic is called *correspondence theory*.

Let α be a formula of modal logic. With α, we identify a class of frames $\mathscr{C}_\alpha$ as follows:

> $F = (W, R) \in \mathscr{C}_\alpha$ iff for every valuation V over W, for every world $w \in W$ and for every substitution instance β of α, $((W, R), V), w \models \beta$.

In other words, when defining $\mathscr{C}_\alpha$, we interpret α as a template, much like an axiom scheme. Notice that for any frame $F = (W, R)$ which does not belong to $\mathscr{C}_\alpha$, we can find a valuation V, a world w and a substitution instance β of α such that $((W, R), V), w \not\models \beta$.

Characterising classes of frames We say a class of frames $\mathscr{C}$ is *characterised* by the formula α if $\mathscr{C} = \mathscr{C}_\alpha$.

We now look at some examples of frame conditions which can characterised by formulas of modal logic.

Proposition 2.4 *The class of reflexive frames is characterised by the formula* $\Box\alpha \supset \alpha$.

Proof: We first show that every reflexive frame belongs to $\mathscr{C}_{\Box\alpha\supset\alpha}$. Let $M = ((W, R), V)$ be a model where R is reflexive. Consider any world $w \in W$. Suppose that $M, w \models \Box\alpha$. We have to show that $M, w \models \alpha$ as well. Since $M, w \models \Box\alpha$, every R-neighbour of w satisfies α. But R is reflexive, so w is an R-neighbour of itself. Hence, $M, w \models \alpha$.

Conversely, we show that every non-reflexive frame does not belong to $\mathscr{C}_{\Box\alpha\supset\alpha}$. Let $F = (W, R)$ be a frame where for some $w \in W$, it is not the case that $w \; R \; w$. Choose a proposition p and define a valuation V as follows: $V(w) = \emptyset$ and $V(w') = \{p\}$ for all $w' \neq w$. Clearly, $(F, V), w \models \Box p$ but $(F, V), w \not\models p$. Hence w fails to satisfy the substitution instance $\Box p \supset p$ of the formula $\Box\alpha \supset \alpha$. ⊣

Proposition 2.5 *The class of transitive frames is characterised by the formula* $\Box\alpha \supset \Box\Box\alpha$.

Proof: We first show that every transitive frame belongs to $\mathscr{C}_{\Box\alpha\supset\Box\Box\alpha}$. Let $M = ((W,R),V)$ be a model where R is transitive. Consider any world $w \in W$. Suppose that $M, w \vDash \Box\alpha$. We have to show that $M, w \vDash \Box\Box\alpha$ as well.

For this, we have to show that every R-neighbour w' of w satisfies $\Box\alpha$. Consider any R-neighbour w' of w. If w' has no R-neighbours, then it is trivially the case that $M, w' \vDash \Box\alpha$. On the other hand, if w' has R-neighbours, then we must show that each R-neighbour of w' satisfies α. Let w'' be an R-neighbour of w'. Since $w\ R\ w'$ and $w'\ R\ w''$, by transitivity w'' is also an R-neighbour of w. Since we assumed that $M, w \vDash \Box\alpha$, it must be the case that $M, w'' \vDash \alpha$, as required.

Conversely, we show that every non-transitive frame does not belong to $\mathscr{C}_{\Box\alpha\supset\Box\Box\alpha}$. Let $F = (W,R)$ be a frame where for some $w, w', w'' \in W$, $w\ R\ w'$ and $w'\ R\ w''$ but it is not the case that that $w\ R\ w''$. Choose a proposition p and define a valuation V as follows:

$$V(\widehat{w}) = \begin{cases} \{p\} & \text{if } w\ R\ \widehat{w} \\ \emptyset & \text{otherwise} \end{cases}$$

Since w'' is not an R-neighbour of w, $V(w'') = \emptyset$. This means that $M, w' \nvDash \Box p$, for w'' is an R-neighbour of w' and $M, w'' \nvDash p$. Therefore, $M, w \nvDash \Box\Box p$, since w' is an R-neighbour of w. On the other hand, $M, w \vDash \Box p$ by the definition of V. Hence, $M, w \nvDash \Box p \supset \Box\Box p$, which is an instance of $\Box\alpha \supset \Box\Box\alpha$. ⊣

The characteristic formula for transitivity can dually be written $\Diamond\Diamond\alpha \supset \Diamond\alpha$. This form represents transitivity more intuitively—the formula says that if $w\ R\ w'\ R\ w''$ and w'' satisfies α, there exists an R-neighbour $\widehat{w}$ of w satisfying α. If R is transitive, w'' is a natural candidate for $\widehat{w}$. Similarly, $\alpha \supset \Diamond\alpha$ is the dual (and more appealing) form of the characteristic formula for reflexivity. We have used the $\Box$ forms of these formulas because they are more standard in the literature.

Proposition 2.6 *The class of symmetric frames is characterised by the formula* $\alpha \supset \Box\Diamond\alpha$.

Proof: We first show that every symmetric frame belongs to $\mathscr{C}_{\alpha\supset\Box\Diamond\alpha}$. Let $M = ((W,R),V)$ be a model where R is symmetric. Consider any world $w \in W$. Suppose that $M, w \vDash \alpha$. We have to show that $M, w \vDash \Box\Diamond\alpha$ as well.

For this, we have to show that every R-neighbour w' of w satisfies $\Diamond\alpha$. Consider any R-neighbour w' of w. Since R is symmetric, w is an R-neighbour of w'. We assumed that $M, w \vDash \alpha$ so $M, w' \vDash \Diamond\alpha$, as required.

Conversely, we show that every non-symmetric frame does not belong to $\mathscr{C}_{\alpha\supset\Box\Diamond\alpha}$. Let $F = (W,R)$ be a frame where for some $w, w' \in W$, $w\ R\ w'$ but it is not the case that that $w'\ R\ w$. Choose a proposition p and define a valuation V as follows:

$$V(\widehat{w}) = \begin{cases} \emptyset & \text{if } w'\ R\ \widehat{w} \\ \{p\} & \text{otherwise} \end{cases}$$

By construction $M, w' \nvDash \Diamond p$. Hence, since wRw', $M, w \nvDash \Box\Diamond p$. On the other hand, $M, w \vDash p$ by the definition of V, so $M, w \nvDash p \supset \Box\Diamond p$, which is an instance of the formula $\alpha \supset \Box\Diamond\alpha$. ⊣

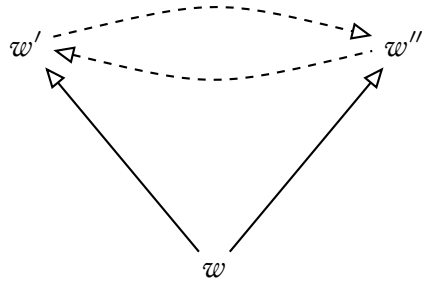


Figure 2: The Euclidean condition

We say that an accessibility relation R over W is *Euclidean* if for all $w, w', w'' \in W$, if $w\ R\ w'$ and $w\ R\ w''$ then $w'\ R\ w''$ and $w''\ R\ w'$ (see Figure 2).

Proposition 2.7 *The class of Euclidean frames is characterised by the formula* $\Diamond\alpha \supset \Box\Diamond\alpha$.

Proof: We first show that every Euclidean frame belongs to $\mathscr{C}_{\Diamond\alpha\supset\Box\Diamond\alpha}$. Let $M = ((W,R),V)$ be a model where R is Euclidean. Consider any world $w \in W$. Suppose that $M, w \vDash \Diamond\alpha$. We have to show that $M, w \vDash \Box\Diamond\alpha$ as well.

Let w' be an R-neighbour of w. We must show that $M, w' \vDash \Diamond\alpha$. Since $M, w \vDash \Diamond\alpha$, there must exist w_α such that $w\ R\ w_\alpha$ and $M, w_\alpha \vDash \alpha$. Since R is Euclidean, $w'\ R\ w_\alpha$ as well, so $M, w' \vDash \Diamond\alpha$ as required.

Conversely, we show that every non-Euclidean frame does not belong to $\mathscr{C}_{\Diamond\alpha\supset\Box\Diamond\alpha}$. Let $F = (W,R)$ be a frame where for some $w, w', w'' \in W$, $w\ R\ w'$ and $w\ R\ w''$ but one of $w'\ R\ w''$ and $w''\ R\ w'$ fails to hold. Without loss of generality, assume that it is not the case that $w''\ R\ w'$.

Choose a proposition p and define a valuation V such that $V(w') = \{p\}$ and $V(\widehat{w}) = \emptyset$ for all $\widehat{w} \neq w'$. Then, since $w\ R\ w'$, $M, w \vDash \Diamond p$ by the definition of V. On the other hand, by construction $M, w'' \nvDash \Diamond p$, so $M, w \nvDash \Box\Diamond p$. So, $M, w \nvDash \Diamond p \supset \Box\Diamond p$, which is an instance of $\Diamond\alpha \supset \Box\Diamond\alpha$. ⊣

Notice that if R is Euclidean, for all w', if there exists w such that $w\ R\ w'$, then $w'\ R\ w'$. It is not difficult to verify that if R is reflexive and Euclidean then R is in fact an equivalence relation. On the other hand, if R is symmetric and transitive then it is also Euclidean.

A frame (W,R) is said to be *converse well-founded* if for all nonempty subsets X of W, there exists a maximal element x of X, i.e. x is in X and for all y in X, it is not the case that $x\ R\ y$.

Proposition 2.8 *The class of transitive, converse well-founded frames is characterised by the formula* $\Box(\Box\alpha \supset \alpha) \supset \Box\alpha$.

Proof: We first show that every transitive and converse well-founded frame is a model of $\Box(\Box\alpha \supset \alpha) \supset \Box\alpha$, i.e., it belongs to $\mathscr{C}_{\Box(\Box\alpha\supset\alpha)\supset\Box\alpha}$. Let $M = ((W,R),V)$ be a model where R is transitive and converse well-founded. Consider any world $w \in W$. Suppose that $M, w \vDash \Box(\Box\alpha \supset \alpha)$. We have to show that $M, w \vDash \Box\alpha$ as well.

For this, we have to show that every R-neighbour w' of w satisfies α. Consider any R-neighbour w' of w. Since w satisfies $\Box(\Box\alpha \supset \alpha)$, w' satisfies $\Box\alpha \supset \alpha$. Thus, to show that every R-neighbour w' of w satisfies α it suffices to show that w' satisfies $\Box\alpha$.

Consider the set X of worlds x such that $w\ R\ x$. Since R is transitive, whenever x is an element of X and $x\ R\ y$, we also have $w\ R\ y$ and hence y is in X. A path in W is any finite sequence $\rho = w_0, w_1, \ldots, w_n$ of worlds ($n \geq 0$) such that for all $i : 0 < i \leq n$, $w_i\ R\ w_{i+1}$. The length of such a path, denoted $len(\rho)$, is defined to be n. A path $\rho = w_0, w_1, \ldots, w_n$ is said to be an x-path (for $x \in W$) if $x = w_0$. For any node $x \in W$, define the *height* of x, denoted $ht(x)$ to be $sup\{len(\rho) \mid \rho$ is an x-path$\}$. The height of a given world is in general an ordinal. But the following useful property holds: whenever $x\ R\ y$ then $ht(y) < ht(x)$.

For all $x \in X$, we prove by transfinite induction on $ht(x)$ that x satisfies $\Box\alpha$ (and hence α). The base case is when $ht(x)$ is 0, which means that there is no $y \in W$ such that $x\ R\ y$. But then x vacuously satisfies $\Box\alpha$. For the induction step, consider an arbitrary world x in X. For all $y \in W$ such that $x\ R\ y$, $y \in X$ and $ht(y)$ is *strictly less than* $ht(x)$. Therefore by the induction hypothesis every R-neighbour y of x satisfies $\Box\alpha$ (and hence α), and hence x satisfies $\Box\alpha$ (and hence α).

Thus every R-neighbour w' of w satisfies α, and hence w satisfies $\Box\alpha$.

Conversely, consider a frame $F = (W, R)$ which is not transitive. This means that there are three worlds of W, w, w', and w'' such that $w\ R\ w'$, $w'\ R\ w''$, but not $w\ R\ w''$. Choose a proposition p and define a valuation V as follows:

$$V(\widehat{w}) = \begin{cases} \{p\} & \text{if } w\ R\ \widehat{w} \text{ and } \widehat{w} \neq w' \\ \emptyset & \text{otherwise} \end{cases}$$

Clearly, for all $\widehat{w}$ in W such that $w\ R\ \widehat{w}$ and $w' \neq \widehat{w}$, $\widehat{w}$ satisfies p and hence $\Box p \supset p$. On the other hand w' does not satisfy $\Box p$ (since it has an R-neighbour, namely w'', which does not satisfy p) and hence satisfies $\Box p \supset p$ vacuously. Since all R-neighbours of w satisfy $\Box p \supset p$, w satisfies $\Box(\Box p \supset p)$. On the other hand, clearly w does not satisfy $\Box p$.

Consider now a frame $F = (W, R)$ which is transitive but not converse well-founded. This means that there is a subset X of W with no R-maximal world, i.e. for all x in X, there is a y in X such that $x\ R\ y$. Choose a world w in X, choose a proposition p and define a valuation V as follows:

$$V(\widehat{w}) = \begin{cases} \emptyset & \text{if } \widehat{w} \in X \\ \{p\} & \text{otherwise} \end{cases}$$

Clearly, for all $\widehat{w} \notin X$, $\widehat{w}$ satisfies p and hence $\Box p \supset p$. On the other hand, every $\widehat{w}$ in X has an R-neighbour in X (which does not satisfy p) and hence $\widehat{w}$ does not satisfy $\Box p$ and thus satisfies $\Box p \supset p$. Thus w satisfies $\Box(\Box p \supset p)$. But clearly w does not satisfy $\Box p$. ⊣

***Exercise** 2.9* What classes of frames are characterised by the following formulas?

(i) $\Diamond\alpha \supset \Box\alpha$.

(ii) $\Diamond\alpha \supset \Diamond\Diamond\alpha$.

(iii) $\alpha \supset \Box\alpha$. ⊣

Are there natural classes of frames which *cannot* be characterised in modal logic? We will see later that irreflexive frames form one such class. But first, we return to the notions of satisfiability and validity and look for a completeness result.

2.4 Axiomatising valid formulas

Validity revisited We said earlier that a formula α is valid if for every frame $F = (W, R)$, every model $M = (F, V)$ and every world w, $M, w \vDash \alpha$. In light of our discussion of correspondence theory we can refine this notion by restricting the range over which we consider frames.

Let $\mathscr{C}$ be a class of frames. We say that a formula α is $\mathscr{C}$-valid if for every frame $F = (W, R)$ from the class $\mathscr{C}$, for every model $M = (F, V)$ and for every world w, $M, w \vDash \alpha$. We denote the fact that α is $\mathscr{C}$-valid by $\vDash_{\mathscr{C}} \alpha$.

Let $\mathscr{F}$ represents the class of all frames. Then, the set of $\mathscr{F}$-valid formulas is the same as the set of valid formulas according to our earlier definition. In other words, the notions $\vDash_{\mathscr{F}} \alpha$ and $\vDash \alpha$ are equivalent.

Dually, we say that a formula α is $\mathscr{C}$-satisfiable if there is a frame $F = (W, R)$ in the class $\mathscr{C}$, a model $M = (F, V)$ and a world w, such that $M, w \vDash \alpha$. Once again, a formula is $\mathscr{F}$-satisfiable iff it is satisfiable according to our earlier definition.

Completeness for the class $\mathscr{F}$

Consider the following axiom system.

Axiom System K

Axioms

(Ao) All tautologies of propositional logic.
(K) $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$.

Inference Rules

(MP) $\dfrac{\alpha, \alpha \supset \beta}{\beta}$ (G) $\dfrac{\alpha}{\Box\alpha}$

The axiom (Ao) is an abbreviation for any set of axioms which are sound and complete for Propositional Logic—in particular, we could instantiate (Ao) with the axioms (A1)–(A3) of the system *AX* discussed in the previous section.

As usual, we say that α is a *thesis* of System K[3], denoted $\vdash_K \alpha$, if we can derive α using the axioms (Ao) and (K) and the inference rules (MP) and (G). Once again, we will omit the subscript and write $\vdash \alpha$ if there is no confusion about which axiom system we are referring to.

[3] The name K is derived from Saul Kripke.

The result we want to establish is the following.

Theorem 2.10 *For all formulas α, $\vdash_K \alpha$ iff $\models_{\mathscr{F}} \alpha$.*

As usual, one direction of the proof is easy.

Lemma 2.11 (Soundness of System K) *If $\vdash_K \alpha$ then $\models_{\mathscr{F}} \alpha$.*

Proof: As we observed in the previous section, it suffices to show that each axiom is $\mathscr{F}$-valid and that the inference rules preserve $\mathscr{F}$-validity. This is precisely what we exhibited in Example 2.2 and Exercise 2.3. ⊣

As in Propositional Logic, we use a Henkin-style argument to show that every $\mathscr{F}$-valid formula is derivable using System K.

Consistency As before, we say that a formula α is *consistent* with respect to System K if $\nvdash_K \neg\alpha$. A finite set of formulas $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is consistent if the conjunction $\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_n$ is consistent. Finally, an arbitrary set of formulas X is consistent if every finite subset of X is consistent.

Our goal is to prove the following.

Lemma 2.12 *Let α be a formula which is consistent with respect to System K. Then, α is $\mathscr{F}$-satisfiable.*

As we saw in the case of Propositional Logic, this will yield as an immediate corollary the result which we seek:

Corollary 2.13 (Completeness for System *K*) *Let α be a formula which is $\mathscr{F}$-valid. Then, $\vdash_K \alpha$.*

Maximal Consistent Sets

As before, we say that a set of formulas X is a *maximal consistent set* or MCS if X is consistent and for all $\alpha \notin X$, $X \cup \{\alpha\}$ is inconsistent. As we saw earlier, by Lindenbaum's Lemma, every consistent set of formulas can be extended to an MCS.

We will once again use the properties of MCSs established in Lemma 1.14. In addition, the following properties of MCSs will prove useful.

Lemma 2.14 *Let X be a maximal consistent set.*

(i) If β is a substitution instance of an axiom, then $\beta \in X$.

(ii) If $\alpha \supset \beta \in X$ and $\alpha \in X$, then $\beta \in X$.

Proof: The proof is routine and is left as an exercise. ⊣

When we studied propositional logic, we saw that each maximal consistent set defines a "propositional world". In modal logic, we have to construct frames with many propositional worlds. In fact, we generate a frame with *all* possible worlds, with a suitable accessibility relation.

Canonical model The *canonical frame* for System K is the pair $F_K = (W_K, R_K)$ where:

- $W_K = \{X \mid X \text{ is an MCS}\}$.
- If X and Y are MCSs, then $X\ R_K\ Y$ iff $\{\alpha \mid \Box\alpha \in X\} \subseteq Y$.

The *canonical model* for System K is given by $M_K = (F_K, V_K)$ where for each $X \in W_K$, $V_K(X) = X \cap \mathscr{P}$.

***Exercise** 2.15* We can dually define R_K using the modality $\Diamond$ rather than $\Box$. Verify that $X\ R_K\ Y$ iff $\{\Diamond\alpha \mid \alpha \in Y\} \subseteq X$. ⊣

The heart of the completeness proof is the following lemma.

Lemma 2.16 *For each MCS $X \in W_K$ and for each formula $\alpha \in \Phi$, $M_K, X \vDash \alpha$ iff $\alpha \in X$.*

Proof: As usual, the proof is by induction on the structure of α.
Basis: If $\alpha = p \in \mathscr{P}$, then $M_K, X \vDash p$ iff $p \in V_K(X)$ iff $p \in X$, by the definition of V_K.

Induction step:

$\alpha = \neg\beta$: Then $M_K, X \vDash \neg\beta$ iff $M_K, X \nvDash \beta$ iff (by the induction hypothesis) $\beta \notin X$ iff (by the fact that X is an MCS) $\neg\beta \in X$.

$\alpha = \beta \vee \gamma$: Then $M_K, X \vDash \beta \vee \gamma$ iff $M_K, X \vDash \beta$ or $M_K, X \vDash \gamma$ iff (by the induction hypothesis) $\beta \in X$ or $\gamma \in X$ iff (by the fact that X is an MCS) $\beta \vee \gamma \in X$.

$\alpha = \Box\beta$: We analyse this case in two parts:
($\Leftarrow$) Suppose that $\Box\beta \in X$. We have to show that $M_K, X \vDash \Box\beta$. Consider any MCS Y such that $X\ R_K\ Y$. Since $\Box\beta \in X$, from the definition of R_K it follows that $\beta \in Y$. By the induction hypothesis $M_K, Y \vDash \beta$. Since the choice of Y was arbitrary, $M_K, X \vDash \Box\beta$.
($\Rightarrow$) Suppose that $M_K, X \vDash \Box\beta$. We have to show that $\Box\beta \in X$. Suppose that $\Box\beta \notin X$. Then, since X is an MCS, $\neg\Box\beta \in X$. We show that this leads to a contradiction.

Claim $Y_0 = \{\gamma \mid \Box\gamma \in X\} \cup \{\neg\beta\}$ is consistent.

If we assume the claim, we can extend Y_0 to an MCS Y. Clearly, $X\ R_K\ Y$. Since $\neg\beta \in Y$, $\beta \notin Y$. By the induction hypothesis, $M_K, Y \nvDash \beta$. This means that $M_K, X \nvDash \Box\beta$ which contradicts our initial assumption that $M_K, X \vDash \Box\beta$.
To complete the proof, we must verify the claim.

Proof of claim Suppose that Y_0 is not consistent. Then, there exists $\{\gamma_1, \gamma_2, \ldots, \gamma_n\}$, a finite subset of Y_0, such that $\gamma_1 \wedge \gamma_2 \wedge \cdots \wedge \gamma_n \wedge \neg\beta$ is inconsistent. Let us denote $\gamma_1 \wedge \gamma_2 \wedge \cdots \wedge \gamma_n$ by $\tilde{\gamma}$.

We then have the following sequence of derivations:

$\vdash \neg(\tilde{\gamma} \wedge \neg\beta)$	By the definition of consistency
$\vdash \neg\tilde{\gamma} \vee \beta$	Tautology of propositional logic (Axiom Ao)
$\vdash \tilde{\gamma} \supset \beta$	Definition of $\supset$
$\vdash \Box(\tilde{\gamma} \supset \beta)$	Inference rule G
$\vdash \Box\tilde{\gamma} \supset \Box\beta$	Axiom K plus one application of MP
$\vdash \neg(\Box\tilde{\gamma} \wedge \neg\Box\beta)$	Tautology of propositional logic (Axiom Ao)

We can easily show that $\vdash \Box(\gamma \wedge \delta) \equiv (\Box\gamma \wedge \Box\delta)$.

In one direction, since $\vdash \gamma \wedge \delta \supset \gamma$ is a tautology of propositional logic, we can use the rule G to get $\vdash \Box(\gamma \wedge \delta \supset \gamma)$. From axiom K and one application of MP, $\vdash \Box(\gamma \wedge \delta) \supset \Box\gamma$. Symmetrically, it follows that $\vdash \Box(\gamma \wedge \delta) \supset \Box\delta$. So, $\vdash \Box(\gamma \wedge \delta) \supset (\Box\gamma \wedge \Box\delta)$.

Conversely, $\vdash \gamma \supset (\delta \supset (\gamma \wedge \delta))$ from propositional logic. By applying axiom K and MP a couple of times, we obtain $\vdash \Box\gamma \supset (\Box\delta \supset \Box(\gamma \wedge \delta))$, from which it follows that $\vdash (\Box\gamma \wedge \Box\delta) \supset \Box(\gamma \wedge \delta)$.

We can extend this argument to show that $\vdash \Box(\delta_1 \wedge \delta_2 \wedge \cdots \wedge \delta_n) \equiv (\Box\delta_1 \wedge \Box\delta_2 \wedge \cdots \Box\delta_n)$ for all n.

From the last line in our derivation above, it then follows that $\vdash \neg(\Box\gamma_1 \wedge \Box\gamma_2 \wedge \cdots \Box\gamma_n \wedge \neg\Box\beta)$. Thus the set $\{\Box\gamma_1, \Box\gamma_2, \ldots, \Box\gamma_n, \neg\Box\beta\}$ is inconsistent. But this is a finite subset of X, which means that X is itself inconsistent, contradicting the fact that X is an MCS.

From the preceding result, the proof of Lemma 2.12 is immediate.

Proof of Lemma 2.12: Let α be a formula which is consistent with respect to System K. By Lindenbaum's Lemma, α can be extended to a maximal consistent set X_α. By the preceding result $M, X_\alpha \vDash \alpha$, so α is $\mathscr{F}$-satisfiable. ⊣

Once we have proved Lemma 2.12, we immediately obtain a proof of completeness (Corollary 2.13) using exactly the same argument as in propositional logic.

It is worth pointing out one important difference between the canonical model constructed for System K and the models constructed when proving completeness for propositional logic. In propositional logic, to satisfy a consistent formula α, we build a valuation v which depends on α. On the other hand, the construction of the canonical model for System K is *independent* of the choice of α. Thus, *every* consistent formula α is satisfied within the model M_K.

Completeness for other classes of frames

Can we axiomatise the set of $\mathscr{C}$-valid formulas for a class of frames $\mathscr{C}$ which is properly included in $\mathscr{F}$? To do this, we use the characteristic formulas which we looked at when discussing correspondence theory.

Reflexive frames

System T is the set of axioms obtained by adding the following axiom scheme to System K.

(T) $\Box\alpha \supset \alpha$

Lemma 2.17 *System T is sound and complete with respect to the class of reflexive frames.*

Proof: To show that System T is sound with respect to reflexive frames, we only need to verify that the new axiom (T) is sound for this class of frames—the other axioms and rules from System K continue to be sound. The soundness of axiom (T) follows from Proposition 2.4.

To show completeness, we must argue that every formula which is consistent with respect to System T can be satisfied in a model based on a reflexive frame. To establish this, we follow the proof of completeness for System K and build a canonical model $M_T = ((W_T, R_T), V_T)$ for System T which satisfies the property described in Lemma 2.12. We just need to verify that the resulting frame (W_T, R_T) is reflexive.

For any MCS X, we need to verify that $X\ R_T\ X$ or, in other words, that $\{\alpha \mid \Box\alpha \in X\} \subseteq X$. Consider any formula $\Box\alpha \in X$. Since $\Box\alpha \supset \alpha$ is an axiom of System T, $\Box\alpha \supset \alpha \in X$, by Lemma 2.14 (i). From Lemma 2.14 (ii), it then follows that $\alpha \in X$, as required. ⊣

Transitive frames

System *4* is the set of axioms obtained by adding the following axiom scheme to System K.

(4) $\Box\alpha \supset \Box\Box\alpha$

Lemma 2.18 *System 4 is sound and complete with respect to the class of transitive frames.*

Proof: We know that the axiom (4) is sound for the class of transitive frames from Proposition 2.5. This establishes the soundness of System *4*.

To show completeness, we must argue that every formula which is consistent with respect to System *4* can be satisfied in a model based on a transitive frame. Once again, we can build a canonical model $M_4 = ((W_4, R_4), V_4)$ for System *4* which satisfies the property described in Lemma 2.12. We just need to verify that the resulting frame (W_4, R_4) is transitive.

In other words, if X, Y, Z are MCSs such that $X\ R_4\ Y$ and $Y\ R_4\ Z$, we need to verify that $X\ R_4\ Z$—that is, we must show that $\{\alpha \mid \Box\alpha \in X\} \subseteq Z$. Consider any formula $\Box\alpha \in X$. Since $\Box\alpha \supset \Box\Box\alpha$ is an axiom of System *4*, it follows from Lemma 2.14 that $\Box\Box\alpha \in X$. Since $X\ R_4\ Y$, it must be the case that $\Box\alpha \in Y$. Further, since $Y\ R_4\ Z$ it must be the case that $\alpha \in Z$, as required. ⊣

***Exercise** 2.19* The System B is obtained by adding the following axiom to System K.

(B) $\alpha \supset \Box\Diamond\alpha$.

Verify that System B is sound and complete with respect to symmetric frames. ⊣

Combinations of frame conditions

By combining the characteristic formulas for different frame conditions, we obtain completeness for smaller classes of frames.

Reflexive and transitive frames

The System *S4* is obtained by adding the axioms (T) (for reflexivity) and (4) (for transitivity) to System *K*.

Lemma 2.20 *System S4 is sound and complete with respect to the class of reflexive and transitive frames.*

Proof: Since System *T* is sound for the class of reflexive frames and System *4* is sound for the class of transitive frames, it follows that System *S4* is sound for the class of reflexive *and* transitive frames.

To show completeness, as usual we build a canonical model $M_{S4} = ((W_{S4}, R_{S4}), V_{S4})$ satisfying the property in Lemma 2.12. Using the argument in the proof of Lemma 2.17, it follows that R_{S4} is reflexive. Similarly, from the proof of Lemma 2.18 it follows that R_{S4} is transitive. ⊣

Equivalence relations

The System *S5* is obtained by adding the following axioms to System *K*.

(T) $\Box\alpha \supset \alpha$
(5) $\Diamond\alpha \supset \Box\Diamond\alpha$.

We have already seen that (T) is the axiom for reflexivity, while (5) characterises Euclidean frames.

Exercise *2.21*

(i) Show that System *S5* is sound and complete for the class of frames whose accessibility relation is an equivalence relation.

(ii) Show that the axioms (4) and (B) can be derived in System *S5*. ⊣

2.5 Bisimulations and expressiveness

Intuitively, it is clear that models which have "similar" structure satisfy the same modal logic formulas. For instance, if we choose the same valuation for all worlds in the two frames shown in Figure 2.5, it seems evident that no formula can distinguish the resulting pair of models.

To formalise this notion, we introduce bisimulations.

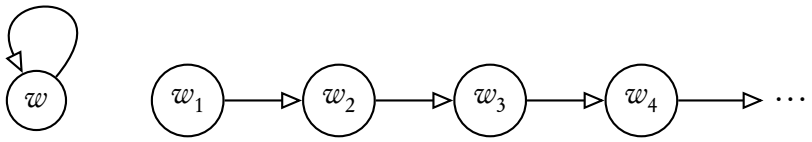


Figure 3: A pair of similar frames

Bisimulation Let $M_1 = ((W_1, R_1), V_1)$ and $M_2 = ((W_2, R_2), V_2)$ be a pair of models. A *bisimulation* is a relation $\sim \subseteq W_1 \times W_2$ satisfying the following conditions.

(i) If $w_1 \sim w_2$ and $w_1 R_1 w_1'$ then there exists w_2' such that $w_2 R_2 w_2'$ and $w_1' \sim w_2'$.

(ii) If $w_1 \sim w_2$ and $w_2 R_2 w_2'$ then there exists w_1' such that $w_1 R_1 w_1'$ and $w_1' \sim w_2'$.

(iii) If $w_1 \sim w_2$ then $V_1(w_1) = V_2(w_2)$.

Notice that the empty relation is a trivial example of a bisimulation. Two worlds which are related by a bisimulation satisfy exactly the same formulas.

Lemma 2.22 *Let $\sim$ be a bisimulation between $M_1 = ((W_1, R_1), V_1)$ and $M_2 = ((W_2, R_2), V_2)$. For all $w_1 \in W_1$ and $w_2 \in W_2$, if $w_1 \sim w_2$, then for all formulas α, $M_1, w_1 \models \alpha$ iff $M_2, w_2 \models \alpha$.*

Proof: As usual, the proof is by induction on the structure of α.

Basis: Suppose $\alpha = p \in \mathscr{P}$. By the definition of bisimulations, we know that $V_1(w_1) = V_2(w_2)$. Hence, $M_1, w_1 \models p$ iff $M_2, w_2 \models p$.

Induction step: The propositional cases $\alpha = \neg\beta$ and $\alpha = \beta \vee \gamma$ are easy, so we omit them and directly consider the case $\alpha = \Box\beta$.

($\Rightarrow$) Suppose that $M_1, w_1 \models \Box\beta$. We must show that $M_2, w_2 \models \Box\beta$ as well. For this, we must argue that $M_2, w_2' \models \beta$ for each world w_2' such that $w_2 R_2 w_2'$. Since $\sim$ is a bisimulation, for each such w_2' there exists a world w_1' such that $w_1 R_1 w_1'$ and $w_1' \sim w_2'$. Since $M_1, w_1 \models \Box\beta$, it follows that $M_1, w_1' \models \beta$. Since $w_1' \sim w_2'$, by the induction hypothesis, it follows that $M_2, w_2' \models \beta$. Since w_2' was an arbitrarily chosen R_2-neighbour of w_2, we have $M_2, w_2 \models \Box\beta$, as required.

($\Leftarrow$) Suppose that $M_2, w_2 \models \Box\beta$. We must show that $M_1, w_1 \models \Box\beta$ as well. The argument is symmetric to the earlier one and we omit the details. $\dashv$

We can use bisimulations to show that certain classes of frames *cannot* be characterised in modal logic.

Lemma 2.23 *The class of irreflexive frames cannot be characterised in modal logic.*

Proof: Let α be a formula that characterises the class of irreflexive frames. Consider the pair of frames in Figure 2.5. Since the first frame is not irreflexive, there should be a valuation V and an instance β of α such that β is not satisfied at w under V.

Let us define a valuation V' on the second model such that for each w_i, $V'(w_i) = V(w)$. We can clearly set up a bisimulation between the two models by relating w to each of the worlds w_i. This means that w satisfies exactly the same formulas as each of the worlds w_i. In particular, β is not satisfied at each w_i. This is a contradiction because the second model is irreflexive and β is an instance of the formula α which we claimed was a characteristic formula for irreflexive frames. ⊣

***Exercise** 2.24* We say that a frame (W, R) is "non-connected" if there are worlds w and w' such that it is not the case that $w(R \cup R^{-1})^* w'$. In other words, we convert (W, R) into an undirected graph by ignoring the orientation of edges in R. The frame is "non-connected" if there are two nodes in the resulting undirected graph which are not reachable from each other.

Show that there is no axiom which characterises the class of "non-connected" frames. ⊣

Antisymmetry

We have seen that irreflexivity cannot be characterised in modal logic. Another natural frame condition which is beyond the expressive power of modal logic is *antisymmetry*. Recall that a relation R on W is antisymmetric if $w\ R\ w'$ and $w'\ R\ w$ imply that $w = w'$.

Lemma 2.25 *Let α be a formula which is satisfiable over the class of reflexive and transitive frames. Then, α is satisfiable in a model based on an reflexive, transitive and antisymmetric frame.*

Proof: Let $M = ((W, R), V)$ be a model where R is reflexive and transitive. We describe a technique called *bulldozing*, due to Krister Segerberg, for constructing a new model $\widehat{M} = ((\widehat{W}, \widehat{R}), \widehat{V})$, where $\widehat{R}$ is reflexive, transitive and antisymmetric, such that $\widehat{M}$ and M satisfy the *same* formulas.

Consider the frame (W, R). If R is not antisymmetric, there are two worlds w and w' in W such that $w\ R\ w'$ and $w'\ R\ w$. The idea is to break each loop of this kind by making infinitely many copies of w and w' and arranging these copies alternately in a chain. We then verify that the new model which we construct is bisimilar to the original model.

Formally, we say that $X \subseteq W$ is a *cluster* if $X \times X \subseteq R$—in a cluster, every world can "see" every other world.

Let Cl be the class of *maximal* clusters of W—that is, $X \in Cl$ if X is cluster and for each $w \notin X$, $(X \cup \{w\}) \times (X \cup \{w\}) \not\subseteq R$. Since R is reflexive, every singleton $\{w\}$ is a cluster. It follows that the set Cl of maximal clusters is not empty and that every world $w \in W$ belongs to some maximal cluster in Cl. In fact, W is partitioned into maximal clusters.

For each $X \in Cl$, define $W_X = X \times \mathbb{N}$, where $\mathbb{N}$ is the set $\{0, 1, 2, \ldots\}$ of natural numbers. Thus W_X contains infinitely many copies of each world from X. For each set W_X, we define an accessibility

relation within W_X. For this, we first fix an arbitrary total order $\leq_X$ on X. For $X \in Cl$, $R_X \subseteq W_X \times W_X$ is then defined as follows:

$$\begin{aligned} R_X = \quad & \{((w,i),(w,i)) \mid w \in X \text{ and } i \in \mathbb{N}\} \\ \cup \quad & \{((w,i),(w',i)) \mid w, w' \in X \text{ and } w \leq_X w'\} \\ \cup \quad & \{((w,i),(w',j)) \mid w, w' \in X \text{ and } i < j\} \end{aligned}$$

We then define a relation across maximal clusters based on the original accessibility relation R:

$$R' = \bigcup \{(W_X \times W_Y) \mid X \neq Y \text{ and for some } w \in X \text{ and } w' \in Y,\ w\ R\ w'\}$$

Finally, we can define the new frame $(\widehat{W}, \widehat{R})$ corresponding to (W, R).

- $\widehat{W} = \bigcup_{X \in Cl} W_X$.
- $\widehat{R} = R' \cup \bigcup_{X \in Cl} R_X$.

It can be verified that $\widehat{R}$ is reflexive, transitive and antisymmetric (Exercise 2.26).

Each world in $\widehat{W}$ is of the form (w,i) where $w \in X$ for some maximal cluster $X \in Cl$ and $i \in \mathbb{N}$. We extend $(\widehat{W}, \widehat{R})$ to a model by defining $\widehat{V}((w,i)) = V(w)$ for all $w \in W$ and $i \in \mathbb{N}$.

We define a relation $\sim \subseteq \widehat{W} \times W$ as follows:

$$\sim = \{((w,i), w) \mid w \in W,\ i \in \mathbb{N}\}$$

We claim that $\sim$ is a bisimulation between $\widehat{M}$ and M. From the definition of $\widehat{V}$, we have $\widehat{V}((w,i)) = V(w)$ for all $w \in W$ and $i \in \mathbb{N}$, so the third condition in the definition of bisimulations is trivially satisfied.

Suppose that $(w,i) \sim w$ and $(w,i)\ \widehat{R}\ (w',j)$. We must show that $w\ R\ w'$. If w and w' belong to the same maximal cluster X, then $w\ R\ w'$ because all elements in X are R-neighbours of each other. On the other hand, if $w \in X$ and $w' \in Y$ for distinct clusters X and Y, it must be the case that $(w,i)\ R'\ (w',j)$. This means that we have $w_1 \in X$ and $w_1' \in Y$ such that $w_1\ R\ w_1'$. Since $w\ R\ w_1$ and $w_1'\ R\ w'$, from the transitivity of R it follows that $w\ R\ w'$.

Conversely, suppose that $(w,i) \sim w$ and $w\ R\ w'$. We exhibit a world (w',j) such that $(w,i)\ \widehat{R}\ (w',j)$. If w and w' belong to the same maximal cluster X, we just choose (w',j) such that $i < j$. Then, by the definition of R_X, $(w,i)\ R_X\ (w',j)$, so $(w,i)\ \widehat{R}\ (w',j)$ as well. On the other hand, if $w \in X$ and $w' \in Y$ for distinct maximal clusters X and Y, then $(x,i)\ R'\ (y,j)$ for all $j \in \mathbb{N}$, so once again we can pick a (w',j) such that $(x,i)\ \widehat{R}\ (y,j)$.

Thus, $\sim$ is a bisimulation between $\widehat{M}$, whose frame is antisymmetric and transitive, and M, whose frame is transitive. Hence, for any world $w \in W$ and any formula α, $M, w \vDash \alpha$ iff $\widehat{M}, (w,i) \vDash \alpha$ for all $i \in \mathbb{N}$. In other words, every formula which is satisfiable in the class of transitive frames is also satisfiable in the class of antisymmetric and transitive frames. ⊣

***Exercise** 2.26* Show that the relation $\widehat{R}$ constructed in the proof of Lemma 2.25 is reflexive, transitive and antisymmetric. ⊣

Corollary 2.27 *The class of antisymmetric frames cannot be characterised in modal logic.*

Proof: Let α be a formula characterizing the class of antisymmetric frames. Let (W,R) be a frame where R is reflexive and transitive but not antisymmetric. Then, there exists an instance β of α and a valuation V over (W,R) such that $M,w \models \neg\beta$ for some $w \in W$. By Lemma 2.25, we can convert M into a model $\widehat{M} = ((\widehat{W},\widehat{R}),\widehat{V})$ where $\widehat{R}$ is reflexive, transitive and antisymmetric, such that $M,\widehat{w} \models \neg\beta$ for some $\widehat{w} \in \widehat{W}$. This is a contradiction, since β was assumed to be an instance of the formula α which characterises antisymmetric frames. ⊣

We have already seen that the system *S4* is sound and complete for the class of reflexive, transitive frames. This class is very close to the class of partial orders, which are ubiquitous in computer science. The fact that antisymmetry cannot be characterised in modal logic means that modal logic cannot distinguish between reflexive and transitive frames (often called *preorders*) and reflexive, transitive and antisymmetric frames (or *partial orders*).

Corollary 2.28 *The system S4 is sound and complete for the class of partial orders.*

Proof: Since partial orders are reflexive and transitive, *S4* is certainly sound for this class of frames. We already know that every formula which is consistent with respect to *S4* is satisfiable in a preorder. The bulldozing construction described in the proof of Lemma 2.25 shows that every formula satisfiable over a preorder is also satisfiable over a partial order. ⊣

2.6 Decidability: Filtrations and the finite model property

Though we have looked at sound and complete axiomatisations of different classes of frames, we have yet to establish any results concerning decidability. The basic technique for showing decidability is to prove that any formula which is satisfiable is in fact satisfiable in a *finite* model.

Finite model property Let A be an axiom system which is sound and complete with respect to a class of frames $\mathscr{C}$. The system A has the *finite model property* if for all formulas $\alpha, \nvdash_A \alpha$ implies there is a model $M = (F,V)$ based on a finite frame $F = (W,R) \in \mathscr{C}$ such that for some $w \in W$, $M,w \models \neg\alpha$.

Since A is sound and complete for the class $\mathscr{C}$, this is equivalent to demanding that any formula which is satisfiable in the class $\mathscr{C}$ is in fact satisfiable in a model based on a finite frame from the class $\mathscr{C}$.

Assume that we can effectively decide whether or not a given finite frame belongs to the class $\mathscr{C}$, we can then systematically enumerate all finite models built from the class $\mathscr{C}$. As a consequence, the finite model property allows us to enumerate the set of formulas satisfiable within the class $\mathscr{C}$. On the other hand, the completeness of the axiom system A allows us to enumerate the set of formulas which are valid in this class of frames.

To check whether a formula α is valid, we interleave these enumerations. If α is valid, it will be enumerated as a thesis of the system A. On the other hand, if α is not valid, its negation $\neg\alpha$ must be satisfiable, so $\neg\alpha$ will appear in the enumeration of formulas satisfiable over $\mathscr{C}$. Thus, the finite model property yields a decision procedure for validity (and, dually, satisfiability).

Subformulas Let α be formula. The set of *subformulas of* α, denoted $sf(\alpha)$, is the smallest set of formulas such that:

- $\alpha \in sf(\alpha)$.
- If $\neg\beta \in sf(\alpha)$ then $\beta \in sf(\alpha)$.
- If $\beta \vee \gamma \in sf(\alpha)$ then $\beta \in sf(\alpha)$ and $\gamma \in sf(\alpha)$.
- If $\Box\beta \in sf(\alpha)$ then $\beta \in sf(\alpha)$.

***Exercise** 2.29* Show that the size of the set $sf(\alpha)$ is bounded by the length of α. More formally, for a formula α, define $|\alpha|$, the length of α, to be the number of symbols in α. Show that if $|\alpha| = n$ then $|sf(\alpha)| \leq n$. Give an example where $|sf(\alpha)| < |\alpha|$. ⊣

For a set X of formulas, we write $sf(X)$ to denote the set $\bigcup_{\alpha\in X} sf(\alpha)$. A set of formulas X is said to be *subformula-closed* (or just *sf-closed*) if $X = sf(X)$.

Let $M = ((W,R),V)$ and $M' = ((W',R'),V')$ be a pair of models. We have already seen that if we can set up a bisimulation $\sim$ between M and M', then for each pair of worlds $(w,w') \in \sim$, the worlds w and w' satisfy the same formulas. Often, we are willing to settle for a weaker relationship between w and w'—we do not require them to agree on *all* formulas, but only on formulas from a fixed set X. For sf-closed subsets X, this can be achieved using filtrations.

Filtrations Let $M = ((W,R),V)$ and $M' = ((W',R'),V')$ be a pair of models and X an sf-closed set of formulas. An *X-filtration* from M to M' is a function $f : W \to W'$ such that:

(i) For all $w, w' \in W$, if $w \; R \; w'$ then $f(w) \; R \; f(w')$.

(ii) The map f is surjective.

(iii) For all $p \in \mathscr{P} \cap X$, $p \in V(w)$ iff $p \in V'(f(w))$.

(iv) If $(f(w), f(w')) \in R'$, then for each formula of the form $\Box\alpha$ in X, if $M, w \vDash \Box\alpha$ then $M, w' \vDash \alpha$.

In a filtration, we have a weaker requirement on the inverse image of f than in a bisimulation. If $f(w)R'f(w')$, we do not demand that $w \; R \; w'$. We only insist that w and w' be "semantically" related upto the formulas in X. It is quite possible that $(w,w') \notin R$ and hence for some $\Box\beta \notin X$, $M, w \vDash \Box\beta$ while $M, w' \nvDash \beta$.

Lemma 2.30 *Let f be an X-filtration from $M = ((W,R),V)$ to $M' = ((W',R'),V')$ where X is an sf-closed set of formulas. Then, for all $\alpha \in X$ and for all $w \in W$, $M, w \models \alpha$ iff $M', f(w) \models \alpha$.*

Proof: The proof is by induction on the structure of α.
Basis If $\alpha = p \in \mathscr{P} \cap X$, then $M, w \models p$ iff $p \in V(w)$ iff (by the definition of X-filtrations) $p \in V'(f(w))$ iff $M, f(w) \models p$.

Induction step The propositional cases $\alpha = \neg\beta$ and $\alpha = \beta \vee \gamma$ are easy, so we omit them and directly consider the case $\alpha = \Box\beta$.

($\Rightarrow$) Suppose $M, w \models \Box\beta$. To show that $M', f(w) \models \Box\beta$, we must show that for each w' with $f(w)R'w'$, $M', w' \models \beta$. Fix an arbitrary w' such that $f(w)R'w'$. Since f is surjective, there is a world $w'' \in W$ such that $w' = f(w'')$. From the last clause in the definition of filtrations, it follows that $M, w'' \models \beta$. Since X is sf-closed, $\beta \in X$. From the induction hypothesis, we have $M', f(w'') \models \beta$ or, in other words, $M', w' \models \beta$. Since w' was an arbitrary R'-neighbour of $f(w)$, it follows that $M', f(w) \models \Box\beta$.

($\Leftarrow$) Suppose that $M', f(w) \models \Box\beta$. To show that $M, w \models \Box\beta$, we must show that for each w' with $w \; R \; w'$, $M, w' \models \beta$. Fix an arbitrary w' such that $w \; R \; w'$. From the first clause in the definition of filtrations, it follows that $f(w)R'f(w')$. Since $M', f(w) \models \Box\beta$, it must be the case that $M', f(w') \models \beta$. Since $\beta \in X$, from the induction hypothesis we have $M, w' \models \beta$. Since w' was an arbitrary R-neighbour of w, it follows that $M, w \models \Box\beta$. ⊣

Recall that our goal is to establish the finite model property for a class of frames $\mathscr{C}$—whenever a formula α is satisfiable over $\mathscr{C}$, then there is a model for α based on a *finite* frame from the class $\mathscr{C}$.

Our strategy will be as follows: given a formula α and an arbitrary model M for α, define an sf-closed set of formulas X_α and a finite model M_α such that $\alpha \in X_\alpha$ and there is an X_α-filtration from M to M_α. Lemma 2.30 then tells us that α is satisfied in M_α. Since this procedure applies uniformly to all satisfiable formulas α over the given class of frames, it follows that this class of frames has the finite model property.

Defining X_α is easy—we set $X_\alpha = sf(\alpha)$. To construct M_α, we have to define a frame (W_α, R_α) and a valuation $V_\alpha : W_\alpha \to 2^{\mathscr{P}}$.

We define W_α and V_α in a uniform manner for all classes of frames. To define W_α, we begin with the following equivalence relation $\simeq_\alpha$ on W: $w \simeq_\alpha w'$ if for each $\beta \in X_\alpha$, $M, w \models \beta$ iff $M, w' \models \beta$. In other words, $w \simeq_\alpha w'$ iff the worlds w and w' satisfy exactly the same formulas from the set X_α. We use $[w]$ represent the equivalence class of w with respect to the relation $\simeq$—that is, $[w] = \{w' \mid w' \simeq_\alpha w\}$.

Let $W_\alpha = \{[w] \mid w \in W\}$. Observe that W_α is finite whenever X_α is finite. Since $X_\alpha = sf(\alpha)$, we know that X_α is finite (recall Exercise 2.29).

Defining V_α is simple: for each $[w] \in W_\alpha$, $V_\alpha([w]) = \bigcap_{w' \in [w]} V(w')$.

Defining R_α is more tricky: in general, this relation has to be defined taking into account the class of frames under consideration. We now show how to define "suitable" R_α for some of the classes of frames for which we have already shown complete axiomatisations.

Lemma 2.31 *The axiom system K has the finite model property.*

Proof: Recall that system K is sound and complete for the class $\mathscr{F}$ of *all* frames. From our discussion of the finite model property, it suffices to show that any formula satisfiable over $\mathscr{F}$ is in fact satisfiable over a finite frame in $\mathscr{F}$.

Let α be a satisfiable formula and let $M = ((W,R),V)$ be a model for α—-that is, for some $w_\alpha \in W$, $M, w_\alpha \models \alpha$. Let $X_\alpha = sf(\alpha)$ and define W_α and V_α as described earlier. Define R_α as follows:

$$R_\alpha = \{([w],[w']) \mid \text{For each formula } \beta \in X_\alpha, \text{ if } M, w \models \Box\beta \text{ then } M, w' \models \beta\}$$

Let $M_\alpha = ((W_\alpha, R_\alpha), V_\alpha)$.

Fix the function $f : W \to W_\alpha$ such that $w \mapsto [w]$ for each $w \in W$. We claim that f is an X_α-filtration from M to M_α—for this, we have to verify that f satisfies properties (i)–(iv) in the definition of filtrations.

It is clear that f is surjective (property (ii)).

To verify property (iii) we have to show that for each $p \in \mathscr{P} \cap X_\alpha$ and for each $w \in W$, $p \in V(w)$ iff $p \in V_\alpha([w])$. Since the worlds in $[w]$ agree on all formulas in X_α, it follows that $p \in V(w)$ iff for each $w' \simeq_\alpha w$, $p \in V(w')$ iff $p \in \bigcap_{w' \in [w]} V(w')$ iff (by the definition of V_α) $p \in V_\alpha([w])$.

Property (i) demands that $(w,w') \in R$ implies $([w],[w']) \in R_\alpha$. By the definition of R_α, $([w],[w']) \in R_\alpha$ if for each $\beta \in X_\alpha$, whenever $M, w \models \Box\beta$, $M, w' \models \beta$ as well. This is immediate from the fact that $(w,w') \in R$.

Finally, property (iv) states that whenever $([w],[w']) \in R_\alpha$, for each formula $\Box\beta \in X_\alpha$, if $M, w \models \Box\beta$ then $M, w' \models \beta$. This follows directly from the definition of R_α.

Having established that f is an X_α-filtration from M to M_α, it follows that $M_\alpha, [w_\alpha] \models \alpha$. Thus M_α is a finite model for α, as required. ⊣

Lemma 2.32 *The axiom system T has the finite model property.*

Proof: Recall that system T is sound and complete for the class of reflexive frames. Let α be a formula satisfiable at a world w_α in a model $M = ((W,R),V)$ where (W,R) is a reflexive frame. We have to exhibit a finite model for α based on a reflexive frame.

Define X_α and $M_\alpha = ((W_\alpha, R_\alpha), V_\alpha)$ as in the proof of Lemma 2.31. We have already seen that $f : w \mapsto [w]$ then defines an X_α-filtration from M to M_α. To complete the proof of the present lemma, it suffices to show that the frame (W_α, R_α) is reflexive.

Since R is reflexive, we have $(w,w) \in R$ for each $w \in W$. By property (i) of filtrations, $(w,w) \in R$ implies $([w],[w]) \in R_\alpha$. Since f is surjective, it then follows that R_α is reflexive as well. (Notice that this argument actually establishes that *any* filtration from a reflexive model M to a model M' preserves reflexivity.) ⊣

Lemma 2.33 *The axiom system S4 has the finite model property.*

Proof: Recall that *S4* is sound and complete for the class of reflexive and transitive frames. Let α be a formula satisfiable at a world w_α in a model $M = ((W,R),V)$ where (W,R) is reflexive and transitive. We have to exhibit a finite model for α based on a reflexive and transitive frame.

Let $X_\alpha = sf(\alpha)$ and define W_α and V_α in terms of $\simeq_\alpha$ as usual. Let R_α be defined as follows:

$$R_\alpha = \{([w],[w']) \mid \text{For each formula } \Box\beta \in X_\alpha.\ \text{if } M,w \vDash \Box\beta \text{ then } M,w' \vDash \Box\beta.\}$$

Let $M_\alpha = ((W_\alpha, R_\alpha), V_\alpha)$.

As usual, we define $f : W \to W_\alpha$ by $w \mapsto [w]$. We have already seen that such a function satisfies properties (ii) and (iii) in the definition of a filtration.

We have to verify that f satisfies properties (i) and (iv) with the new definition of R_α. To show property (i), we have to verify that if $(w,w') \in R$ then $([w],[w']) \in R_\alpha$. Suppose that $M,w \vDash \Box\beta$. Since (W,R) is transitive, $M,w \vDash \Box\beta \supset \Box\Box\beta$, so $M,w \vDash \Box\Box\beta$ as well. Since (w,w'), $M,w' \vDash \Box\beta$. Thus $([w],[w']) \in R_\alpha$.

For property (iv), we have to show that if $([w],[w']) \in R_\alpha$ then for each formula of the form $\Box\beta$ in X_α, if $M,w \vDash \Box\beta$, then $M,w' \vDash \beta$. From the definition of R_α, we know that if $M,w \vDash \Box\beta$, then $M,w' \vDash \Box\beta$ as well. Since (W,R) is reflexive, $M,w' \vDash \Box\beta \supset \beta$, so $M,w' \vDash \beta$ as required.

Having established that f is an X_α-filtration from M to M_α, it remains to prove that the frame (W_α, R_α) is reflexive and transitive. Recall that (W,R) is assumed to be a reflexive and transitive frame. We have already remarked in the proof of the previous lemma that any filtration from a reflexive model preserves reflexivity, so it is immediate that (W_α, R_α) is a reflexive frame.

To show transitivity, suppose that $([w_1],[w_2])$ and $([w_2],[w_3])$ belong to R_α. We have to show that $([w_1],[w_3]) \in R_\alpha$ as well. This means that for each formula $\Box\beta$ in X_α, we have to show that if $M,w_1 \vDash \Box\beta$ then $M,w_3 \vDash \Box\beta$. Suppose that $M,w_1 \vDash \Box\beta$. Since $([w_1],[w_2]) \in R_\alpha$, we know that $M,w_2 \vDash \Box\beta$. Now, since $([w_2],[w_3]) \in R_\alpha$, it follows that $M,w_3 \vDash \Box\beta$ as well. ⊣

***Exercise** 2.34*

(i) Recall that the axiom system B is sound and complete for the class of symmetric frames. Show that B has the finite model property. Define R_α as follows:

$$R_\alpha = \{([w],[w']) \mid \text{For each formula } \Box\beta \in X_\alpha,\quad \begin{array}{ll}(\text{i}) & \text{if } M,w \vDash \Box\beta \text{ then } M,w' \vDash \beta \\ (\text{ii}) & \text{if } M,w' \vDash \Box\beta \text{ then } M,w \vDash \beta\}\end{array}$$

(ii) Recall that the axiom system $S5$ is sound and complete for the class of frames based on equivalence relations. Show that $S5$ has the finite model property. Define R_α as follows:

$$R_\alpha = \{([w],[w']) \mid \text{For each formula } \Box\beta \in X_\alpha,\ M,w \vDash \Box\beta \text{ iff } M,w' \vDash \Box\beta\}$$

⊣

Small model property In all the finite models we have constructed, we have defined W_α to be the set of equivalence classes generated by the relation $\simeq_\alpha$. Since the size of $sf(\alpha)$ is bounded by $|\alpha|$, it follows that $|W_\alpha|$ is bounded by $2^{|\alpha|}$. Thus, when we establish the finite model property using the equivalence relation $\simeq_\alpha$, we in fact derive a bound on the size of a finite model for α. As a result, we establish a stronger property, which we call the *small model property*.

More formally, we say that a class of frames $\mathscr{C}$ has the small model property if there is a function $f_{\mathscr{C}} : \mathbb{N} \to \mathbb{N}$ such that for each formula α satisfiable over the class $\mathscr{C}$, there is a model for α over $\mathscr{C}$ whose size is bounded by $f_{\mathscr{C}}(|\alpha|)$. For instance, in the examples we have seen, $f_{\mathscr{C}}(|\alpha|) = 2^{|\alpha|}$.

The small model property gives us a more direct decidability argument—to check if α is satisfiable, we just have to enumerate all models of size less than $f_{\mathscr{C}}(|\alpha|)$. To show that this is possible, we first observe that the number of frames in this subclass is bounded. To bound the number of models based on this finite set of frames, notice that it suffices to consider valuations restricted to the finite set of atomic propositions which occur in α. Thus given a finite frame, there are only finitely many different valuations possible over that frame.

This decision procedure has the advantage of giving us a bound on the complexity of the decision problem. This bound is just the bound on the number of different models which can be generated whose size is less than $f_{\mathscr{C}}(|\alpha|)$.

***Exercise** 2.35* In the examples we have seen (axiom systems *K*, *T* etc.) verify that the satisfiability of a formula α can be checked in time which is doubly exponential in $|\alpha|$. ⊣

2.7 *Labelled transition systems and multi-modal logic*

Transition systems A *transition system* is a pair $(S, \to)$ where S is a set of *states* and $\to \subseteq S \times S$ is a *transition relation*. Transition systems are a general framework to describe computing systems. States describe configurations of the system—for instance, the contents of the disk, memory and registers of a computer at a particular instant. The transition relation then describes when one configuration can follow another—for instance the effect of executing a machine instruction which affects some of the memory, register or disk locations and leaves the rest of the configuration untouched.

It is clear that a transition system has exactly the same structure as a frame (W, R) in modal logic. Hence, we can use modal logic to describe properties of transition systems. This is one of the main reasons why modal logic is interesting to computer scientists.

Often, we are interested in a more structured representation of the configuration space of a computing system—in particular, we not only want to record that a transition is possible from a configuration s to a configuration s' but we also want to keep track of the "instruction" which caused this change of configuration. This leads us to the notion of labelled transition systems.

Labelled transition systems A *labelled transition system* is a triple $(S, \Sigma, \to)$ where S is a set of states, Σ is a set of actions and $\to \subseteq S \times \Sigma \times S$ is a *labelled transition relation*.

The underlying structure in a finite automaton is a familiar example of a labelled transition system, where the set of states is finite.

How can we reason about labelled transition systems in the framework of modal logic? One option is to ignore the labels and consider the derived transition relation $\Rightarrow = \{(s,s') \mid \exists a \in \Sigma : (s,a,s') \in \rightarrow\}$. We can then reason about the frame $(S, \Rightarrow)$ using the modalities $\Box$ and $\Diamond$. This approach is clearly not satisfactory because we have lost all information about the labels of actions within our logic. A more faithful translation involves the use of multi-modal logics.

Multi-modal logics A *multi-relational frame* is a structure $(W, R_1, R_2, \ldots, R_n)$ where R_i is a binary relation on W for each $i \in \{1,2,\ldots,n\}$. A multi-relational frame can be viewed as the superposition of n normal frames (W,R_1), (W,R_2), $\ldots$, (W,R_n), all defined with respect to the same set of worlds.

To reason about a multi-relational frame, we define a *multi-modal logic* whose syntax consists of a set $\mathscr{P}$ of atomic propositions, the boolean connectives $\neg$ and $\vee$ and a set of n modalities $\Box_1$, $\Box_2$, $\ldots$, $\Box_n$.

To define the semantics of multi-modal logic, we first fix a valuation $V : W \rightarrow 2^{\mathscr{P}}$ as before. We then define the satisfaction relation $M, w \models \alpha$. The propositional cases are the same as for standard modal logic. The only difference is in the semantics of the modalities. For each $i \in \{1,2,\ldots,n\}$, we define

$$M, w \models \Box_i \alpha \text{ iff for each } w' \in W, \text{ if } w \; R_i \; w' \text{ then } M, w' \models \alpha$$

Thus, the modalities $\{\Box_i\}_{i \in \{1,2,\ldots,n\}}$ are used to "independently" reason about the relations $\{R_i\}_{i \in \{1,2,\ldots,n\}}$. We can then use the theory we have developed to describe properties of each of these relations. For instance, the multi-relational frames where the axioms $\Box_3 \alpha \supset \alpha$ and $\Box_7 \alpha \supset \Box_7 \Box_7 \alpha$ are valid correspond to the class where R_3 is reflexive and R_7 is transitive. We can express interdependencies between different relations using formulas which combine these modalities. For instance, the formula $\alpha \supset \Diamond_5 \Diamond_2 \beta$ indicates that a world which satisfies α has an R_5-neighbour which in turn has an R_2-neighbour where β holds.

We have seen how to characterise classes of frames using formulas from modal logic. We can extend this idea in a natural way to characterise classes of multi-relational frames.

***Exercise** 2.36* Consider the class of multi-relational frames (W, R_1, R_2) where $R_2 = R_1^{-1}$. Describe axioms to characterise this class. (*Hint:* The combined relation $R_1 \cup R_2$ is a symmetric relation on W. Work with suitable modifications of axiom (B). You may use more than one axiom.) $\dashv$

To reason about labelled transition systems in this framework, we have to massage the structure $(S, \Sigma, \rightarrow)$ into a multi-relational frame. To achieve this, we define a relation $\rightarrow_a \; \subseteq S \times S$ for each $a \in \Sigma$ as follows:

$$\rightarrow_a = \{(s,s') \mid (s,a,s') \in \rightarrow\}$$

It is then clear that the multi-relational frame $(S, \{\rightarrow_a\}_{a \in \Sigma})$ describes the same structure as the original labelled transition system $(S, \Sigma, \rightarrow)$.

To reason about the structure $(S, \{\rightarrow_a\}_{a\in\Sigma})$, we have modalities $\Box_a$ (read as *Box a*) and $\Diamond_a$ (read as *Diamond a*) for each $a \in \Sigma$. Traditionally, the modality $\Box_a$ is written $[a]$ and the modality $\Diamond_a$ is written $\langle a \rangle$.

When reasoning about labelled transition systems, the set of atomic propositions $\mathscr{P}$ corresponds to properties which distinguish one configuration of the system from each other. For instance, we could have an atomic proposition to denote that "memory location 27 is unused" or that "the printer is busy". In these notes, we will not go into the details of how to model a computing system in terms of such a logic.

Assuming we have an abstract encoding of system properties in terms of atomic propositions, we can now reason about the dynamic behavior of the system. For instance, we can assert $M, s \vDash [c]\langle b \rangle \alpha$ to denote that in the state s, any c-transition will lead to a state from where we can use a b-transition to realise the property described by α. In particular, if α is just the constant $\top$, this formula asserts that a b-transition is enabled after any c-transition.

Unfortunately, we still do not have the expressive power we need to make non-trivial statements about programs. For instance, we cannot say that after a c-transition, we can *eventually* reach a state where a b-transition is enabled. Or that we have reached a portion of the state space where *henceforth* only a and d transitions are possible.

For this, we need to move from modal logic to dynamic logic, which is the topic of discussion in the next section.

3 Dynamic Logic

Dynamic logic is a multi-modal logic where the modalities are indexed not by uninterpreted letters, but by *programs*, which have structure. The relationship between different programs also forms an integral part of the logic.

3.1 Syntax

As in propositional logic, we begin with a countably infinite set of atomic propositions $\mathscr{P} = \{p_0, p_1, \ldots\}$ and two logical connectives $\neg$ (read as *not*) and $\vee$ (read as *or*). We also begin with a countably infinite set of *atomic actions* $\mathscr{A} = \{a_0, a_1, \ldots\}$.

The set Φ of formulas of dynamic logic and the set Π of programs are simultaneously defined by induction as the smallest sets satisfying the following:

- Every atomic proposition p is a member of Φ.
- If α is a member of Φ, so is $(\neg\alpha)$.
- If α and β are members of Φ, so is $(\alpha \vee \beta)$.
- If α is a member of Φ and π is a member of Π, then $([\pi]\alpha)$ is a member of Φ.
- Every atomic action a is a member of Π.
- If π_1 and π_2 are members of Π, so are $(\pi_1 + \pi_2)$ and $(\pi_1 \cdot \pi_2)$.
- If π is a member of Π, so is (π^*).
- If α is a member of Φ, $(\alpha?)$ is a member of Π.

As before, we omit parentheses if there is no ambiguity. The derived propositional connectives $\wedge$, $\supset$ and $\equiv$ are defined as before. In addition, we have a derived modality $\langle\pi\rangle$ which is *dual* to the modality $[\pi]$, defined as follows: $\langle\pi\rangle\alpha \stackrel{\text{def}}{=} \neg[\pi]\neg\alpha$.

Informally, $[\pi]\alpha$ is true in a world w iff all worlds w' which one ends up in after executing program π in w satisfies α. The programs $\pi_1 + \pi_2$, $\pi_1 \cdot \pi_2$, and π^* denote nondeterministic choice between π_1 and π_2, sequential composition of π_1 and π_2, and arbitrary iteration of π, respectively. The program $\alpha?$ executed at world w is just a *skip* if α is true at w and an *abort* otherwise.

3.2 Semantics

Frames A *frame* is just a labelled transition system $F = (W, \mathscr{A}, \rightarrow)$. For each a in $\mathscr{A}$, define $\xrightarrow{a} \subseteq W \times W$ to be the set of pairs (w, w') such that (w, a, w') belongs to $\rightarrow$. If $w \xrightarrow{a} w'$ we say that w' is an a-neighbour of w.

Models A model is a pair $M = (F, V)$ where $F = (W, \mathscr{A}, \rightarrow)$ is a frame and $V : W \rightarrow 2^{\mathscr{P}}$ is a *valuation*.

Satisfaction The notion of truth is localised to each world in a model. We write $M, w \models \alpha$ to denote that α is true at the world w in the model M. The satisfaction relation and the relations $\xrightarrow{\pi}$ for each π in Π are defined by simultaneous induction as follows. We say that w' is a π-neighbour of w if $w \xrightarrow{\pi} w'$.

$$
\begin{array}{lll}
M, w \models p & \text{iff} & p \in V(w) \text{ for } p \in \mathscr{P} \\
M, w \models \neg\alpha & \text{iff} & M, w \not\models \alpha \\
M, w \models \alpha \vee \beta & \text{iff} & M, w \models \alpha \text{ or } M, w \models \beta \\
M, w \models [\pi]\alpha & \text{iff} & \text{for each } w' \in W, \text{ if } w \xrightarrow{\pi} w' \text{ then } M, w' \models \alpha \\
w \xrightarrow{\pi_1 + \pi_2} w' & \text{iff} & w \xrightarrow{\pi_1} w' \text{ or } w \xrightarrow{\pi_2} w' \\
w \xrightarrow{\pi_1 \cdot \pi_2} & \text{iff} & \text{for some } w'' \in W, w \xrightarrow{\pi_1} w'' \text{ and } w'' \xrightarrow{\pi_2} w' \\
w \xrightarrow{\pi^*} w' & \text{iff} & w \xrightarrow{\pi}{}^{*} w', \text{ where } R^* \text{ denotes the reflexive transitive closure of } R \\
w \xrightarrow{\alpha?} w' & \text{iff} & w = w' \text{ and } M, w \models \alpha
\end{array}
$$

Thus, $M, w \models [\pi]\alpha$ if every π-neighbour of w satisfies α. Notice that if w is π-isolated—that is, there is no world w' such that $w \xrightarrow{\pi} w'$—then $M, w \models [\pi]\alpha$ for *every* formula α. We say that a sequence of worlds $w_0, w_1, \ldots, w_n$ $(n \geq 0)$ is a π-path if $w_i \xrightarrow{\pi} w_{i+1}$ for all i such that $0 \leq i < n$. Such a path is said to be of length n. It is said to be from w to w' if $w_0 = w$ and $w_n = w'$. w' is said to be π-reachable from w if there is a π-path from w to w'. Notice that $w \xrightarrow{\pi^*} w'$ iff w' is π-reachable from w. Thus $M, w \models [\pi^*]\alpha$ iff every π-reachable world w' of w satisfies α.

Satisfiability and validity As usual, we say that α is *satisfiable* if there exists a frame $F = (W, \mathscr{A}, \rightarrow)$ and a model $M = (F, V)$ such that $M, w \models \alpha$ for some $w \in W$. The formula α is *valid*, written $\models \alpha$, if for every frame $F = (W, \mathscr{A}, \rightarrow)$, for every model $M = (F, V)$ and for every $w \in W$, $M, w \models \alpha$.

***Example** 3.1* Here are some examples of valid formulas in dynamic logic.

(i) Every substitution instance of a tautology of propositional logic is valid. The details are trivial.

(ii) The formula $[\pi](\alpha \supset \beta) \supset ([\pi]\alpha \supset [\pi]\beta)$ is valid. Consider a model $M = ((W, \mathscr{A}, \rightarrow), V)$ and a world $w \in W$. Suppose that $M, w \models [\pi](\alpha \supset \beta)$. We must argue that $M, w \models [\pi]\alpha \supset [\pi]\beta$. Let $M, w \models [\pi]\alpha$. Then we must show that $M, w \models [\pi]\beta$. In other words, we must show that every π-neighbour w' of w satisfies β. Since we assumed $M, w \models [\pi](\alpha \supset \beta)$, we know that $M, w' \models \alpha \supset \beta$. Moreover, since $M, w \models [\pi]\alpha$, $M, w' \models \alpha$. By the semantics of the connective $\supset$, it follows that $M, w' \models \beta$, as required.

(iii) The formula $[\pi_1 + \pi_2]\alpha \equiv ([\pi_1]\alpha \wedge [\pi_2]\alpha)$ is valid. Consider a model $M = ((W, \mathscr{A}, \rightarrow), V)$ and a world $w \in W$. Now $M, w \models [\pi_1 + \pi_2]\alpha$ iff (by semantics) $M, w' \models \alpha$ for all $\pi_1 + \pi_2$-neighbours w' of w iff (by definition of $\xrightarrow{\pi_1 + \pi_2}$) $M, w' \models \alpha$ for all w' that are either π_1-neighours or π_2-neighbours of w iff (by semantics) $M, w \models ([\pi_1]\alpha \wedge [\pi_2]\alpha)$.

(iv) The formula $[\pi_1 \cdot \pi_2]\alpha \equiv [\pi_1][\pi_2]\alpha$ is valid. Consider a model $M = ((W, \mathscr{A}, \rightarrow), V)$ and a world $w \in W$. Now $M, w \vDash [\pi_1 \cdot \pi_2]\alpha$ iff (by semantics) $M, w' \vDash \alpha$ for all $\pi_1 \cdot \pi_2$-neighbours w' of w iff (by definition of $\xrightarrow{\pi_1 \cdot \pi_2}$) $M, w' \vDash \alpha$ for all w' that are π_2-neighours of some π_1-neighbour w'' of w iff (by semantics) $M, w'' \vDash [\pi_2]\alpha$ for all π_1-neighbours w'' of w iff (by semantics) $M, w \vDash [\pi_1][\pi_2]\alpha$.

(v) The formula $[\pi^*]\alpha \equiv \alpha \wedge [\pi][\pi^*]\alpha$ is valid. Consider a model $M = ((W, \mathscr{A}, \rightarrow), V)$ and a world $w \in W$. Now $M, w \vDash [\pi^*]\alpha$ iff (by semantics) every world w' π-reachable from w satisfies α iff (by definition of π-reachability) w satisfies α and for all π-neighbours w'' of w, all worlds w' π-reachable from w'' satisfy α iff (by semantics) w satisfies α and every π-neighbour w'' of w satisfies $[\pi^*]\alpha$ iff (by semantics, again) $M, w \vDash \alpha \wedge [\pi][\pi^*]\alpha$.

(vi) The formula $(\alpha \wedge [\pi^*](\alpha \supset [\pi]\alpha)) \supset [\pi^*]\alpha$ is valid. Consider a model $M = ((W, \mathscr{A}, \rightarrow), V)$ and a world $w \in W$. Suppose $M, w \vDash \alpha$ and $M, w \vDash [\pi^*](\alpha \supset [\pi]\alpha)$. For any world w' of W that is π-reachable from w, define the π-height of w' (with respect to w) as the length of the shortest π-path from w to w'. We prove by induction on the π-height that every world w' π-reachable from w satisfies α, thereby showing that $M, w \vDash [\pi^*]\alpha$. Consider any world w' whose π-height is zero. It follows that $w' = w$ and therefore $M, w' \vDash \alpha$. Consider any world w' whose π-height is a non-zero number n. Clearly, there is a world w'' with π-height $n - 1$ such that $w'' \xrightarrow{\pi} w'$. Now, by induction hypothesis, $M, w'' \vDash \alpha$. But since $M, w \vDash [\pi^*](\alpha \supset [\pi]\alpha)$, it follows that $M, w'' \vDash \alpha \supset [\pi]\alpha$. Therefore $M, w'' \vDash [\pi]\alpha$, and hence $M, w' \vDash \alpha$.

(vii) The formula $[\alpha?]\beta \equiv (\alpha \supset \beta)$ is valid. Consider a model $M = ((W, \mathscr{A}, \rightarrow), V)$ and a world $w \in W$. Now $M, w \vDash [\alpha?]\beta$ iff $M, w' \vDash \beta$ for all $\alpha?$-neighbours w' of w iff (since $w \xrightarrow{\alpha?} w'$ iff $w = w'$ and $M, w \vDash \alpha$) whenever w satisfies α it also satisfies β iff (by semantics) $M, w \vDash \alpha \supset \beta$.

(viii) Suppose that α is valid. Then, $[\pi]\alpha$ must also be valid. Consider any model $M = ((W, \mathscr{A}, \rightarrow), V)$ and any $w \in W$. To check that $M, w \vDash [\pi]\alpha$ we have to verify that every π-neighbour of w satisfies α. Since α is valid, $M, w' \vDash \alpha$ for all $w' \in W$. So, every π-neighbour of w does satisfy α and $M, w \vDash [\pi]\alpha$.

3.3 Axiomatising valid formulas

Consider the following axiom system.

Axioms

(A0) All tautologies of propositional logic.
(A1) $[\pi](\alpha \supset \beta) \supset ([\pi]\alpha \supset [\pi]\beta)$.
(A2) $[\pi_1 + \pi_2]\alpha \equiv ([\pi_1]\alpha \wedge [\pi_2]\alpha)$.
(A3) $[\pi_1 \cdot \pi_2]\alpha \equiv [\pi_1][\pi_2]\alpha$.
(A4) $[\pi^*]\alpha \equiv (\alpha \wedge [\pi][\pi^*]\alpha)$.
(A5) $(\alpha \wedge [\pi^*](\alpha \supset [\pi]\alpha)) \supset [\pi^*]\alpha$.
(A6) $[\alpha?]\beta \equiv (\alpha \supset \beta)$.

(MP) $\dfrac{\alpha, \alpha \supset \beta}{\beta}$ (G) $\dfrac{\alpha}{[\pi]\alpha}$

As usual, we say that α is a *thesis*, denoted $\vdash \alpha$, if we can derive α using the axioms (A0) to (A6) and the inference rules (MP) and (G). It is easily seen that $\vdash [\pi](\alpha \wedge \beta) \equiv ([\pi]\alpha \wedge [\pi]\beta)$.

The result we want to establish is the following.

Theorem 3.2 *For all formulas α, $\vdash \alpha$ iff $\models \alpha$.*

As usual, one direction of the proof is easy.

Lemma 3.3 (Soundness) *If $\vdash \alpha$ then $\models \alpha$.*

Proof: As we observed earlier, it suffices to show that each axiom is valid and that the inference rules preserve validity. This is precisely what we exhibited in Example 3.1. ⊣

As in Propositional Logic and Modal Logic, we use a Henkin-style argument to show that every valid formula is derivable in our axiom system, but we do not construct a canonical model. It is technically much simpler to directly construct a finite model for each consistent formula.

Consistency We say that a formula α is *consistent* if $\not\vdash \neg\alpha$. A finite set of formulas $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is consistent if the conjunction $\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_n$ is consistent. Finally, an arbitrary set of formulas X is consistent if every finite subset of X is consistent.

Our goal is to prove the following.

Lemma 3.4 *Every consistent formula is satisfiable.*

As we saw in the case of Propositional Logic, this will yield as an immediate corollary the result we seek:

Corollary 3.5 (Completeness for dynamic logic) *Let α be a valid formula. Then $\vdash \alpha$.*

Atoms

Instead of working with maximal consistent sets as we did in modal logic, we work with certain subsets of subformulas of the formula of interest. We first make precise the notion of subformula of a formula. The definition is not completely obvious – it has some aspects which are motivated by the proof of completeness. *For convenience, in the rest of the section, we will fix a consistent formula α_0 and try to construct a model in which it is satisfied.*

Subformulas Let α be formula. The set of *subformulas of* α, denoted $sf(\alpha)$, is the smallest set of formulas such that:

- $\alpha \in sf(\alpha)$.
- If $\neg\beta \in sf(\alpha)$ then $\beta \in sf(\alpha)$.
- If $\beta \vee \gamma \in sf(\alpha)$ then $\beta \in sf(\alpha)$ and $\gamma \in sf(\alpha)$.
- If $[a]\beta \in sf(\alpha)$ (for $a \in \mathscr{A}$) then $\beta \in sf(\alpha)$.
- If $[\pi_1 + \pi_2]\beta \in sf(\alpha)$ then $[\pi_1]\beta \in sf(\alpha)$ and $[\pi_2]\beta \in sf(\alpha)$.
- If $[\pi_1 \cdot \pi_2]\beta \in sf(\alpha)$ then $[\pi_1][\pi_2]\beta \in sf(\alpha)$.
- If $[\pi^*]\beta \in sf(\alpha)$ then $[\pi][\pi^*]\beta \in sf(\alpha)$ and $\beta \in sf(\alpha)$.
- If $[\beta?]\gamma \in sf(\alpha)$ then $\beta \in sf(\alpha)$ and $\gamma \in sf(\alpha)$.

***Exercise** 3.6* Show that the size of the set $sf(\alpha)$ is bounded by the square of the length of α. More formally, for a formula α, define $|\alpha|$, the length of α, to be the number of symbols in α. Show that if $|\alpha| = n$ then $|sf(\alpha)| \leq n^2$. Give an example where $|sf(\alpha)| < |\alpha|^2$. ⊣

It is convenient in what follows to work with negation-closed sets of formulas. For any formula α, we define $\overline{\alpha}$ to be β if α is of the form $\neg\beta$, and $\neg\alpha$ otherwise. We define the *closure* of a formula α, denoted $cl(\alpha)$, to be the set $\{\beta, \overline{\beta} \mid \beta \in sf(\alpha)\}$. Note that the size of $cl(\alpha)$ is at most twice that of $sf(\alpha)$. In what follows, we will freely use the fact that $\vdash \neg\alpha \equiv \overline{\alpha}$, and loosely talk of $\neg\alpha$ belonging to a particular set when we actually mean that $\overline{\alpha}$ belongs to that set. For the rest of the section, we fix cl to be $cl(\alpha_0)$.

An *atom* is a maximal consistent subset of cl—it is a consistent subset A of cl such that for all $\alpha \notin cl$, $A \cup \{\alpha\}$ is not consistent. It can be easily seen that the atoms are exactly sets of the form $X \cap cl$ for some MCS X. The set of all atoms is denoted by AT.

As we saw earlier, by Lindenbaum's Lemma, every consistent set of formulas can be extended to an MCS. In particular, there is an MCS X containing α_0, and hence (by the observation in the previous paragraph), an atom A_0 containing α_0.

We will use the following properties of atoms.

Lemma 3.7 *Let A be an atom. Then:*

(i) For all formulas $\neg\alpha \in cl$, $\alpha \notin A$ iff $\neg\alpha \in A$.

(ii) For all formulas $\alpha \vee \beta \in cl$, $\alpha \vee \beta \in A$ iff $\alpha \in A$ or $\beta \in A$.

(iii) If $\alpha \in cl$ is a thesis, then $\alpha \in A$.

(iv) If A is an atom and if $A \cup \{\alpha\}$ is consistent for $\alpha \in cl$, then $\alpha \in A$.

(v) *If* $\vdash (\alpha_1 \wedge \cdots \wedge \alpha_n) \supset \beta$, $\alpha_i \in A$ *for each* $i \leq n$ *and* $\beta \in cl$, *then* $\beta \in A$.

(vi) *For all formulas* $[\pi_1 + \pi_2]\alpha \in cl$, $[\pi_1 + \pi_2]\alpha \in A$ *iff* $[\pi_1]\alpha \in A$ *and* $[\pi_2]\alpha \in A$.

(vii) *For all formulas* $[\pi_1 \cdot \pi_2]\alpha \in cl$, $[\pi_1 \cdot \pi_2]\alpha \in A$ *iff* $[\pi_1][\pi_2]\alpha \in A$.

(viii) *For all formulas* $[\pi^*]\alpha \in cl$, $[\pi^*]\alpha \in A$ *iff* $\alpha \in A$ *and* $[\pi][\pi^*]\alpha \in A$.

(ix) *For all formulas* $[\alpha?]\beta \in cl$, $[\alpha?]\beta \in A$ *iff* $\alpha \notin A$ *or* $\beta \in A$.

Proof: The proof is routine and is left as an exercise. ⊣

For any finite set of formulas $A = \{\alpha_1, \ldots, \alpha_n\}$, define $\widehat{A}$ to be $\alpha_1 \wedge \cdots \wedge \alpha_n$, and for any finite collection $V = \{A_1, \ldots, A_m\}$ of finite sets of formulas, define $\widehat{V}$ to be $\widehat{A_1} \vee \cdots \vee \widehat{A_m}$. We first present the following useful properties related to $\widehat{AT}$.

Lemma 3.8 $\vdash \widehat{AT}$.

Proof: Let $AT = \{A_1, \ldots, A_r\}$. Suppose it is not the case that $\vdash \widehat{AT}$. Then $\neg\widehat{AT}$ is consistent. In other words, $\neg\widehat{A_1} \wedge \cdots \wedge \neg\widehat{A_r}$ is consistent. By Lindenbaum's lemma, there is a maximal consistent set X such that $\neg\widehat{A_1} \wedge \cdots \wedge \neg\widehat{A_r} \in X$. This means that for all $i : 1 \leq i \leq r$, $\neg\widehat{A_i} \in X$. Let $B = X \cap cl$. Since X is a maximal consistent set, B is an atom, i.e. $B \in AT$. Let it be A_k for some $k : 1 \leq k \leq r$. But then $\widehat{A_k} \in X$, contradicting the consistency of X. Therefore it cannot be the case that $\neg\widehat{AT}$ is consistent, and so $\vdash \widehat{AT}$. ⊣

Lemma 3.9 *Let* $U \subseteq AT$ *and let* $V = AT \setminus U$. *Then* $\vdash \widehat{U} \equiv \neg\widehat{V}$.

Proof: Let $U = \{A_1, \ldots, A_m\}$ and $V = \{B_1, \ldots, B_n\}$. Further for all $i : 1 \leq i \leq m$ and $j : 1 \leq j \leq n$, let $\alpha_{ij} \in A_i \setminus B_j$. The following derivation shows that if $\vdash \widehat{U} \vee \widehat{V}$ then $\vdash \widehat{U} \equiv \neg\widehat{V}$. (The line number ℓ_{ij} denotes $2n(i-1)+2j+1$ and the line number ℓ'_{ij} denotes $\ell_{ij}+1$ for $1 \leq i \leq m$ and $1 \leq j \leq n$. Note that $\ell_{11} = 3$.)

1.	$\widehat{U} \vee \widehat{V}$	Assumption.
2.	$\neg\widehat{V} \supset \widehat{U}$	1, PL.
ℓ_{11}.	$(\widehat{A_1} \supset \alpha_{11}) \wedge (\widehat{B_1} \supset \neg\alpha_{11})$	$\alpha_{11} \in A_1 \setminus B_1$.
ℓ'_{11}.	$\widehat{A_1} \supset \neg\widehat{B_1}$	ℓ_{11}, PL.
	...	
ℓ_{ij}.	$(\widehat{A_i} \supset \alpha_{ij}) \wedge (\widehat{B_j} \supset \neg\alpha_{ij})$	$\alpha_{ij} \in A_i \setminus B_j$.
ℓ'_{ij}.	$\widehat{A_i} \supset \neg\widehat{B_j}$	ℓ_{ij}, PL.
	...	
ℓ_{mn}.	$(\widehat{A_m} \supset \alpha_{mn}) \wedge (\widehat{B_n} \supset \neg\alpha_{mn})$	$\alpha_{mn} \in A_m \setminus B_n$.
ℓ'_{mn}.	$\widehat{A_m} \supset \neg\widehat{B_n}$	ℓ_{mn}, PL.
$\ell'_{mn}+1$.	$(\widehat{A_1} \vee \cdots \vee \widehat{A_m}) \supset (\neg\widehat{B_1} \wedge \cdots \wedge \neg\widehat{B})$	$\ell'_{11}, \ldots, \ell'_{mn}$, PL.
$\ell'_{mn}+2$.	$\widehat{U} \supset \neg\widehat{V}$	$\ell'_{mn}+1$, def. of $\widehat{U}, \widehat{V}$, PL.
$\ell'_{mn}+3$.	$\widehat{U} \equiv \neg\widehat{V}$	2, $\ell'_{mn}+2$, PL.

Now it follows from definitions of U and V and by Lemma 3.8 that $\vdash \widehat{U} \vee \widehat{V}$. From the above derivation $\vdash \widehat{U} \equiv \neg\widehat{V}$. ⊣

Lemma 3.10 *Let $\alpha \in cl$, and let U denote the set $\{A \in AT \mid \alpha \in A\}$. Then $\vdash \alpha \equiv \widehat{U}$.*

Proof: Let U be $\{A_1, \ldots, A_m\}$ and let $V = AT \setminus U$ be $\{B_1, \ldots, B_n\}$. Then we have the following derivation.

1.	$(\widehat{A_1} \supset \alpha) \wedge \cdots \wedge (\widehat{A_m} \supset \alpha)$	$\alpha \in A_i$ for $1 \leq i \leq m$.
2.	$\widehat{U} \supset \alpha$	1, def. of $\widehat{U}$, PL.
3.	$(\widehat{B_1} \supset \neg\alpha) \wedge \cdots \wedge (\widehat{B_n} \supset \neg\alpha)$	$\alpha \notin B_j$, and hence $\neg\alpha \in B_j$ for $1 \leq j \leq n$.
4.	$(\alpha \supset \neg\widehat{B_1}) \wedge \cdots \wedge (\alpha \supset \neg\widehat{B_n})$	3, PL
5.	$\alpha \supset \neg\widehat{V}$	4, def. of $\widehat{V}$, PL.
6.	$\widehat{U} \equiv \neg\widehat{V}$	Lemma 3.9.
7.	$\alpha \supset \widehat{U}$	5, 6, PL.
8.	$\alpha \equiv \widehat{U}$	2, 7, PL.

This completes the proof of the lemma. ⊣

Lemma 3.11 *Suppose α and β are formulas and π is a program such that for all $A \in AT$, either $\vdash \alpha \supset [\pi]\neg\widehat{A}$ or $\vdash \widehat{A} \supset \beta$. Then $\vdash \alpha \supset [\pi]\beta$.*

Proof: Let AT be $\{A_1, \ldots, A_r\}$. Consider an arbitrary atom A_i. If $\vdash \widehat{A_i} \supset \beta$, then we have the following sequence of derivations.

$\vdash \widehat{A_i} \supset \beta$ Assumption
$\vdash [\pi](\widehat{A_i} \supset \beta)$ Applying rule (G)
$\vdash \alpha \supset [\pi](\widehat{A_i} \supset \beta)$ from the above, by propositional logic

If, on the other hand, $\vdash \alpha \supset [\pi]\neg\widehat{A_i}$, then we have the following sequence of derivations.

$\vdash \neg\widehat{A_i} \supset (\widehat{A_i} \supset \beta)$ Propositional Logic
$\vdash [\pi]\neg\widehat{A_i} \supset [\pi](\widehat{A_i} \supset \beta)$ Axiom (A1), Rule (G), and propositional logic
$\vdash \alpha \supset [\pi]\neg\widehat{A_i}$ Assumption
$\vdash \alpha \supset [\pi](\widehat{A_i} \supset \beta)$ Propositional Logic

Thus, for all $i : 1 \leq i \leq r$, we have $\vdash \alpha \supset [\pi](\widehat{A_i} \supset \beta)$. Now by propositional logic and the fact that $\vdash [\pi](\gamma \wedge \delta) \equiv ([\pi]\gamma \wedge [\pi]\delta)$, we immediately see that $\vdash \alpha \supset [\pi]((\widehat{A_1} \supset \beta) \wedge \cdots \wedge (\widehat{A_r} \supset \beta))$. It easily follows that $\vdash \alpha \supset [\pi]((\widehat{A_1} \vee \cdots \vee \widehat{A_r}) \supset \beta)$. But then $\widehat{AT} = \widehat{A_1} \vee \cdots \vee \widehat{A_r}$, and by Lemma 3.8 $\vdash \widehat{AT}$, so it immediately follows that $\vdash \alpha \supset [\pi]\beta$, as desired. ⊣

The atom graph for α_0

The atom graph The *atom graph for α_0* is defined to be $F = (AT, \mathscr{A}, \rightarrow)$ where, for all $A, B \in AT$ and $a \in \mathscr{A}$, $A \xrightarrow{a} B$ iff $\widehat{A} \wedge \langle a \rangle \widehat{B}$ is consistent.

The *atom model* is given by $M = (F, V)$ where for each $A \in AT$, $V(A) = A \cap \mathscr{P}$. Given M, the various $\xrightarrow{\pi}$'s for different programs π is defined in the standard manner, as described in Subsection 3.2.

The heart of the completeness proof is the following lemma.

Lemma 3.12 *For each atom $A \in AT$ and for each formula $\alpha \in cl$, $M, A \vDash \alpha$ iff $\alpha \in A$. In particular, $M, A_0 \vDash \alpha_0$.*

Proof: The proof is by induction on the length of the formula. We precisely define the length of a formula below. The notion is carefully defined to ensure that as many formulas in $sf(\alpha)$ as possible end up having length strictly less than that of α. The notions $|\alpha|$ for a formula α and $|\pi|$ for a program π are defined by simultaneous induction as follows: $|p| = 1$ for $p \in P$, $|\neg\alpha| = |\alpha| + 1$, $|\alpha \vee \beta| = |\alpha| + |\beta| + 1$, $|[\pi]\alpha| = |\pi| + |\alpha|$; $|a| = 1$ for $a \in \mathscr{A}$, $|\pi_1 + \pi_2| = |\pi_1 \cdot \pi_2| = |\pi_1| + |\pi_2| + 1$, $|\pi^*| = |\pi| + 1$, $|\alpha?| = |\alpha| + 1$.

Note that the definition ensures that $|[\pi_1]\alpha| < |[\pi_1 + \pi_2]\alpha|$ and $|[\pi_1][\pi_2]\alpha| < |[\pi_1 \cdot \pi_2]\alpha|$, for instance. It can be easily checked that all appeals to the induction hypothesis in the following proof are proper.

In what follows, we prove three claims by simultaneous induction.

(i) For each atom $A \in W$ and for each formula $\alpha \in cl$, $M, A \vDash \alpha$ iff $\alpha \in A$.

(ii) For any two atoms A and B, and any program π which "occurs" in α_0—more formally, any π such that $[\pi]\alpha \in cl$ for some α—if $A \xrightarrow{\pi} B$ and $[\pi]\alpha \in A$, then $\alpha \in B$.

(iii) For any two atoms A and B, and any program π which occurs in α_0, if $\widehat{A} \wedge \langle\pi\rangle\widehat{B}$ is consistent then $A \xrightarrow{\pi} B$.

Proof of (i)
Basis: If $\alpha = p \in \mathscr{P} \cap cl$, then $M, A \vDash p$ iff $p \in V(A)$ iff $p \in A$, by the definition of V.

Induction step:

$\alpha = \neg\beta \in cl$: Then $M, A \vDash \neg\beta$ iff $M, A \nvDash \beta$ iff (by the induction hypothesis) $\beta \notin A$ iff (by the fact that A is an atom) $\neg\beta \in A$.

$\alpha = \beta \vee \gamma \in cl$: Then $M, A \vDash \beta \vee \gamma$ iff $M, A \vDash \beta$ or $M, A \vDash \gamma$ iff (by the induction hypothesis) $\beta \in A$ or $\gamma \in A$ iff (by the fact that A is an atom) $\beta \vee \gamma \in X$.

$\alpha = [\pi]\beta \in cl$: We analyse this case in two parts:

($\Leftarrow$) Suppose that $[\pi]\beta \in A$. We have to show that $M, A \vDash [\pi]\beta$. Consider any atom B such that $A \xrightarrow{\pi} B$. By (ii), we know that $\beta \in B$. By induction hypothesis, $M, B \vDash \beta$. Since B is an arbitrary π-neighbour of A, $M, A \vDash [\pi]\beta$, as desired.

($\Rightarrow$) Suppose $M, A \vDash [\pi]\beta$. This means that for all atoms B such that $A \xrightarrow{\pi} B$, $M, B \vDash \beta$. In other words, for all atoms B such that $M, B \nvDash \beta$, it is not the case that $A \xrightarrow{\pi} B$. By (iii), this implies that for all such B, $\vdash \widehat{A} \supset [\pi]\neg\widehat{B}$. By induction hypothesis on β, $M, B \nvDash \beta$ iff $\beta \notin B$. Thus our earlier statement is equivalent to saying that for all atoms B such that $\beta \notin B$, $\vdash \widehat{A} \supset [\pi]\neg\widehat{B}$. By the properties of atoms, this is the same as saying that for all atoms B such that $\neg\beta \in B$, $\vdash \widehat{A} \supset [\pi]\neg\widehat{B}$. By propositional logic, axiom (A1) and rule (G), we can see that $\vdash \widehat{A} \supset [\pi]\neg\widehat{U}$, where U is the set of all atoms which contain $\neg\beta$. But by Lemma 3.10, we see that $\vdash \widehat{U} \equiv \neg\beta$. Therefore $\vdash \widehat{A} \supset [\pi]\beta$. But $\widehat{A}$ is a conjunction of formulas belonging to A and $[\pi]\beta \in cl$, and A is an atom, so it follows that $[\pi]\beta \in B$, as desired.

Proof of (ii)
Basis: Suppose $\pi = a \in \mathscr{A}$ and A and B are atoms. Let $[a]\alpha \in A$ and $\alpha \notin B$. Then $\neg\alpha \in B$. Now it is easy to see that $\vdash \widehat{A} \supset [a]\alpha$ and $\vdash \widehat{B} \supset \neg\alpha$. Thus $\vdash \alpha \supset \neg\widehat{B}$, and hence by rule (G), $\vdash [a]\alpha \supset [a]\neg\widehat{B}$. Therefore it follows that $\vdash \widehat{A} \supset [a]\neg\widehat{B}$. But this means that it is not the case that $A \xrightarrow{a} B$. Thus we see that if $A \xrightarrow{a} B$ and $[a]\alpha \in A$ then $\alpha \in B$.

Induction step:

$\pi = \pi_1 + \pi_2$: For any atom A, $[\pi_1 + \pi_2]\alpha \in A$ iff $[\pi_1]\alpha \in A$ and $[\pi_2]\alpha \in A$. Now $A \xrightarrow{\pi_1+\pi_2} B$ iff $A \xrightarrow{\pi_1} B$ or $A \xrightarrow{\pi_2} B$. In either case it follows from induction hypothesis that $\alpha \in B$.

$\pi = \pi_1 \cdot \pi_2$: For any atom A, $[\pi_1 \cdot \pi_2]\alpha \in A$ iff $[\pi_1][\pi_2]\alpha \in A$. Now $A \xrightarrow{\pi_1\cdot\pi_2} B$ iff there exists another atom C such that $A \xrightarrow{\pi_1} C$ and $C \xrightarrow{\pi_2} B$. Now by induction hypothesis it follows that $[\pi_2]\alpha \in C$ and again by induction hypothesis it follows that $\alpha \in B$.

$\pi = \pi_1^*$: For any atom A, $[\pi_1^*]\alpha \in A$ iff $\alpha \in A$ and $[\pi_1][\pi_1^*]\alpha \in A$. Consider any atom such that $A \xrightarrow{\pi_1^*} B$. This means that there exists a sequence of atoms $A_0, \ldots, A_k$ ($k \geq 0$) such that $A = A_0$, $B = A_k$

and for all $i : 0 \leq i < k$, $A_i \xrightarrow{\pi_1} A_{i+1}$. We prove by induction that $[\pi_1^*]\alpha \in A_i$ for all $i : 0 \leq i \leq k$. In particular, $[\pi_1^*]\alpha \in A_k = B$ and hence $\alpha \in B$, as desired.

Now for the induction. Clearly $[\pi_1^*]\alpha \in A_0$. Suppose $[\pi_1^*]\alpha \in A_i$. Then $[\pi_1][\pi_1^*]\alpha \in A_i$. But since $A_i \xrightarrow{\pi_1} A_{i+1}$, we can apply the induction hypothesis on $\xrightarrow{\pi_1}$ to conclude that $[\pi_1^*]\alpha \in A_{i+1}$, as desired.

$\pi = \beta?$: For any atom A, $[\beta?]\alpha \in A$ iff $\beta \notin A$ or $\alpha \in A$. By applying (i) on β, $\beta \notin A$ iff $M, A \not\models \beta$. Now $A \xrightarrow{\beta?} B$ iff $M, A \models \beta$ and $A = B$. This tells us that $\beta \in A$ and hence it has to be the case that $\alpha \in A = B$.

Proof of (iii)
Basis: For $a \in \mathscr{A}$, it immediately follows from the definition of $\xrightarrow{a}$ that whenever $\widehat{A} \wedge \langle a \rangle \widehat{B}$, $A \xrightarrow{a} B$.

Induction step:

$\pi = \pi_1 + \pi_2$: Suppose $\pi_1 + \pi_2$ occurs in α_0. We prove the desired claim in the contrapositive form. It is not the case that $A \xrightarrow{\pi_1+\pi_2} B$ iff it is not the case that $A \xrightarrow{\pi_1} B$ and it is not the case that $A \xrightarrow{\pi_2} B$. But by induction hypothesis, this implies that $\vdash \widehat{A} \supset [\pi_1]\neg\widehat{B}$ and $\vdash \widehat{A} \supset [\pi_2]\neg\widehat{B}$. It immediately follows from Axiom (A2) that $\vdash \widehat{A} \supset [\pi_1 + \pi_2]\neg\widehat{B}$.

$\pi = \pi_1 \cdot \pi_2$: Suppose $\pi_1 \cdot \pi_2$ occurs in α_0. We prove the desired claim in the contrapositive form. It is not the case that $A \xrightarrow{\pi_1 \cdot \pi_2} B$ iff it is not the case that there exists an atom C such that $A \xrightarrow{\pi_1} C$ and $C \xrightarrow{\pi_2} B$. But by induction hypothesis, this implies that for all atoms C, $\vdash \widehat{A} \supset [\pi_1]\neg\widehat{C}$ or $\vdash \widehat{C} \supset [\pi_2]\neg\widehat{B}$. But we can appeal to Lemma 3.11 now—with $\widehat{A}$ in place of α, $[\pi_2]\neg\widehat{B}$ in place of β, and π_1 in place of π—and conclude that $\vdash \widehat{A} \supset [\pi_1][\pi_2]\neg\widehat{B}$. But now it follows from Axiom (A3) that $\vdash \widehat{A} \supset [\pi_1 \cdot \pi_2]\neg\widehat{B}$, as desired.

$\pi = \alpha?$: Suppose $\alpha?$ occurs in α_0. We prove the desired claim in the contrapositive form. It is not the case that $A \xrightarrow{\alpha?} B$ iff it is either the case that $M, A \not\models \alpha$ or it is the case that $A \neq B$. In the first case, by (i) applied to α, $\alpha \notin A$ and hence $\neg\alpha \in A$ (A being an atom and α being in *cl*). Therefore $\vdash \widehat{A} \supset \neg\alpha$ and, sure enough, $\vdash \widehat{A} \supset (\alpha \supset \neg\widehat{B})$. In the second case, it is clear that there is some $\beta \in A$ such that $\neg\beta \in B$. It therefore follows that $\vdash \widehat{A} \supset \neg\widehat{B}$ and therefore $\vdash \widehat{A} \supset (\alpha \supset \neg\widehat{B})$. So in both cases it is clear that $\vdash \widehat{A} \supset (\alpha \supset \neg\widehat{B})$. But by Axiom (A6), this is the same as saying that $\vdash \widehat{A} \supset [\alpha?]\neg\widehat{B}$, as desired.

$\pi = \pi_1^*$: Suppose π_1^* occurs in α_0. We prove the desired claim in the contrapositive form. Suppose it is not the case that $A \xrightarrow{\pi_1^*} B$. Define U to be the set of all π_1-reachable worlds from A, i.e. $U = \{C \in AT \mid A \xrightarrow{\pi_1^*} C\}$. Clearly $B \neq C$ for all $C \in U$. Now for any two distinct atoms C and D, it is easy to see that $\vdash \widehat{C} \supset \neg\widehat{D}$. Therefore it follows that $\vdash \widehat{U} \supset \neg\widehat{B}$. Now suppose we prove that $\vdash \widehat{U} \supset [\pi_1]\widehat{U}$. Then it follows by axiom (A5), rule (G), and propositional logic that $\vdash \widehat{U} \supset [\pi_1^*]\widehat{U}$. But $\vdash \widehat{A} \supset \widehat{U}$ (since $A \in U$) and $\vdash \widehat{U} \supset \neg\widehat{B}$. Therefore by axiom (A1), rule (G) and propositional logic it follows that $\vdash \widehat{A} \supset [\pi_1^*]\neg\widehat{B}$, as desired. It is only left to verify the following claim.

Claim $\vdash \widehat{U} \supset [\pi_1]\widehat{U}$.

Proof Let $V = AT \setminus U$. By Lemma 3.9, $\vdash \widehat{U} \equiv \neg\widehat{V}$. Thus it suffices to show that $\vdash \widehat{U} \supset [\pi_1]\neg\widehat{V}$. Consider any $C \in U$ and $D \in V$. Then it is clear that D is *not* a π_1-neighbour of Y. (If it were, then by definition of U, D would also belong to U, which is a contradiction.) The fact that D is not a π_1-neighbour of C and the induction hypothesis on $\xrightarrow{\pi_1}$ immediately imply that $\widehat{C} \wedge \langle\pi_1\rangle\widehat{D}$ is not consistent. In other words, $\vdash \widehat{C} \supset [\pi_1]\neg\widehat{D}$. But this holds for every $C \in U$ and $D \in V$. Thus by axiom (A1), rule (G) and propositional logic, $\vdash \widehat{U} \supset [\pi_1]\neg\widehat{V}$, and the claim follows.

This completes the proof of Lemma 3.12, and hence of Lemma 3.4. ⊣

Once we have proved Lemma 3.4, we immediately obtain a proof of completeness (Corollary 3.5) using exactly the same argument as in propositional logic. Note that we not only have completeness but also the small model property for dynamic logic, as follows: whenever α is satisfiable it is consistent (by soundness), whence it is satisfied in the atom model for α (which is of size at most $2^{2\cdot|\alpha|^2}$). Thus we also see that the satisfiability problem for dynamic logic is decidable.

4 First-Order Logic

Consider typical structures which we come across in mathematics and computer science—graphs, groups, monoids, rings, fields, ... A graph, for instance, is a set of vertices with a binary relation on this set which defines the edges. A group is a set equipped with a special constant (identity) and a binary function on the set which is associative. In general, all these structures consist of an underlying set of elements together with relations and functions defined over this set which satisfy certain properties.

First-order logic provides a natural framework for talking about such structures. In first-order logic, we begin by fixing abstract symbols to denote relations, functions and constants. These can then be combined using the usual propositional connectives built up from $\neg$ (not) and $\vee$ (or). In addition, first-order logic provides the means to *quantify* over elements[4] in the structure—we have the existential quantifier $\exists$ (read as "there exists") and its dual, the universal quantifier $\forall$ (read as "for all"). The logic also has the symbol $\equiv$, denoting equality, as a primitive construct.[5]

Defining the precise syntax and semantics of first-order logic is a little more involved than for propositional or modal logics. Before getting into the details, let us look at an informal example.

Groups in first-order logic As we know, a group is a structure $(G, +, 0)$ where G is a set, $0 \in G$ is a special element called the *identity* and $+ : G \times G \to G$ is a binary operation such that the following properties hold:

- The operation $+$ is associative.
- The constant 0 is a right-identity for the operation $+$.
- Every element in G has a right-inverse—that is, for each $x \in G$ we can find another element $y \in G$ such that $x + y = 0$.

To formalise this in first-order logic, we have to first fix the symbols in the language. We choose a function symbol op which takes two arguments and a constant symbol ε. We can then write the following formulas.

(G1) $\forall x \forall y \forall z \; op(op(x, y), z) \equiv op(x, op(y, z))$

(G2) $\forall x \; op(x, \varepsilon) \equiv x$

(G3) $\forall x \exists y \; op(x, y) \equiv \varepsilon$

[4] The "first" in first-order logic refers to the limitation placed on the quantifiers. In first-order logic, we can only quantify over single elements of the underlying set. In second-order logic, we can quantify over functions and relations. In third-order logic we can quantify over sets of function etc.

[5] We use $\equiv$ in the logical language rather than $=$ to avoid any confusion between syntactic references to equality and "real" equality over sets.

To assign meaning to these formulas, we fix a set S and map the symbol *op* to a binary function f on S and ε to an element s of S. The symbol $\equiv$ is *assumed* to be interpreted as equality over the set S. The formula (G1) then captures the fact that the function f denoted by *op* is associative. The next formula expresses that the element s denoted by ε acts as a right identity for the function f. The last formula postulates the existence of a right inverse for each element in S. If the set S, the function f assigned to *op* and the element s assigned to ε "satisfy" the formulas (G1)–(G3), we say that the *structure* (S, f, s) is a *model* for (G1)–(G3). It should be clear that any model (S, f, s) of (G1)–(G3) is in fact a group. Conversely, any group $(G, +, 0)$ can be made a model of (G1)–(G3) by assigning $+$ to be the function[6] denoted by *op* and 0 to be the element denoted by ε. Thus, in a precise logical sense, the formulas (G1)–(G3) describe groups: a structure (S, f, s) is a group iff it is a model of (G1)–(G3).

Our goal is to explore the extent to which first-order logic can capture properties of mathematical structures. While several properties can be naturally described in the logic, we shall see that various useful properties cannot. In the process of arriving at these results, we shall formally analyse first-order logic as we have done other logics so far—we shall explore issues such as compactness, completeness and decidability.

4.1 Syntax

First-order languages To define the formulas of first-order logic, we have to first fix the underlying *language*. A *first-order language* is a triple $L = (R, F, C)$ where $R = \{r_1, r_2, \ldots\}$ is a countable set of *relation symbols*, $F = \{f_1, f_2, \ldots\}$ is a countable set of *function symbols* and $C = \{c_1, c_2, \ldots\}$ is a countable set of *constant symbols*. Each symbol $r \in R$ and $f \in F$ is associated with an *arity*, denoted $\#(r)$ or $\#(f)$, indicating how many arguments the symbol takes. We also fix a countable set $Var = \{v_1, v_2, \ldots\}$ of variables. We shall use $x, y, z, \ldots$ to denote typical elements of Var.

The set of first-order formulas over a first-order language L is built up from *atomic formulas* using the propositional connectives $\neg$ and $\vee$ and the existential quantifier $\exists$. To define atomic formulas, we first have to define the *terms* of the language L.

Terms Let $L = (R, F, C)$ be a first-order language. The set of *terms* over L is the smallest set satisfying the following conditions:

- Every constant symbol $c \in C$ is a term.
- Every variable $x \in Var$ is a term.
- Let $t_1, t_2, \ldots, t_n$ be terms over L and let $f \in F$ be a function symbol of arity n. Then $f(t_1, t_2, \ldots, t_n)$ is a term.

A term which does not contain any variables is called a *closed* term. Notice that if L contains no function symbols, then the only terms over L are constants from C and variables from Var.

[6] Notice that though we normally write the group operation $+$ in infix notation as $x + y$, it is just a binary function and can just as well be written $+(x, y)$.

As we described before in an informal way, to define the semantics of first-order logic we have to fix a structure with respect to which the formulas of the language are interpreted. This interpretation will map each term to a unique element of the set underlying the structure. It is helpful to think of terms as the "names" which we can generate within L to talk about elements in the structure we are interested in.

Atomic formulas Let $L = (R, F, C)$ be a first-order language. The *atomic formulas* over L are defined as follows:

- Let $r \in R$ be a relation symbol of arity n and let $t_1, t_2, \ldots, t_n$ be terms over L. Then, $r(t_1, t_2, \ldots, t_n)$ is an atomic formula.
- Let t_1 and t_2 be terms over L. Then, $t_1 \equiv t_2$ is an atomic formula.

Atomic formulas play the role of atomic propositions in propositional logic. The first type of atomic formula asserts that the n-tuple denoted by $\langle t_1, t_2, \ldots, t_n \rangle$ is part of the relation denoted by r while the second type of atomic formula asserts that two different terms t_1 and t_2 are in fact just different "names" for the same element. Both these types of statements can be unambiguously labelled as true or false once we have fixed a structure and the interpretation of the symbols in the language within that structure.

Formulas Having defined the atomic formulas, we can then define Φ_L, the set of *first-order formulas* over L. The set Φ_L is the smallest set satisfying the following conditions:

- Every atomic formula over L belongs to Φ_L.
- If $\varphi \in \Phi_L$ then $\neg\varphi \in \Phi_L$.
- If $\varphi, \psi \in \Phi_L$ then $\varphi \vee \psi \in \Phi_L$.
- If $\varphi \in \Phi_L$ and $x \in \mathit{Var}$, then $\exists x\ \varphi \in \Phi_L$.

As usual, we may use parentheses to disambiguate the structure of a formula. We can define derived propositional connectives $\wedge$, $\supset$ and $\equiv$ using $\neg$ and $\vee$ in the standard way. In addition, we define the dual of $\exists$ as follows:

$$\forall x\ \varphi \stackrel{\text{def}}{=} \neg\exists x\ \neg\varphi$$

4.2 Semantics

As we saw informally earlier, to give meaning to a first-order formula over a language $L = (R, F, C)$, we have to fix a set S and assign a relation over S to each relation symbol in R, a function over S to each function symbol in F and an element of S to each constant symbol in C.

First-order structures Let $L = (R, F, C)$ be a first-order language. A *first-order structure* for L is a pair $\mathscr{M} = (S, \iota)$ where S is a *non-empty* set and ι is a function defined over $R \cup F \cup C$ such that:

- For each relation symbol $r \in R$ with $\#(r) = n$, $\iota(r)$ is an n-ary relation over S—that is, $\iota(r) \subseteq S^n$.
- For each function symbol $f \in F$ with $\#(f) = n$, $\iota(f)$ is an n-ary function over S—that is, $\iota(f) : S^n \to S$.
- For each constant symbol $c \in C$, $\iota(c)$ is an element of S—that is, $\iota(c) \in S$.

For convenience, we often denote $\iota(r)$, $\iota(f)$ and $\iota(c)$ by $r^{\mathscr{M}}$, $f^{\mathscr{M}}$ and $c^{\mathscr{M}}$ respectively. We also refer to a first-order structure for L as an *L-structure*.

Once we define a first-order structure, we fix the meaning of the symbols in the first-order language. However, we also have to assign meanings to the variables in Var. Once this is done, we can assign meaning to all formulas in the language.

Interpretation Let $L = (R, F, C)$ be a first-order language. An *interpretation* of L is a pair $\mathscr{I} = (\mathscr{M}, \sigma)$ where $\mathscr{M} = (S, \iota)$ is a first-order structure for L and $\sigma : Var \to S$ is an *assignment* of elements of S to variables in Var. In informal usage, we say that an interpretation or a structure has a certain cardinality when we mean that the associated underlying set has that cardinality.

Let $\sigma : Var \to S$ be an assignment. We denote by $\sigma[x_1 \mapsto s_1, x_2 \mapsto s_2, \ldots, x_n \mapsto s_n]$ the modified assignment σ' where $\sigma'(x_i) = s_i$ for $i \in \{1, 2, \ldots, n\}$ and $\sigma'(z) = \sigma(z)$ for all variables $z \notin \{x_1, x_2, \ldots, x_n\}$. For an interpretation $\mathscr{I} = (\mathscr{M}, \sigma)$, we use $\mathscr{I}[x_1 \mapsto s_1, x_2 \mapsto s_2, \ldots, x_n \mapsto s_n]$ to denote the modified interpretation $(\mathscr{M}, \sigma[x_1 \mapsto s_1, x_2 \mapsto s_2, \ldots, x_n \mapsto s_n])$.

We mentioned earlier that terms are names for elements in the structure. We can now make this statement precise. Once we fix an interpretation $\mathscr{I}$, each term t over L maps to a unique element $t^{\mathscr{I}}$ of S. Let $\mathscr{I} = (\mathscr{M}, \sigma)$ where $\mathscr{M} = (S, \iota)$. Then:

- If t is a constant $c \in C$, $t^{\mathscr{I}} = c^{\mathscr{M}}$.
- If t is a variable $x \in Var$, $t^{\mathscr{I}} = \sigma(x)$.
- If t is of the form $f(t_1, t_2, \ldots, t_n)$ where $f \in F$, then $t^{\mathscr{I}} = f^{\mathscr{M}}(t_1^{\mathscr{I}}, t_2^{\mathscr{I}}, \ldots, t_n^{\mathscr{I}})$.

Satisfaction relation Let $L = (R, F, C)$ be a first-order language and let $\mathscr{I}$ be an interpretation for L. The notion of a formula $\varphi \in \Phi_L$ being satisfied under the interpretation $\mathscr{I} = (\mathscr{M}, \sigma)$ is denoted $\mathscr{I} \vDash \varphi$ and is defined as follows:

- $\mathscr{I} \vDash t_1 \equiv t_2$ if $t_1^{\mathscr{I}} = t_2^{\mathscr{I}}$.
- $\mathscr{I} \vDash r(t_1, t_2, \ldots, t_n)$ if $(t_1^{\mathscr{I}}, t_2^{\mathscr{I}}, \ldots, t_n^{\mathscr{I}}) \in r^{\mathscr{M}}$.
- $\mathscr{I} \vDash \neg\varphi$ if $\mathscr{I} \nvDash \varphi$.
- $\mathscr{I} \vDash \varphi \vee \psi$ if $\mathscr{I} \vDash \varphi$ or $\mathscr{I} \vDash \psi$.
- $\mathscr{I} \vDash \exists x\ \varphi$ if there is an element $s \in S$ such that $\mathscr{I}[x \mapsto s] \vDash \varphi$.

***Exercise** 4.1* Verify that the semantics of $\forall x\ \varphi$ is as follows:

$$\mathscr{I} \vDash \forall x\ \varphi \text{ if for each element } s \in S,\ \mathscr{I}[x \mapsto s] \vDash \varphi.$$

⊣

As usual, we say that a first-order formula $\varphi \in \Phi_L$ is *satisfiable* if there is an interpretation $\mathscr{I}$ based on an L-structure $\mathscr{M}$ such that $\mathscr{I} \vDash \varphi$. Similarly, a formula $\varphi \in \Phi_L$ is *valid* if for every L-structure $\mathscr{M}$ and every interpretation $\mathscr{I}$ based on $\mathscr{M}$, $\mathscr{I} \vDash \varphi$. A *model* of φ is an interpretation satisfying φ.

Bound and free variables Before looking at examples of how to describe properties of structures in first-order logic, let us look closer at the role that variables play in defining the meaning of a formula.

As we saw above, we need to augment an L-structure $\mathscr{M}$ with an assignment σ in order fully specify the meaning of formulas. In principle, σ fixes a value for all variables in Var. However, for a fixed formula φ, we only need to know the values fixed by σ for those variables mentioned in φ.

More precisely, we only need σ to fix values of variables which are not "quantified" within φ. In a formula of the form $\exists x\ \psi$ or $\forall x\ \psi$, the value assigned by σ to x is irrelevant in fixing the meaning of the overall formula—the semantics of the quantifiers forces us to look at *all* possible assignments for x in order to give meaning to the formula.

Formally, in a formula of the form $\exists x\ \psi$ the *scope* of the quantifier $\exists x$ is the formula ψ. We say that a variable x is *free* in φ if it does not occur within the scope of a quantifier $\exists x$. Otherwise, x is said to be bound. For a formula φ, the set of free variables of φ, denoted $FV(\varphi)$, is defined inductively as follows:

- If φ is an atomic formula $r(t_1, t_2, \ldots, t_n)$, $FV(\varphi)$ is the set of variables which are mentioned in $\{t_1, t_2, \ldots, t_n\}$.
- If φ is an atomic formula $t_1 \equiv t_2$, $FV(\varphi)$ is the set of variables which are mentioned in $\{t_1, t_2\}$.
- $FV(\neg\varphi) = FV(\varphi)$.
- $FV(\varphi \vee \psi) = FV(\varphi) \cup FV(\psi)$.
- $FV(\exists x\ \varphi) = FV(\varphi) \setminus \{x\}$.

In the rest of the notes, we often write $\varphi(x_1, x_2, \ldots, x_k)$ to denote the fact that $FV(\varphi) \subseteq \{x_1, x_2, \ldots, x_k\}$.

The following proposition, analogous to Proposition 1.7 of Propositional Logic, formalises the fact that the meaning of a formula does not depend on that portion of the assignment which lies outside its set of free variables.

Proposition 4.2 *Let L be a first-order language and $\varphi \in \Phi_L$. Let $\mathscr{M}$ be an L-structure and σ, σ' be a pair of assignments which agree on $FV(\varphi)$. Then $(\mathscr{M}, \sigma) \vDash \varphi$ iff $(\mathscr{M}, \sigma') \vDash \varphi$.*

In other words, to give meaning to a formula $\varphi(x_1, x_2, \ldots, x_n)$, it is sufficient to fix a structure $\mathscr{M}$ and an assignment for the variables $x_1, x_2, \ldots, x_n$ which are potentially free in φ, rather than specifying an assignment σ over all variables. Thus, we can write $(\mathscr{M}, [x_1 \mapsto s_1, x_2 \mapsto s_2, \ldots, x_n \mapsto s_n]) \vDash \varphi$ to indicate that $(\mathscr{M}, \sigma) \vDash \varphi$ for every assignment σ which assigns x_i the value s_i for $i \in \{1, 2, \ldots, n\}$.

Sentences A *sentence* is a first-order formula with no free variables. The formulas (G1)–(G3) which we wrote earlier to describe properties of groups are all sentences. From the preceding discussion, it is clear that the meaning of a sentence is fixed once we fix an L-structure for the language L—assignments play no role in defining the meaning of a sentence.

Corollary 4.3 *Let L be a first-order language and $\varphi \in \Phi_L$ a sentence. Let $\mathcal{M}$ be an L-structure and σ, σ' any pair of assignments. Then, $(\mathcal{M}, \sigma) \models \varphi$ iff $(\mathcal{M}, \sigma') \models \varphi$.*

In other words, for a sentence φ and an L-structure $\mathcal{M}$ it makes sense to directly write $\mathcal{M} \models \varphi$. As usual, if X is a set of sentences, we write $\mathcal{M} \models X$ to denote that $\mathcal{M} \models \varphi$ for each sentence $\varphi \in X$..

Logical consequence We formalise the notion of logical consequence in first-order logic in the same way that we have for propositional logic. Let X be a set of first-order sentences over L. We say that a sentence φ is a *logical consequence* of X, denoted $X \models \varphi$, if it is the case that for every structure $\mathcal{M}$, if $\mathcal{M} \models X$ then $\mathcal{M} \models \varphi$.

Thus, for instance, the first-order formulas which are valid over all groups are just those formulas which are logical consequences of the sentences (G1)–(G3) which we used to characterise groups.

We end this section with some notation about variables and some assumptions about substitution. Given a formula $\varphi(x_1, x_2, \ldots, x_n)$, where $\{x_1, x_2, \ldots, x_n\} \subseteq FV(\varphi)$, and terms $t_1, t_2, \ldots, t_n$, the formula $\varphi(x_1, x_2, \ldots, x_n)[x_1 \mapsto t_1, x_2 \mapsto t_2, \ldots, x_n \mapsto t_n]$ is obtained by substituting uniformly for x_i by t_i in φ for $i \in \{1, 2, \ldots, n\}$. In the process, it may be that a variable in one of the terms t_i accidentally "intrudes" into the scope of a quantifier in φ. For instance, consider the formula $\varphi(x) = \exists y\ \neg(x \equiv y)$ and $t = y$. If we blindly substitute x by t, we end up with the formula $\exists y\ \neg(y \equiv y)$, which is clearly not what was intended. In such cases, we assume that the bound variables in φ are renamed to avoid clashes—in the preceding example, $\exists y\ \neg(x \equiv y)\ [x \mapsto y]$ would result in a formula of the form $\exists z\ \neg(y \equiv z)$. We shall not go into the precise definition of this renaming operation, but it should be intuitively clear from the example. Henceforth, we implicitly assume that such renaming is performed whenever we substitute a term for a free variable in a formula. We frequently abbreviate the formula $\varphi(x_1, x_2, \ldots, x_n)[x_1 \mapsto t_1, x_2 \mapsto t_2, \ldots, x_n \mapsto t_n]$ as $\varphi(t_1, t_2, \ldots, t_n)$.

4.3 Formalisations in first-order logic

We have seen, informally, how to represent groups in terms of first-order logic. Now that we have the precise syntax and semantics of the logic in place, let us look at some more examples of how to describe properties of structures in the logic.

Groups revisited

As we saw earlier, the three sentences (G1)–(G3) characterize groups, in the sense that any structure $\mathcal{M} = (S, f, s)$ which is a model for (G1)–(G3) defines a group over the set S with group operation f and identity s.

In groups, the cancellation law holds. This says that for any three elements x, y, z in the group, if $x \circ z = y \circ z$, then $x = y$. Recall that the language we chose for groups consisted of a binary function symbol op and a constant ε. In this language, the cancellation law can be stated as follows:

$$\varphi_c \stackrel{\text{def}}{=} \forall x\ \forall y\ \forall z\ (op(x,z) \equiv op(y,z) \supset x \equiv z)$$

Since the cancellation law φ_c holds in all groups, we would expect that (G1), (G2), (G3) $\vDash \varphi_c$.

An element g in a group $(G, +, 0)$ such that $g \neq 0$ and $\underbrace{g + g + \cdots + g}_{n \text{ times}} = 0$ is said to be of order n. We can formulate the fact that a group has no elements of order two as follows:

$$\psi \stackrel{\text{def}}{=} \neg\exists x\ (\neg(x \equiv \varepsilon) \wedge op(x,x) \equiv \varepsilon)$$

In other words, if $\mathscr{M} = (S, f, s)$ is a model for (G1)–(G3) and $\mathscr{M} \vDash \psi$, then $\mathscr{M}$ is a group which has no elements of order two.

An abelian group is one in which the group operation is commutative. This is simple to state:

(Ab) $\forall x\ \forall y\ op(x,y) \equiv op(y,x)$

Thus, the set of sentences {(G1), (G2), (G3), (Ab)} characterize abelian groups.

Lest we get the impression that all interesting properties of groups can be captured easily in first-order logic, let us consider torsion groups. A group $(G, +, 0)$ is said to be a *torsion group* if every element of G has finite order—that is, for each $g \in G$, there is a natural number $n \geq 1$ such that $\underbrace{g + g + \cdots + g}_{n \text{ times}} = 0$. To formalize this in a "natural way", we would have to write a formula of the form

$$\forall x\ (x \equiv \varepsilon \vee op(x,x) \equiv \varepsilon \vee op(op(x,x),x) \equiv \varepsilon \vee \cdots)$$

This is an infinite formula and is not permitted by our syntax. We shall show later that we cannot capture this property in first-order logic, even if we are permitted an *infinite* set of formulas to replace this single formula of infinite width.

Equivalence relations

Let r be a binary relation symbol in the language. We can force r to be interpreted as an equivalence relation through the following three sentences.

- $\forall x\ r(x,x)$
- $\forall x\ \forall y\ (r(x,y) \equiv r(y,x))$
- $\forall x\ \forall y\ \forall z\ ((r(x,y) \wedge r(y,z)) \supset r(x,z))$

It should be clear that in any structure $\mathscr{M}$, these three sentences would force $r^{\mathscr{M}}$ to be reflexive, symmetric and transitive.

Order

Ordered structures occur frequently in mathematics. A *strict linear order* $<$ over a set S is a non-empty binary relation which is irreflexive and transitive and which has the property that any two distinct elements in S are related by $<$. For instance, the less-than ordering over the set of natural numbers is a strict linear order.

Using the same symbol $<$ to denote the ordering relation within our language, we can axiomatise linear order using the following sentences.

- $\forall x\ \neg(x < x)$
- $\forall x\ \forall y\ \forall z\ ((x < y \wedge y < z) \supset x < z)$
- $\forall x\ \forall y\ (x < y \vee x = y \vee y < x)$

Fields

Recall that a field is a structure $(F, +, \cdot, 0, 1)$ where:

- $(F, +, 0)$ is an abelian group.
- $\cdot$ is a associative, commutative operation over F with identity 1 such that $0 \neq 1$ and every element other than 0 has a right-inverse with respect to $\cdot$.
- The operation $\cdot$ distributes over the operation $+$.

***Exercise** 4.4* Using a first-order language with two binary function symbols and two constants, axiomatise fields. ⊣

Questions of cardinality

We can make assertions about the size of structures in first-order logic. Consider the sentence

$$\varphi_{\geq 2} \stackrel{\text{def}}{=} \exists x\ \exists y\ \neg(x \equiv y)$$

Clearly, any structure which models $\varphi_{\geq 2}$ must have at least two distinct elements in the underlying set. We can easily generalize this formula to $\varphi_{\geq n}$ for any natural number n as follows:

$$\varphi_{\geq n} \stackrel{\text{def}}{=} \exists x_1 \exists x_2 \cdots \exists x_n \bigvee_{i \neq j} \neg(x_i \equiv x_j)$$

Conversely, the negation $\neg\varphi_{\geq 2} = \forall x\ \forall y\ (x \equiv y)$ asserts that the underlying structure has at most one element. (In fact, since we only deal with non-empty structures, $\neg\varphi_{\geq 2}$ asserts that the structure has exactly one element).

We can thus combine formulas of the form $\varphi_{\geq n}$ and $\neg\varphi_{\geq m}$ to tightly bound the range of elements in the structure.

Alternatively, we can use the infinite family of sentences $\{\varphi_{\geq 2}, \varphi_{\geq 3}, \ldots\}$ to specify that the structure we are interested is not finite.

Modal logic as a fragment of first-order logic

As a final example of formalisation in first-order logic, let us look at how to embed modal logic within first-order logic. In order to achieve this, we have to show how to translate models and formulas of modal logic into first-order logic in such a way that a model $M = (F, V)$ satisfies a formula α iff the structure $\mathscr{M}$ corresponding to M satisfies the formula $\hat{\alpha}$ corresponding to α.

Let $\mathscr{P} = \{p_0, p_2, p_2, \ldots\}$ be the set of atomic propositions which are used in defining the formulas of modal logic. The first-order language L which we use to embed formulas over $\mathscr{P}$ will have a binary relation symbol r, to describe the underlying modal frame, and unary relation symbols $\{P_0, P_1, P_2, \ldots\}$ to describe the valuation. Thus, an L-structure $\mathscr{M}$ would consist of a set S, which constitute the "possible worlds", together with a relation $r^{\mathscr{M}}$, describing the accessibility relation, and subsets $P_0^{\mathscr{M}}, P_1^{\mathscr{M}}, P_2^{\mathscr{M}}, \ldots$ describing the valuations $V(p_0), V(p_1), V(p_2), \ldots$.

We inductively define a translation $\{\alpha \mapsto \hat{\alpha}(x)\}$, where x is a variable, for all modal logic formulas over $\mathscr{P}$ as follows:

- For $p_i \in \mathscr{P}$, $\hat{p}_i(x) \stackrel{\text{def}}{=} P_i(x)$, where x is a variable.
- If $\alpha = \neg\beta$, then $\hat{\alpha}(x) \stackrel{\text{def}}{=} \neg\hat{\beta}(x)$.
- If $\alpha = \beta \vee \gamma$, then $\hat{\alpha}(x) \stackrel{\text{def}}{=} \hat{\beta}(x) \vee \hat{\gamma}(x)$.
- If $\alpha = \Box\beta$, then $\hat{\alpha}(x) \stackrel{\text{def}}{=} \forall y\ (r(x,y) \supset \hat{\beta}(y))$.

Proposition 4.5 *Let α be a modal logic formula over $\mathscr{P}$. Then, α is satisfiable iff $\hat{\alpha}(x)$ is first-order satisfiable.*

Proof: ($\Rightarrow$) Suppose that $M = (F, V)$, with $F = (W, R)$ such that for $w \in W$, $M, w \vDash \alpha$. We use W as the underlying set of our structure $\mathscr{M}$ and set $r^{\mathscr{M}} = R$ and $P_i^{\mathscr{M}} = V(p_i)$ for each $p_i \in \mathscr{P}$. We can then establish that $(\mathscr{M}, [x \mapsto w]) \vDash \hat{\alpha}(x)$ by induction on the structure of α.

For brevity, we only consider one case in detail, when α is of the form $\Box\beta$. Recall that $\hat{\alpha}(x)$ is then given by $\forall y\ (r(x,y) \supset \hat{\beta}(y))$. Since $M, w \vDash \Box\beta$, we know that for all elements $w' \in W$ such that $w\ R\ w'$, $M, w' \vDash \beta$. From the induction hypothesis and the fact that $r^{\mathscr{M}} = R$, it follows that for each y such that $[y \mapsto w']$ and $w\ R\ w'$, $(\mathscr{M}, [y \mapsto w']) \vDash \hat{\beta}(y)$. From the semantics of the universal quantifier, it then follows that $(\mathscr{M}, [x \mapsto w]) \vDash \hat{\alpha}(x)$.

($\Leftarrow$) Conversely, suppose that there is a structure $\mathscr{M}$ based on a set S such that for some $s \in S$, $(M, [x \mapsto s]) \vDash \hat{\alpha}(x)$. We must show that α is satisfiable. We fix our frame to be $(S, r^{\mathscr{M}})$ and for each $p_i \in \mathscr{P}$, we fix $V(p_i) = P_i^{\mathscr{M}}$. Once again, by induction on the structure of α, we can establish that $M, s \vDash \alpha$. We omit the details. $\dashv$

Our translation from modal logic to first-order logic allows us to reduce some questions about modal logic to the framework of modal logic. For instance, by the preceding proposition, questions about the satisfiability or validity of a formula α in modal logic can be phrased in terms of the first-order satisfiability or first-order validity of the corresponding formula $\hat{\alpha}(x)$.

We can even reduce more sophisticated questions to first-order logic. For instance, if we want to check whether a formula α is satisfiable over a frame whose accessibility relation is an equivalence relation, we can check the simultaneous satisfiability of $\hat{\alpha}(x)$ along with the three first-order sentences we saw earlier which capture the fact that the relation r is an equivalence relation. In general, questions about "relativised satisfiability" can be reduced to first-order logic whenever the properties demanded of the accessibility relation can be captured using first-order sentences.

We can even talk about satisfiability with respect to classes of frames which cannot be axiomatised in modal logic—for instance, the sentence $\forall y\,(\neg r(y,y))$ describes the class of irreflexive frames, which cannot be described in modal logic. In other words, the formula $\forall y\,(\neg r(y,y)) \wedge \hat{\alpha}(x)$ is satisfiable iff α is satisfiable over an irreflexive frame.

The disadvantage with reducing questions about modal logic to first-order logic is that first-order logic is too powerful from a computational point of view—for instance, we shall observe later that satisfiability is undecidable for first-order logic. On the other hand, we showed that for many systems of modal logic, satisfiability is in fact decidable.

***Exercise** 4.6* Let L be a finite first-order language and let $\mathscr{M}$ be a finite L-structure. Show that there is an L-sentence $\varphi_{\mathscr{M}}$ the models of which are precisely the L-structures isomorphic to $\mathscr{M}$. ⊣

***Exercise** 4.7*

(i) Let $L = \{+, \times, 0\}$ where $+$ and $\times$ are binary function symbols and 0 is a constant symbol. Consider the L-structure $(\mathbb{R}, +, \times, 0)$, where $\mathbb{R}$ is the set of real numbers with the conventional interpretation of $+$, $\times$ and 0 as addition, multiplication and zero.

Show that the relation $<$ ("less than") is elementary definable in $(\mathbb{R}, +, \times, 0)$—that is, there is a formula $\varphi(x,y)$ over L such that for all a, b in $\mathbb{R}$, $((\mathbb{R}, +, \times, 0), [x \mapsto a, y \mapsto b]) \models \varphi(x,y)$ iff $a < b$.

(ii) Let $L = \{+, 0\}$. Show that the relation $<$ is *not* elementary definable in $(\mathbb{R}, +, 0)$.

(Hint: Work with a suitable automorphism of $(\mathbb{R}, +, 0)$—that is, a suitable isomorphism of $(\mathbb{R}, +, 0)$ onto itself). ⊣

***Exercise** 4.8* Let $L = \{r\}$, where r is a binary relation. Formalize the following notions using sentences over L.

(i) r is an equivalence relation with at least two equivalence classes.

(ii) r is an equivalence relation with an equivalence class containing more than one element. ⊣

***Exercise** 4.9* A set M of natural numbers is called a *spectrum* if there is a language L and a sentence φ over L such that

$$M = \{n \mid \varphi \text{ has a model containing exactly } n \text{ element}\}$$

Show that:

(i) Every finite subset of $\{1, 2, 3, \ldots\}$ is a spectrum.

(ii) For every $m \geq 1$, the set of numbers greater than 0 which are divisible by m is a spectrum.

(iii) The set of squares greater than 0 is a spectrum.

(iv) The set of nonprime numbers greater than 0 is a spectrum.

(v) The set of prime numbers is a spectrum. ⊣

4.4 Satisfiability: Henkin's reduction to propositional logic

When is a set $X \subseteq \Phi_L$ of sentences in a first-order language L satisfiable—in other words, when can we find an L-structure $\mathscr{M}$ such that for each $\varphi \in X$, $\mathscr{M} \models \varphi$? Henkin proposed a solution to this question which essentially reduces the problem to one of satisfiability in a propositional framework.

For the rest of this discussion, we assume that we are working with a fixed first-order language $L = (R, F, C)$.

Let r be a binary relation symbol in L and t_1, t_2 be a pair of terms. It is immediate that the formula $r(t_1, t_2) \wedge \neg(r(t_1, t_2))$ is *not* satisfiable—we can treat $r(t_1, t_2)$ as an atomic proposition and recognize that this is an instance of an unsatisfiable propositional formula. How about a formula of the form $\forall x\ r(x, t_2) \wedge \neg\exists x\ r(x, t_2)$? Since we assumed that all structures are non-empty, we can check that this, too, is not satisfiable. However, there is no immediate way to represent this as an unsatisfiable propositional formula.

Henkin's approach is the following. Expand the language L by adding new constants. Use these new constants to define a special set of formulas and, using this set, blow up the given set X of formulas whose satisfiability we want to check into a larger set X' such that X is satisfiable iff X' is satisfiable. Show that X' is such that its satisfiability can be deduced from its "propositional structure".

We begin by defining a notion of "atomic proposition" with respect to first-order formulas.

Prime formulas A *prime formula* over L is an atomic formula or a formula which begins with the quantifier $\exists$. Let $\mathscr{P}_L$ be the set of prime formulas over L.

***Example** 4.10* In the formula $\exists x\ r(x) \vee t_1 \equiv t_2$, the prime formulas are $\exists x\ r(x)$ and $t_1 \equiv t_2$. In the formula $\forall x\ s(x) \supset \exists x\ s(x)$, after rewriting $\forall$ in terms of $\exists$, we have two prime formulas—$\exists x\ \neg s(x)$ and $\exists x\ s(x)$.

Observe that every formula in Φ_L can be constructed from prime formulas using the propositional connectives $\neg$ and $\vee$. The idea is to treat each distinct prime formula as an independent atomic proposition and deduce the satisfiability of a set $X \subseteq \Phi_L$ from the propositional structure of its prime formulas.

Propositional satisfiability We say a formula $\varphi \in \Phi_L$ is *propositionally satisfiable* if there is a valuation $v : \mathscr{P}_L \to \{\top, \bot\}$ such that the prime formula structure of φ evaluates to $\top$ under v. An *L-tautology* is a formula in Φ_L which evaluates to $\top$ for every propositional valuation to the prime formulas $\mathscr{P}_L$.

***Example** 4.11* The formula $\exists x\ r(x) \vee t_1 \equiv t_2$ is propositionally satisfiable. We can assign either prime formula (or both) independently to $\top$ to satisfy this formula. On the other hand, the formula $\exists x\ r(x) \wedge \neg\exists x\ r(x)$ is not propositionally satisfiable—the formula is built up from a single prime formula and has the structure $p \wedge \neg p$.

The formulas $\exists x\ r(x) \vee \neg(\exists x\ r(x))$ is a tautology over L—the formula is built up from a single prime formula and has the structure $p \vee \neg p$. Another example of a tautology over L is the formula, $\exists x\ r_1(y) \supset \forall y\ r_2(y) \vee \exists x\ r_1(x)$, which is of the form $p \supset q \vee p$.

Proposition 4.12 *Let $\mathscr{I}$ be an L-interpretation. There exists a valuation v of $\mathscr{P}_L$ such that for each formula φ, $\mathscr{I} \models \varphi$ iff $v \models \varphi$.*

Proof: For each prime formula ψ, define $v(\psi) = \top$ if $\mathscr{I} \models \psi$ and $v(\psi) = \bot$ otherwise. Since each first-order formula can be built up from prime formulas using the connectives $\neg$ and $\vee$, the result follows. ⊣

Corollary 4.13 *Let $X \subseteq \Phi_L$ be a set of formulas. If X is first-order satisfible, then X is propositionally satisfiable.*

The converse of the preceding Corollary is false. Consider the following examples.

***Example** 4.14* The set $\{c \equiv d, d \equiv e, \neg(c \equiv e)\}$ is propositionally satisfiable—we can fix a valuation which maps the prime formulas $c \equiv d$ and $d \equiv e$ to $\top$ and $c \equiv e$ to $\bot$. However, it is clearly not first-order satifiable.

The set of formulas $\{\forall x\ (r(x) \supset s(x)), \forall x\ r(x), \exists x\ \neg s(x)\}$, is propositionally satisfiable—once again, the three formulas in the set are made up of different prime formulas whose truth value can be assigned independently to make the whole set propositionally true. However, this set is not first-order satisfiable.

The preceding examples show that the prime formula structure of Φ_L does not accurately capture the effect of the equality relation and the role played by quantifiers in the semantics of first-order logic. Henkin's solution is to add extra formulas which "tie together" formulas connected by the equality relation and quantifiers so that the truth of one formula is linked to the truth of the other.

For instance, if we augment the set $\{c \equiv d, d \equiv e, \neg(c \equiv e)\}$ with the formula $\{(c \equiv d) \wedge (d \equiv e) \supset (c \equiv e)\}$, the set is no longer propositionally satisfiable. The new formula links the truth value of the prime formulas $c \equiv d$ and $d \equiv e$ to that of $c \equiv e$. Clearly the formula we have added is true in any structure, so it has not altered the first-order satisfiability of the original set.

Similarly, consider the second example $\{\forall x\ (r(x) \supset s(x)), \forall x\ r(x), \exists x\ \neg s(x)\}$, which may be rewritten as $\{\neg\exists x\ (r(x) \wedge \neg s(x)), \neg\exists x\ \neg r(x), \exists x\ \neg s(x)\}$.

If a sentence of the form $\exists y\ \varphi(y)$ is satisfied in a structure, we can use a term t to denote the "witnessing" element where φ holds. With this intended interpretation of t, we can append the sentence $\exists y\ \varphi(y) \supset \varphi(t)$ to the set containing $\exists y\ \varphi(y)$ without affecting its satisfiability.

Similarly, a sentence of the form $\neg\exists y\ \varphi(y)$ is satisfiable just in case $\neg\varphi(t)$ holds for *every* term t. Thus, we can expand a set of formulas containing $\neg\exists y\ \varphi(y)$ by a sentence $\neg\exists y\ \varphi(y) \supset \neg\varphi(t)$, where t is an *arbitrary* term, without affecting satisfiability.

If we apply this reasoning to the set $\{\neg\exists x\ (r(x) \wedge \neg s(x)), \neg\exists x\ \neg r(x), \exists x\ \neg s(x)\}$, we first identify a term t to witness the formula $\exists x\ \neg s(x)$ and add the formula $\exists x\ \neg s(x) \supset \neg s(t)$ to the set. Applying the rule for $\neg\exists y\ \varphi(y)$ to the other two formulas, we can then add $\neg\exists x\ (r(x) \wedge \neg s(x)) \supset \neg(r(t) \wedge \neg s(t))$ and $\neg\exists x\ \neg r(x) \supset \neg\neg r(t)$ to the set. A valuation which satisfies the three original formulas in the set must now also make the set $\{\neg(r(t) \wedge \neg s(t)), \neg\neg r(t), \neg s(t)\}$ true. This simplifies to $\{\neg r(t) \vee s(t), r(t), \neg s(t)\}$, which is not propositionally satisfiable. In other words, the expanded set is not propositionally satisfiable, which reflects the fact that the original set of three formulas was not first-order satisfiable.

Adding equality formulas, as we did in the first example, is not a problem. However, in the second case, we need to have a term to denote the witnessing element for each sentence $\exists y\ \varphi(y)$ in our set. It may be the case that the original language L does not have enough terms to cover all existential sentences of this form! In general, we have to expand the language in order to ensure that we do not run out of terms.

The Witnessing Expansion of L

Let $L = (R, F, C)$ be the original language, with $X \subseteq \Phi_L$ the set of sentences whose satisfiability we want to establish. We shall systematically add new constants to L in order to ensure that we have enough terms in the language to "name" all witnessing elements for existential sentences. Formally, we inductively define new sets of constants $C_0, C_1, \ldots$ as follows:

- Let $C_0 = \emptyset$ and let $L_0 = L$.
- Assume we have defined C_n. Let $L_n = (R, F, C \cup C_1 \cup C_2 \cup \cdots \cup C_n)$. For each formula $\varphi(x)$ of $\Phi_{L_n} \setminus \Phi_{L_{n-1}}$, with exactly one free variable x, let $c_{\varphi(x)}$ be a new constant, called the *witnessing constant* of the sentence $\exists x\ \varphi(x)$.

  Let C_{n+1} be the set of such constants generated by $\Phi_{L_n} \setminus \Phi_{L_{n-1}}$.

Let $C_H = \bigcup_{i \geq 0} C_i$ and let $L_H = (R, F, C \cup C_H)$.

Henkin and quantifier axioms

- The *Henkin axioms* are sentences over L_H of the form $\exists x\ \varphi(x) \supset \varphi(c_{\varphi(x)})$.
- The *quantifier axioms* are sentences over L_H of the form $\varphi(t) \supset \exists x\ \varphi(x)$, where t is a closed term over L_H.

It is clear that the quantifier axioms are true in any structure and are hence first-order valid. On the other hand, the Henkin axioms are not automatically true—we need to ensure that the witnessing constants are interpreted properly in the structure in order for the axioms to be true.

Let Φ_H denote the set of all instances of the Henkin axiom and Φ_Q denote the set of all instances of the quantifier axiom over the language L_H.

The equality axioms

Adding the equality axioms is easier. Let L_H be the witnessing expansion of L. To ensure that our propositional valuations respect the notion of equality, we define the following set of axioms capturing properties of equality. The equality axioms are all instances of the following, where t, u, v with or without subscripts are uniformly substituted by arbitrary terms over L_H, f is an arbitrary n-ary function symbol in L and r is an arbitrary n-ary relation symbol in L.

$$\begin{aligned}
& t \equiv t \\
t \equiv u \ &\supset\ u \equiv t \\
(t \equiv u \wedge u \equiv v) \ &\supset\ t \equiv v \\
(t_1 \equiv u_1 \wedge t_2 \equiv u_2 \wedge \cdots \wedge t_n \equiv u_n) \ &\supset\ (f(t_1, t_2, \ldots, t_n) \equiv f(u_1, u_2, \ldots, u_n)) \\
(t_1 \equiv u_1 \wedge t_2 \equiv u_2 \wedge \cdots \wedge t_n \equiv u_n) \ &\supset\ (r(t_1, t_2, \ldots, t_n) \supset r(u_1, u_2, \ldots, u_n))
\end{aligned}$$

Let Φ_{Eq} denote all instances of the equality axioms over L_H. Notice that though these axioms are not, in general, sentences, each formula in Φ_{Eq} is satisfied in *every* interpretation of L_H,

We now have the following lemma, which shows that satisifiability in first-order logic can be reduced to a similar question in propositional logic.

Lemma 4.15 (First-order satisfiability) *Let L be a first-order language and let L_H be the witnessing expansion of L. For any set X of formulas over L, the following are equivalent:*

(i) There is an L-interpretation $\mathscr{I} = (\mathscr{M}, \sigma)$ which is a model for X.

(ii) There is an L_H-interpretation $(\mathscr{M}, \sigma)$ which is a model for X.

(iii) $X \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq}$ is propositionally satisfiable.

Proof: The fact that (i) implies (iii) is easily proved. Let $\mathscr{I} = (\mathscr{M}, \sigma)$ be an L-interpretation which is a model for X, where $\mathscr{M} = (S, \iota)$. Define an L_H-interpretation $\mathscr{I}' = ((S, \iota'), \sigma)$, where ι' is defined on constants as follows: $\iota'(c) = \iota(c)$ for $c \in C$; for $c_{\varphi(x)} \in C_H$, $\iota(c_{\varphi(x)}) = o \in S$ such that $\mathscr{I} \models \varphi(o)$ if $\mathscr{M} \models \exists x\ \varphi(x)$ and o is arbitrary otherwise. It is clear that $\mathscr{I}' \models X \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq}$. It follows now from Proposition 4.12 that $X \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq}$ is propositionally satisfiable.

That (ii) implies (i) is immediate. The details are left as an exercise.

So, what remains is to establish that (iii) implies (ii). In other words, if there is a valuation v of the prime formulas over L_H such that $v \models X \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq}$, we must be able to construct an interpretation $\mathscr{I} = (\mathscr{M}, \sigma)$ where $\mathscr{M} = (S, \iota)$, such that $\mathscr{I} \models X$. We will in fact show that $\mathscr{I}$ has the property that for every formula φ over L_H, $\mathscr{I} \models \varphi$ iff $v \models \varphi$.

The main function of Φ_H is to ensure that if $v \models \exists x\ \varphi(x)$, then $v \models \varphi(c_{\varphi(x)})$ for every existential sentence $\exists x\ \varphi(x)$ over L_H.

To define $\mathscr{I}$, we must

(a) Define the underlying set S.

(b) Fix an interpretation $r^{\mathscr{M}} \subseteq S^n$ for each n-ary relation symbol r in L_H.

(c) Fix an interpetation $f^{\mathscr{M}} : S^n \to S$ for each n-ary function symbol f in L_H.

(d) Fix an interpretation $c^{\mathscr{M}} \in S$ for each constant symbol c in L_H.

(e) Fix an assignment σ.

The construction of $\mathscr{M}$ is as follows.

(a) Let $L_H = (R, F, C \cup C_H)$. To fix S, we define an equivalence relation $\simeq$ on terms over L_H by

$$t \simeq u \text{ iff } v \models t \equiv u$$

The equality axioms guarantee that $\simeq$ is in fact an equivalence relation. For instance, let us show that $\simeq$ is transitive. Suppose that $t \simeq u$ and $u \simeq w$. As an instance of the third equality axiom we have $t \equiv u \wedge u \equiv w \supset t \equiv w$. Since $v \models \Phi_{Eq}$, it must be the case that $v \models t \equiv u \wedge u \equiv w \supset t \equiv w$. Since $t \simeq u$ and $u \simeq w$, $v \models t \equiv u$ and $v \models u \equiv w$. Hence $v \models t \equiv w$ as well, which means that $t \simeq w$ as required. For each constant symbol t, let $[t]$ denote the equivalence class containing t. We define S to be the set $\{[t] \mid t \text{ is a term over } L_H\}$.

(b) Let r be an n-ary relation symbol. Fix $r^{\mathscr{M}} = \{\langle [t_1], [t_2], \ldots, [t_n]\rangle \mid v \models r(t_1, t_2, \ldots, t_n)\}$. To check that this is well-defined, we must verify that whenever $t_1 \simeq u_1$, $t_2 \simeq u_2$, …, $t_n \simeq u_n$ and $v \models r(t_1, t_2 \ldots, t_n)$ then $v \models r(u_1, u_2 \ldots, u_n)$ as well. As an instance of the last equality axiom, we have

$$t_1 \equiv u_1 \wedge t_2 \equiv u_2 \wedge \cdots \wedge t_n \equiv u_n \supset (r(t_1, t_2 \ldots, t_n) \supset r(u_1, u_2 \ldots, u_n))$$

Since $t_i \simeq u_i$ for $i \in \{1, 2, \ldots, n\}$, we have $v \models t_i \equiv u_i$ for $i \in \{1, 2, \ldots, n\}$. We also know that $v \models r(t_1, t_2, \ldots, t_n)$. Since $v \models \Phi_{Eq}$, it must then be the case that $v \models r(u_1, u_2, \ldots, u_n)$ as required.

(c) Let $t_1, \ldots, t_n$ be terms over L_H and f an n-ary function symbol in F. We define $f^{\mathscr{M}}([t_1], \ldots, [t_n])$ to be $[f(t_1, \ldots, t_n)]$.

To check that $f^{\mathscr{M}}$ is well-defined, we have to show that if $t_i \simeq u_i$ for $i \in \{1, 2, \ldots, n\}$, then $f(t_1, t_2, \ldots, t_n) \simeq f(u_1, u_2, \ldots, u_n)$. Let $t_i \simeq u_i$ for $i \in \{1, 2, \ldots, n\}$. This implies that $v \vDash t_i \equiv u_i$ for each i. From the fourth equality axiom, it then follows that $v \vDash f(t_1, t_2, \ldots, t_n) \equiv f(u_1, u_2, \ldots, u_n)$, so $f(t_1, t_2, \ldots, t_n) \simeq f(u_1, u_2, \ldots, u_n)$

(d) For $c \in C \cup C_H$, let $c^{\mathscr{M}} = [c]$.

(e) For $x \in \mathit{Var}$, let $\sigma(x) = [x]$.

This completes the construction of $\mathscr{M}$ and, at the same time, establishes that for atomic sentences φ, $\mathscr{M} \vDash \varphi$ iff $v \vDash \varphi$. Indeed, $\mathscr{I} \vDash r(t_1, \ldots, t_n)$ iff (by semantics) $\langle [t_1], \ldots, [t_n] \rangle \in r^{\mathscr{M}}$ iff (by definition) $v \vDash r(t_1, \ldots, t_n)$. On the other hand, $\mathscr{I} \vDash t_1 \equiv t_2$ iff (by semantics) $t_1^{\mathscr{I}} = t_2^{\mathscr{I}}$ iff (by definition) $t_1 \simeq t_2$ iff (by definition, again) $v \vDash t_1 \equiv t_2$.

To extend this argument to all sentences φ, we proceed by induction on the structure of φ. The cases where $\varphi = \neg\psi$ and $\varphi = \psi_1 \vee \psi_2$ are straightforward, so suppose that $\varphi = \exists x\ \psi(x)$.

If $(\mathscr{M}, \sigma) \vDash \varphi$ then there is an element s in the underlying set S such that $(\mathscr{M}, \sigma[x \mapsto s]) \vDash \psi(x)$. Since every element in S corresponds to an equivalence class $[t]$ for some term t over L_H, we can find a constant $t_s \in L_H$ such that $t_s^{\mathscr{M}} = s$. Clearly, $(\mathscr{M}, \sigma) \vDash \psi(t_s)$. By the induction hypothesis, $v \vDash \psi(t_s)$. Since $\psi(t_s) \supset \exists x\ \psi(x)$ is a quantifier axiom, we must have $v \vDash \psi(t_s) \supset \exists x\ \psi(x)$ and hence $v \vDash \exists x\ \psi(x)$, as required.

Conversely, suppose that $v \vDash \exists x\ \psi(x)$. Then, since $\exists x\ \psi(x) \supset \psi(c_{\psi(x)})$ is a Henkin axiom, we must have $v \vDash \psi(c_{\psi(x)})$ as well. By the induction hypothesis, it then follows that $\mathscr{I} \vDash \psi(c_{\psi(x)})$. From the semantics of the quantifier $\exists$, we must then have $\mathscr{I} \vDash \exists x\ \psi(x)$. ⊣

***Exercise** 4.16* Let L be a first-order language and let L_H be the witnessing expansion of L. Prove that for any set X of formulas over L, if there is an L_H-interpretation which is a model for X, there is also an L-interpretation which is a model for X. ⊣

4.5 Compactness and the Löwenheim-Skolem Theorem

Using the First-Order Satisfiability Lemma (Lemma 4.15), we can immediately derive some powerful and important results.

Theorem 4.17 (Compactness) *Let X be any set of First-Order formulas and let φ be a formula. Then, $X \vDash \varphi$ iff there is a finite subset $Y \subseteq_{\text{fin}} X$ such that $Y \vDash \varphi$.*

As we saw in the case of Propositional Logic (Page 11), this follows directly once we establish the following finite satisfiability result.

Lemma 4.18 (Finite Satisfiability) *Let L be a First-Order language and let X be a set of formulas over L. Then, X is satisfiable iff every $Y \subseteq_{\text{fin}} X$ is satisfiable.*

Proof: The non-trivial half of the statement is to show that if every $Y \subseteq_{\text{fin}} X$ is satisfiable then X is satisfiable. From the First-Order Satisfiability Lemma, it is sufficient to establish that $(X \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq})$ is propositionally satisfiable. From the Finite Satisfiability Lemma for propositional logic (Lemma 1.20), it suffices to show that every finite subset $(X \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq})$ is propositionally satisfiable. By assumption, each finite subset $Y \subseteq_{\text{fin}} X$ is satisfiable. From the First-Order Satisfiability Lemma, we can then conclude that for each $Y \subseteq_{\text{fin}} X$, $(Y \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq})$ is propositionally satisfiable. Since each finite subset of $(X \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq})$ is contained in $(Y \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq})$ for some $Y \subseteq_{\text{fin}} X$, it then follows that each finite subset of $(X \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq})$ is propositionally satisfiable. Thus, $(X \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq})$ is propositionally satisfiable, or, in other words, X is First-Order satisfiable. ⊣

To derive the Compactness Theorem from the Finite Satisfiability Theorem, we use the same argument as in propositional logic (Page 11).

The next result we derive from the First-Order Satisfiability Lemma has no counterpart in propositional logic.

Theorem 4.19 (Löwenheim-Skolem) *Let L be a first-order language and let X be a set of formulas over L.*

(i) If L is finite or countable, then if X is satisfiable, X is satisfiable in a structure whose underlying set is countable.

(ii) If L is not countable, then if X is satisfiable, X is satisfiable in a structure whose underlying set has a cardinality bounded by the cardinality of L.

Proof: Let us look at the first case in detail. If L is finite or countable, then Φ_L is countable. If X is satisifiable, then it is satisfiable in the structure constructed in the proof of Lemma 4.15. The underlying set in that structure is bounded by the number of constants in L together with the number of constants in the witnessing expansion of L. Recall the construction of C_H, the set of set of witnessing constants for L. Initially, C_1 contains a constant $c_{\varphi(x)}$ for each formula $\varphi(x) \in \Phi_L$. Since Φ_L is countable, so is C_1 and, thus, L_1 is countable. Inductively, assuming that L_n is countable, the same argument establishes that the next set of witnessing constants C_{n+1} is countable. Thus, C_H is the countable union of countable sets and is thus countable.

A similar argument applies in the second case. We omit the details. ⊣

In particular, the Löwenheim-Skolem Theorem says that if L is a countable first-order language, then no set of axioms over L can completely capture the properties of real numbers. Any attempt to describe a theory of real numbers over L will have to admit a countable model.

4.6 A Complete Axiomatisation

Before exploring the semantic consequences of the Compactness and Löwenheim-Skolem Theorems, let us look at an axiomatisation of first-order logic.

Axiom System FOL-AX The axiom system *FOL-AX* consists of three categories axioms and two inference rules.

(A1) All tautologies of propositional logic.

(A2a) $x \equiv x$

(A2b) $t \equiv u \supset (\varphi(t) \equiv \varphi(u))$, where φ is an atomic formula

(A3) $\varphi(t) \supset \exists x\ \varphi(x)$

(MP: Modus Ponens) $$\frac{\varphi,\ \varphi \supset \psi}{\psi}$$

(G: Generalisation) $$\frac{\varphi(x) \supset \psi}{\exists x\ \varphi(x) \supset \psi}, \text{ where } x \notin FV(\psi)$$

As usual, if X is a set of formulas over L, we write $X \vdash \varphi$ to indicate that there is a finite sequence of formulas $\varphi_1, \varphi_2, \ldots, \varphi_n$ such that $\varphi_n = \varphi$ and for each $i \in \{1, 2, \ldots, n\}$, φ_i is either a member of X, an instance of the axioms (A1)–(A3) or is derived from earlier formulas in the sequence using one of the two inference rules.

The following is an interesting lemma:

Lemma 4.20 *All the equality axioms over L can be derived using the above axioms and rules.*

Proof: Consider the equality axiom $t \equiv t$ for some term t. Here is a derivation of it:

1. $x \equiv x$ A2a.
2. $y \equiv y$ A2a.
3. $\neg(x \equiv x) \supset \neg(y \equiv y)$ 1, PL.
4. $\exists\, x \neg(x \equiv x) \supset \neg(y \equiv y)$ 3, rule (G).
5. $\neg(t \equiv t) \supset \exists\, x \neg(x \equiv x)$ A3.
6. $\neg(t \equiv t) \supset \neg(y \equiv y)$ 4,5,PL.
7. $t \equiv t$ 2,6,PL.

Now consider the equality axiom $t \equiv u \supset u \equiv t$. This is easily derivable as follows, where we let $\alpha(x)$ be $x \equiv t$ (note that $\alpha(t)$ is $t \equiv t$ and $\alpha(u)$ is $u \equiv t$):

1. $t \equiv t$ by the earlier derivation.
2. $t \equiv u \supset (\alpha(t) \equiv \alpha(u))$ A2b.
3. $t \equiv u \supset u \equiv t$ 1,2,PL.

Consider $(t \equiv u \wedge u \equiv v) \supset t \equiv v$. Again the following is an easy derivation, letting $\alpha(x)$ be $t \equiv x$ (note that $\alpha(u)$ is $t \equiv u$ and $\alpha(v)$ is $t \equiv v$):

1. $u \equiv v \supset (\alpha(u) \equiv \alpha(v))$ A2b.
2. $(t \equiv u \wedge u \equiv v) \supset t \equiv v$ 1, PL.

Now consider,without loss of generality, a ternary function symbol f and the equality axiom $(t_1 \equiv u_1 \wedge t_2 \equiv u_2 \wedge t_3 \equiv u_3) \supset (f(t_1, t_2, t_3) \equiv f(u_1, u_2, u_3))$. We provide a derivation below. We define $\alpha(x_1, x_2, x_3)$ to be $f(t_1, t_2, t_3) \equiv f(x_1, x_2, x_3)$, where x_1, x_2, x_3 do not occur in $t_1, t_2, t_3, u_1, u_2, u_3$. Also

define $\alpha_1(x_1)$ to be $f(t_1,t_2,t_3) \equiv f(x_1,t_2,t_3)$, $\alpha_2(x_2)$ to be $f(t_1,t_2,t_3) \equiv f(u_1,x_2,t_3)$, and $\alpha_3(x_3)$ to be $f(t_1,t_2,t_3) \equiv f(u_1,u_2,x_3)$. Notice that $\alpha_1(t_1)$ is the same as $f(t_1,t_2,t_3) \equiv f(t_1,t_2,t_3)$. Also notice that $\alpha_3(u_3)$ is just $f(t_1,t_2,t_3) \equiv f(u_1,u_2,u_3)$, so line 6 in the derivation below contains the desired formula. Further note that $\alpha_1(u_1)$ is the same as $\alpha_2(t_2)$ and $\alpha_2(u_2)$ is the same as $\alpha_3(t_3)$.

1. $\alpha_1(t_1)$ instance of $t \equiv t$.
2. $(t_1 \equiv u_1 \wedge t_2 \equiv u_2 \wedge t_3 \equiv u_3) \supset \alpha_1(u_1)$ A2b, 1, PL.
3. $(t_1 \equiv u_1 \wedge t_2 \equiv u_2 \wedge t_3 \equiv u_3) \supset \alpha_2(t_2)$ 2, $\alpha_1(u_1) \equiv \alpha_2(t_2)$, PL.
4. $(t_1 \equiv u_1 \wedge t_2 \equiv u_2 \wedge t_3 \equiv u_3) \supset \alpha_2(u_2)$ 3, A2b, PL.
5. $(t_1 \equiv u_1 \wedge t_2 \equiv u_2 \wedge t_3 \equiv u_3) \supset \alpha_3(t_3)$ 4, $\alpha_2(u_2) \equiv \alpha_3(t_3)$, PL.
6. $(t_1 \equiv u_1 \wedge t_2 \equiv u_2 \wedge t_3 \equiv u_3) \supset \alpha_3(u_3)$ 5, A2b, PL.

Consider, without loss of generality, a binary relation symbol r and the equality axiom $(t_1 \equiv u_1 \wedge t_2 \equiv u_2) \supset (r(t_1,t_2) \supset r(u_1,u_2))$. Let x_1, x_2 not occur in t_1, t_2, u_1, u_2. Now consider the following derivation:

1. $t_1 \equiv u_1 \supset (r(t_1,x_2) \supset r(u_1,x_2))$ A2b.
2. $(t_1 \equiv u_1 \wedge t_2 \equiv u_2) \supset (r(t_1,t_2) \supset r(u_1,u_2))$ 1,A2b,PL.

Thus we see that all the equality axioms are derivable in our axiom system. ⊣

The theorem we are after is the following:

Theorem 4.21 *Let X be a set of formulas over L and φ a sentence over L. Then $X \vdash L$ iff $X \models L$.*

As usual, the proof of this theorem is in two parts, soundness and completeness.

Lemma 4.22 (Soundness) *If $X \vdash \varphi$ then $X \models \varphi$.*

Proof: As usual, this is proved by the length of the derivation $X \vdash \varphi$. It suffices to argue that the axioms (A1)–(A3) are valid and that the rules (MP) and (G) preserve validity.

The validity of (A1) is obvious. We have already observed that (A2) and (A3) are valid when discussing the witnessing expansion used in the proof of Lemma 4.15.

As for the rules, when discussing propositional logic, we have already verified that (MP) preserves validity. Thus, we just need to argue that (G) preserves validity.

Suppose that the formula $\varphi(x) \supset \psi$ is valid, where $x \notin FV(\psi)$. In other words, for any interpretation $(\mathscr{M},\sigma)$, $(\mathscr{M},\sigma) \models \varphi(x) \supset \psi$.

Consider an arbitrary interpretation $(\mathscr{M}',\sigma')$, where $\mathscr{M}' = (S',\iota')$. We must show that if $(\mathscr{M}',\sigma') \models \exists x\ \varphi(x)$ then $(\mathscr{M}',\sigma') \models \psi$ as well.

Suppose that $(\mathscr{M}',\sigma') \models \exists x\ \varphi(x)$. From the semantics of the quantifier $\exists$, $(\mathscr{M}',\sigma') \models \exists x\ \varphi(x)$ iff for some $s \in S'$, $(\mathscr{M}',\sigma'[x \mapsto s]) \models \varphi(x)$. From the validity of $\varphi(x) \supset \psi$, we can conclude that $(\mathscr{M}',\sigma'[x \mapsto s]) \models \psi$. But, $x \notin FV(\psi)$, so $\sigma'[x \mapsto s]$ and σ' agree on $FV(\psi)$. From Proposition 4.2, it follows that $(\mathscr{M}',\sigma') \models \psi$ as well, as required. ⊣

To establish that the axiomatisation *AX-FOL* is complete, we need the following lemma.

Lemma 4.23 *Let X be a set of formulas.*

(i) *If* $X \vdash (\varphi \supset \psi)$ *and* $X \vdash (\neg\varphi \supset \psi)$ *then* $X \vdash \psi$.

(ii) *If* $X \vdash (\varphi \supset \theta) \supset \psi$, *then* $X \vdash (\neg\varphi \supset \psi)$ *and* $X \vdash (\theta \supset \psi)$.

(iii) *If* $x \notin FV(\psi)$ *and* $X \vdash [(\exists y\ \varphi(y) \supset \varphi(x)) \supset \psi]$, *then* $X \vdash \psi$.

Proof:

(i) Since $[(\varphi \supset \psi) \supset ((\neg\varphi \supset \psi) \supset \psi)]$ is a tautology, $X \vdash [(\varphi \supset \psi) \supset ((\neg\varphi \supset \psi) \supset \psi)]$. Given $X \vdash (\varphi \supset \psi)$ and $X \vdash (\neg\varphi \supset \psi)$, we can apply (MP) twice to obtain $X \vdash \psi$.

(ii) This follows from the fact that $[((\varphi \supset \theta) \supset \psi) \supset (\neg\varphi \supset \psi)]$ and $[((\varphi \supset \theta) \supset \psi) \supset (\theta \supset \psi)]$ are tautologies.

(iii) Suppose $X \vdash [(\exists y\ \varphi(y) \supset \varphi(x)) \supset \psi]$, where $x \notin FV(\psi)$. By (ii), $X \vdash (\neg\exists y\ \varphi(y) \supset \psi)$ and $X \vdash \varphi(x) \supset \psi$. We can apply rule (G) to the second formula and rename bound variables to obtain $X \vdash \exists y\ \varphi(y) \supset \psi$. From (i), it then follows that $X \vdash \psi$.

⊣

Lemma 4.24 (Completeness) *If* $X \models \varphi$ *then* $X \vdash \varphi$.

Proof: Suppose that $X \models \varphi$. Then, $X \cup \{\neg\varphi\}$ is not first-order satisfiable. By Lemma 4.15, $X \cup \neg\varphi \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq}$ is not propositionally satisfiable. From the Compactness Theorem for propositional logic, it follows that there is a finite subset $Y \subseteq_{\text{fin}} X \cup \Phi_H \cup \Phi_Q \cup \Phi_{Eq}$ such that $Y \cup \{\neg\varphi\}$ is not propositionally satisfiable.

Let the formulas in Y be listed in the order $\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_m$, such that:

- The sequence $\alpha_1, \alpha_2, \ldots, \alpha_n$ consists of those members of Y which belong to $X \cup \Phi_Q \cup \Phi_{Eq}$, listed in any order.
- The sequence $\beta_1, \beta_2, \ldots, \beta_m$ are those members of Y which belong to Φ_H. These sentences must be listed more carefully.

  Recall that L_H, the witnessing expansion of L, was constructed as the limit of a sequence of languages $L_0 \subsetneq L_1 \subsetneq \cdots$. For each formula ψ over L_H, let the rank of ψ be the least k such that ψ is a formula over L_k.

  The list $\beta_1, \beta_2, \ldots, \beta_m$ is arranged in such a way that the rank of β_i is greater than or equal to the rank of β_{i+1} for each $i \in \{1, 2, \ldots, m-1\}$. Recall that each β_i is of the form $\exists x\ \psi(x) \supset \psi(c_{\psi(x)})$—let us call $c_{\psi(x)}$ the witnessing constant for β_i. By arranging the list $\beta_1, \beta_2, \ldots, \beta_m$ in decreasing rank order, we ensure that the witnessing constant for β_i does not appear in $\beta_{i+1}, \beta_{i+2}, \ldots, \beta_m$ for each $i \in \{1, 2, \ldots, m-1\}$.

Since $Y \cup \{\neg\varphi\}$ is not propositionally satisfiable, we have

$$\alpha_1 \supset (\alpha_2 \supset \cdots \supset (\alpha_n \supset (\beta_1 \supset (\beta_2 \supset \cdots \supset (\beta_m \supset \varphi)\cdots)$$

to be a tautology, so we can derive

$$X \vdash \alpha_1 \supset (\alpha_2 \supset \cdots \supset (\alpha_n \supset (\beta_1 \supset (\beta_2 \supset \cdots \supset (\beta_m \supset \varphi)\cdots)$$

If we replace each witnessing constant in this formula by a distinct variable, the result

$$(\alpha'_1 \supset (\alpha'_2 \supset \cdots \supset (\alpha'_n \supset (\beta'_1 \supset (\beta'_2 \supset \cdots \supset (\beta'_m \supset \varphi')\cdots)$$

is still a tautology. Note however, that φ' is the same as φ, since φ is a formula over L and does not contain any witnessing constants.

Each formula in $\alpha'_1, \alpha'_2, \ldots, \alpha'_n$ is either a member of X or a logical axiom, so we may apply (MP) n times to obtain

$$X \vdash \beta'_1 \supset (\beta'_2 \supset \cdots \supset (\beta'_m \supset \varphi)\cdots)$$

Recall that each formula β'_i is of the form $\exists x\ \psi(x) \supset \psi(y)$, where the variable y does not appear in $\beta'_{i+1}, \beta'_{i+2}, \ldots, \beta'_m, \varphi$. We can thus apply Lemma 4.23 (iii) n times to obtain $X \vdash \varphi$. ⊣

4.7 Variants of the Löwenheim-Skolem Theorem

The Löwenheim-Skolem and Compactness Theorems play a dominant role in the semantics of first-order languages and in applying them to mathematical structures. Here we use the Compactness Theorem to obtain variants of the Löwenheim-Skolem Theorem.

Let us first prove the following easy consequence of the Compactness Theorem.

Theorem 4.25 *Let X be a set of formulas which has arbitrarily large finite models (i.e. for every $n \in \mathbb{N}$ there is a model for X whose cardinality is at least n). Then X also has a countable model.*

Proof: Let $Y \stackrel{\text{def}}{=} X \cup \{\varphi_{\geq n} \mid 2 \leq n\}$ ($\varphi_{\geq n}$ was presented in Subsection 4.3 under the head *Questions of cardinality*). Every model of Y is also a model of X and is infinite in size. Therefore we only need to prove that Y is satisfiable. By the Compactness Theorem it suffices to show that every finite subset Y_0 of Y is satisfiable. Each such Y_0 is a subset of $X_{n_0} \stackrel{\text{def}}{=} X \cup \{\varphi_{\geq n} \mid 2 \leq n \leq n_0\}$ for an appropriate $n_0 \in \mathbb{N}$. But according to hypothesis there is a model for X whose size is at least n_0. This is also a model for X_{n_0} and hence Y_0. Thus we are done. ⊣

We next prove that if a set of formulas has a model of a certain cardinality, it has models of every larger cardinality.

Theorem 4.26 ("Upward" Löwenheim-Skolem Theorem) *Let X be a set of formulas which has an infinite model. Then for every set A there is a model for X which has at least as many elements as A (what we mean is that there is an injective map from A into the underlying set).*

Proof: Let L be the language of X and let C be the set of constants in L. For each $a \in A$ let c_a be a new constant ($c_a \notin C$) such that $c_a \neq c_b$ for distinct $a, b \in A$. Let L' be the language L augmented with the set of constants $\{c_a \mid a \in A\}$. Suppose we show that the set $Y \stackrel{\text{def}}{=} X \cup \{\neg(c_a \equiv c_b) \mid a, b \in A, a \neq b\}$ of L'-formulas is satisfiable. Consider any model $\mathscr{I}$ of Y. Since $\mathscr{I} \models \neg(c_a \equiv c_b)$ for all distinct $a, b \in A$, it is clear that $\mathscr{I}(c_a) \neq \mathscr{I}(c_b)$ for distinct $a, b \in A$. Thus $\{(a, \mathscr{I}(c_a)) \mid a \in A\}$ is an injective map from A into the underlying set of $\mathscr{I}$, and the theorem would be proved.

We now turn our attention to proving that Y is indeed satisfiable. By Compactness it suffices to show that all finite subsets Y_0 of Y are satisfiable. But that is very easy to see. Every such Y_0 is a subset of $Z = X \cup \{\neg(c_{a_i} \equiv c_{a_j}) \mid 1 \leq i, j \leq n, i \neq j\}$ for some appropriate subset $\{a_1, \ldots a_n\}$ of A. Now let $\mathscr{I}$ be some infinite model for X. Clearly we can choose n distinct elements $b_1, \ldots, b_n$ from the underlying set of $\mathscr{I}$. We can now extend $\mathscr{I}$ to L' by setting $\mathscr{I}(a_i) = b_i$ for $i \leq n$, and giving $\mathscr{I}(c_a)$ an arbitrary value, for the other elements a occurring in A. It is easily checked that $\mathscr{I}$, extended as above, is a model for Z—and hence for Y_0. We are done. ⊣

The above theorem can be put to good use in the study of algebraic theories. For instance, let X be the set of group axioms. Since there exist infinite groups, the above theorem says that there exist arbitrarily large groups. Similarly, there are arbitrarily large orderings and arbitrarily large fields. While each of these facts can be derived using algebraic methods specific to the theory, first-order logic provides us with the framework and with methods to state and prove such results in a general form.

4.8 Elementary Classes

For a set of L-formulas we call

$$\text{Mod}^L X \stackrel{\text{def}}{=} \{\mathscr{I} \mid \mathscr{I} \text{ is an } L\text{-structure and } \mathscr{I} \models X\}$$

the *class of models* of X. We drop the superscript when there is no scope for confusion. We also write Mod φ instead of Mod $\{\varphi\}$.

Definition 4.27 *Let $\mathscr{C}$ be a class of L-structures.*

(i) $\mathscr{C}$ is called elementary *if there is an L-formula φ such that $\mathscr{C} = \text{Mod } \varphi$.*

(ii) $\mathscr{C}$ is called Δ-elementary *if there is as set X of L-formulas such that $\mathscr{C} = \text{Mod } X$.*

Every elementary class is Δ-elementary. Conversely, every Δ-elementary class is the intersection of elementary classes. This is because, for any set X of sentences, $\text{Mod } X = \bigcap_{\varphi \in X} \text{Mod } \varphi$.

In the rest of this section we will see some examples of elementary and Δ-elementary classes of structures. We will also see some examples of classes of structures that are not elementary, and some which are not Δ-elementary. This gives us an indication of the expressive power of first-order logic.

For example, the class of fields is elementary since it consists of precisely those models which satisfy the conjunction of the (finitely many) field axioms. The class of ordered fields is also elementary since order can also be characterised using a finite number of axioms. Similarly the class of groups, the class of equivalence relations, the class of partial orderings, the class of directed graphs, etc. are all elementary.

Fields of prime characteristic and of characteristic 0 Let p be a prime. A field F has characteristic p if $\underbrace{1+1+\cdots+1}_{p \text{ times}} = 0$. If there is no prime p for which F has characteristic p, F is said to have characteristic 0. For every prime p the field $\mathbb{Z}/(p)$ of the integers modulo p has characteristic p. The field $\mathbb{R}$ of real numbers has characteristic 0. Let φ_F be the conjunction of all the field axioms, and let χ_p be the formula $\underbrace{1+1+\cdots+1}_{p \text{ times}} \equiv 0$ (we use the 0 and 1 both as constant symbols of the language of fields as well as names of the additive and multiplicative identities of fields). Then the class of fields of characteristic p is exactly the same as $\mathrm{Mod}(\varphi_F \wedge \chi_p)$. Hence this class is elementary. The class of fields of characteristic 0 is Δ-elementary — it is easily seen to be the same as $\mathrm{Mod}(\{\varphi_F\} \cup \{\neg\chi_p \mid p \text{ is prime}\})$. In what follows, we show that it is not elementary.

Let φ be a sentence in the lenaguage of fields which is valid in all fields of characteristic 0, that is

$$\{\varphi_F\} \cup \{\neg\chi_p \mid p \text{ is prime}\} \models \varphi.$$

By the Compactness Theorem there is an n_0 such that

$$\{\varphi_F\} \cup \{\neg\chi_p \mid p \text{ is prime}, p < n_0\} \models \varphi.$$

Hence φ is valid in all fields of characteristic $\geq n_0$. Thus we have proved the following theorem.

Theorem 4.28 *A sentence (in the language of fields) which is valid in all fields of characteristic* 0 *is also valid in all fields whose characteristic is sufficiently large.*

From this we conclude that the class of fields of characteristic 0 is *not elementary*, for otherwise, there would have to be a sentence φ (characterising the class) which is valid precisely in all the fields of characteristic 0.

The class of finite structures and the class of infinite structures It is easily seen that the class of finite L-structures (for a fixed L), the class of finite groups, the class of finite fields are not Δ-elementary. The proof is simple: If, for example, the class of finite groups were of the form $\mathrm{Mod}\, X$, then X would be a set of formulas having arbitrarily large finite models (groups of the form $\mathbb{Z}/(p)$) but no infinite model. That would contradict Theorem 4.25.

On the other hand the corresponding classes of infinite structures is Δ-elementary. In fact, let $\mathscr{C}$ be any Δ-elementary class of structures, characterised by the set of formulas X. Then the class $\mathscr{C}^\infty$ of infinite structures in $\mathscr{C}$ is characterised by $X \cup \{\varphi_{\geq n} \mid n \geq 2\}$.

Torsion groups A group G is called a *torsion group* if every element is of finite order, i.e. if for every $a \in G$ there is an $n \geq 1$ such that $\underbrace{a+a+\cdots+a}_{n \text{ times}} = 0$. An ad-hoc formalization of this property would be

$$\forall x(x \equiv 0 \vee x + x \equiv 0 \vee x + x + x \equiv 0 \vee \cdots).$$

However, we may not form infinitely long disjunctions in first-order logic. Indeed, the class of torsion groups is not even Δ-elementary.

Suppose, for a contradiction, that X is a set of formulas that characterises the class of torsion groups. Let

$$Y \stackrel{\text{def}}{=} X \cup \{\neg(\underbrace{x + x + \cdots + x}_{n \text{ times}} \equiv 0) \mid n \geq 1\}.$$

Every finite subset Y_0 of Y has a model. Choose an n_0 such that $Y_0 \subseteq X \cup \{\neg(\underbrace{x + x + \cdots + x}_{n \text{ times}} \equiv 0) \mid 1 \leq n < n_0\}$. Then every cyclic group of order n_0 is a model of Y_0 if x is interpreted as the generating element. Since every finite subset of Y is satisfiable, Y is also satisfiable. Let $\mathscr{I}$ be a model of Y. Then $\mathscr{I}(x)$ does not have a finite order, showing that $\mathscr{I}$ is a model of X but not a torsion group, a contradiction.

The class of connected graphs A graph $G = (V, E)$ is said to be *connected* if, for arbitrary $a, b \in V$ with $a \neq b$, there are $n \geq 2$ and $a_1, \ldots, a_n \in V$ with

$$a_1 = a, a_n = b, a_i \; E \; a_{i+1} \text{ for } i = 1, \ldots, n-1$$

(i.e., if for any two distinct elements in V there is a path connecting them). For $n \in \mathbb{N}$, the regular $(n+1)$-gon G_n with the vertices $0, \ldots, n$ is a connected graph. More precisely, G_n is the structure (V_n, E_n) with $V_n \stackrel{\text{def}}{=} \{0, \ldots, n\}$ and

$$E_n \stackrel{\text{def}}{=} \{(i, i+1) \mid i < n\} \cup \{(i, i-1) \mid 1 \leq i \leq n\} \cup \{(0, n), (n, 0)\}.$$

We now prove that the class of connected graphs is not Δ-elementary. Assume, towards a contradiction, that a set X of formulas characterises the class of connected graphs. For $n \geq 2$ we set

$$\psi_n \stackrel{\text{def}}{=} \neg(x \equiv y) \wedge \neg \exists x_1 \ldots \exists x_n (x_1 \equiv x \wedge x_n \equiv y \wedge x_1 \; E \; x_2 \wedge \cdots \wedge x_{n-1} \; E \; x_n)$$

and

$$Y \stackrel{\text{def}}{=} X \cup \{\psi_n \mid n \geq 2\}.$$

Then every subset Y_0 of Y has a model: For Y_0 choose an n_0 such that $Y_0 \subseteq X \cup \{\psi_n \mid 2 \leq n < n_0\}$; then $G_{2 \cdot n_0}$ is a model of Y_0 if x is interpreted by 0 and y by n_0. Since every finite subset of Y has a model, Y is also satisfiable. Let $\mathscr{I}$ be a model of Y. Then there is no path connecting $\mathscr{I}(x)$ and $\mathscr{I}(y)$. Therefore $\mathscr{I}$ is a model of X but not a connected graph. This contradicts the assumption on X.

4.9 Elementarily Equivalent Structures

In this section, we look at a new notion of equivalence between structures based on the set of formulas they satisfy. This offers another interesting means of studying the power of first-order formulas. While in the previous section we were concerned with the *expressive power* of first-order logic (i.e., what classes of structures can be charaterised by first-order formulas?), in this section we look at the *distinguishing power* of first-order logic (i.e., when can a first-order formula tell two structures apart?). We can also sometimes prove facts about expressibility using facts about distinguishability. We will see some examples of this later. But first we begin by introducing two new notions.

Definition 4.29

(i) Two structures (for the same language) $\mathcal{M}$ and $\mathcal{M}'$ are called elementarily equivalent *(written: $\mathcal{M} \equiv \mathcal{M}'$) if for every formula φ (in the appropriate language) we have $\mathcal{M} \models \varphi$ iff $\mathcal{M}' \models \varphi$.*

(ii) For an interpretation $\mathcal{M}$ let $\mathrm{Th}(\mathcal{M}) \stackrel{\text{def}}{=} \{\varphi \mid \mathcal{M} \models \varphi\}$. $\mathrm{Th}(\mathcal{M})$ *is called the (first-order)* theory *of $\mathcal{M}$.*

Lemma 4.30 *For two structures $\mathcal{M}$ and $\mathcal{M}'$,*

$$\mathcal{M} \equiv \mathcal{M}' \text{ iff } \mathcal{M}' \models \mathrm{Th}(\mathcal{M}).$$

Proof: If $\mathcal{M} \equiv \mathcal{M}'$ then, since $\mathcal{M} \models \mathrm{Th}(\mathcal{M})$, also $\mathcal{M}' \models \mathrm{Th}(\mathcal{M})$. Conversely, suppose $\mathcal{M}' \models \mathrm{Th}(\mathcal{M})$. Consider a sentence φ. If $\mathcal{M} \models \varphi$ then $\varphi \in \mathrm{Th}(\mathcal{M})$ and hence $\mathcal{M}' \models \varphi$. If, on the other hand, $\mathcal{M} \not\models \varphi$ then $\mathcal{M} \models \neg\varphi$ and thus $\neg\varphi \in \mathrm{Th}(\mathcal{M})$. Hence $\mathcal{M}' \models \neg\varphi$ and therefore $\mathcal{M}' \not\models \varphi$. Thus $\mathcal{M} \equiv \mathcal{M}'$. ⊣

It can be easily seen by a simple (but probably tedious) induction that any two isomorphic structures satisfy the same first-order formulas. In other words, they are elementarily equivalent. The converse is not immediately clear though: Are any two elementarily equivalent structures isomorphic to each other?

Theorem 4.31 *For every structure $\mathcal{M}$, the class $\mathscr{C} = \{\mathcal{M}' \mid \mathcal{M}' \equiv \mathcal{M}\}$ is Δ-elementary; in fact $\mathscr{C} = \mathrm{Mod}\,\mathrm{Th}(\mathcal{M})$. Moreover, $\mathscr{C}$ is the smallest Δ-elementary class which contains $\mathcal{M}$.*

Proof: From Lemma 4.30 it is clear that $\mathcal{M}' \in \mathrm{Mod}\,\mathrm{Th}(\mathcal{M})$ iff $\mathcal{M}' \equiv \mathcal{M}$. Now if $\mathrm{Mod}\,X$ is another Δ-elementary class containing $\mathcal{M}$, then $\mathcal{M} \models X$ and therefore $\mathcal{M}' \models X$ for every $\mathcal{M}'$ elementarily equivalent to $\mathcal{M}$. Hence $\{\mathcal{M}' \mid \mathcal{M}' \equiv \mathcal{M}\} \subseteq \mathrm{Mod}\,X$. ⊣

Theorem 4.32 *If $\mathcal{M}$ is infinite then the class of all structures isomorphic to $\mathcal{M}$ is not Δ-elementary; in other words, no infinite structure can be characterised up to isomorphism by a set of first-order formulas.*

Proof: Let $\mathcal{M} = (S, \iota)$ be an infinite structure. Suppose, towards a contradiction, that X is a set of first-order formulas whose models are exactly the structures isomorphic to $\mathcal{M}$. X has an infinite model, and hence by the upward Löwenheim-Skolem theorem, X has a model $\mathcal{M}'$ with at least as many elements as the power set of S. But then $\mathcal{M}'$ is a model of X but not isomorphic to $\mathcal{M}$, contrary to what we supposed. This proves the theorem. ⊣

If we choose $X = \mathrm{Th}(\mathcal{M})$ in the above proof, then $\mathcal{M}$ and $\mathcal{M}'$ are elementarily equivalent but not isomorphic. This shows that not all elementarily equivalent structures are isomorphic to each other.

Theorem 4.31 tells us that a Δ-elementary class contains, together with any given structure, all elementarily equivalent ones. In certain cases, one can use this to show that a class $\mathscr{C}$ is not Δ-elementary. We simply specify two elementarily equivalent structures, of which one belongs to $\mathscr{C}$, and the other does not. We illustrate this method in the case of archimedean fields.

An ordered field F is called *archimedean* if for every $a \in F$ there is $n \in \mathbb{N}$ such that $a < \underbrace{1+1+\cdots+1}_{n \text{ times}}$.

For example, the ordered field of rational numbers and the ordered field of reals $\mathbb{R}^<$ are archimedean. We show that there is an non-archimedean ordered field elementarily equivalent to the ordered field of real numbers. This will prove the following.

Theorem 4.33 *The class of archimedean fields is not Δ-elementary.*

Proof: Let

$$X \stackrel{\text{def}}{=} \mathrm{Th}(\mathbb{R}^<) \cup \{\mathbf{0} < x, \mathbf{1} < x, \mathbf{2} < x, \ldots\},$$

where $\mathbf{0}, \mathbf{1}, \mathbf{2}, \ldots$ stand for the terms $0, 1, 1+1, \ldots$ in the language of arithmetic. Every finite subset of X is satisfiable, for instance, by an interpretation of the form $(\mathbb{R}^<, \sigma)$, where $\sigma(x)$ is a sufficiently large natural number. By the Compactness Theorem there is a model $(\mathscr{M}', \sigma')$ of X. Since $\mathscr{M}' \models \mathrm{Th}(\mathbb{R}^<)$, $\mathscr{M}'$ is an ordered field elementarily equivalent to $\mathbb{R}^<$, but (as shown by the element $\sigma'(x)$) is not archimedean. ⊣

The use of the Compactness Theorem in the above is typical. We provide further examples of its use below, when we turn our attention to the structure $\mathbb{N}$ of natural numbers, and the structure $\mathbb{N}^<$ of ordered natural numbers. (Note that the signatures of interest here are $\{0, s, +, \cdot\}$ and $\{0, s, +, \cdot, <\}$.) These structures can be characterised up to isomorphism by a finite set of axioms, usually called *Peano's axioms*, which includes a second-order induction axiom. But in what follows, we show that no system of first-order axioms can characterise the structures $\mathbb{N}$ and $\mathbb{N}^<$ up to isomorphism. From this it follows that the induction axiom cannot be formulated as set of first-order formulas.

A theory is said to be *categorical* if all its models are isomorphic to one another. The (second-order) Peano's axiom system is an example of a categorical theory. No first-order theory can be categorical; this is a consequence of the Löwenheim-Skolem theorems (both "upward" and "downward"). It is more interesting to study if a theory is $\aleph$-categorical for a given cardinal number $\aleph$. In particular, we are interested in seeing whether arithmetic is $\aleph_0$-categorical, i.e. if all the countable models of $\mathrm{Th}(\mathbb{N})$ are isomorphic to one another. The following two theorems say that $\mathrm{Th}(\mathbb{N})$ and $\mathrm{Th}(\mathbb{N}^<)$ are not $\aleph_0$-categorical.

Let us introduce another bit of terminology before stating the results. A structure which is elementarily equivalent, but not isomorphic to $\mathbb{N}$ is called a *nonstandard model of arithmetic*.

Theorem 4.34 (Skolem's Theorem) *There is a countable nonstandard model of arithmetic.*

Proof: Let

$$X \stackrel{\text{def}}{=} \mathrm{Th}(\mathbb{N}) \cup \{\neg(x \equiv \mathbf{0}), \neg(x \equiv \mathbf{1}), \neg(x \equiv \mathbf{2}), \ldots\},$$

where $\mathbf{0}, \mathbf{1}, \mathbf{2}, \ldots$ stands for the terms $0, s(0), s(s(0)), \ldots$. Every finite subset of X has a model of the form $(\mathbb{N}, \sigma)$, where $\sigma(x)$ is a sufficiently large natural number. By the Compactness Theorem there is a model $(\mathscr{M}', \sigma')$ of X, which by the countability of the language of arithmetic and the Löwenheim-Skolem theorem we may assume to be at most countable. $\mathscr{M}$ is a structure elementarily equivalent to $\mathbb{N}$. Since for $m \neq n$ the sentence $\neg(\mathbf{m} \equiv \mathbf{n})$ belongs to $\mathrm{Th}(\mathbb{N})$, $\mathscr{M}$ is infinite and hence is countable. $\mathscr{M}$ and

$\mathbb{N}$ are not isomorphic, since an isomorphism from $\mathbb{N}$ onto $\mathscr{M}$ would have to map the interpretation of $\mathbf{n}$ in the structure $\mathbb{N}$ (this turns out to be the number n) to the interpretation of $\mathbf{n}$ in the structure $\mathscr{M}$, and thus $\sigma(x)$ would not belong to the range of the isomorphism at all. ⊣

Considering the set $\mathrm{Th}(\mathbb{N}^<) \cup \{\neg(x \equiv \mathsf{0}), \neg(x \equiv \mathbf{1}), \neg(x \equiv \mathbf{2}), \ldots\}$, we obtain the following theorem.

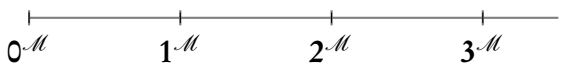
Theorem 4.35 *There is a countable nonstandard model of* $\mathrm{Th}(\mathbb{N}^<)$.

What do nonstandard models of $\mathrm{Th}(\mathbb{N})$ or $\mathrm{Th}(\mathbb{N}^<)$ look like? In the following we gain some insight into the order structure of a nonstandard model $\mathscr{M}$ of $\mathrm{Th}(\mathbb{N}^<)$.

In $\mathbb{N}^<$ the sentences

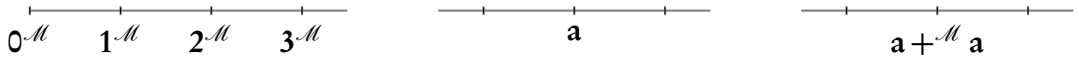
$$\forall x(\mathsf{0} \equiv x \vee \mathsf{0} < x),$$
$$\mathsf{0} < \mathbf{1} \wedge \forall x(\mathsf{0} < x \supset (\mathbf{1} \equiv x \vee \mathbf{1} < x)),$$
$$\mathbf{1} < \mathbf{2} \wedge \forall x(\mathbf{1} < x \supset (\mathbf{2} \equiv x \vee \mathbf{2} < x)), \ldots$$

hold. They say that $\mathsf{0}$ is the smallest element, 1 is the next smallest element after $\mathsf{0}$, 2 is the next smallest element after 1, and so on. Since these sentences also hold in $\mathscr{M}$, the "initial segment" of $\mathscr{M}$ looks as follows:



In addition, S (the underlying set of $\mathscr{M}$) contains a further element, say a, since otherwise $\mathscr{M}$ and $\mathbb{N}^<$ are isomorphic. Furthermore, $\mathbb{N}^<$ satisfies a sentence φ which says that for every element there is an immediate successor and for every element other than $\mathsf{0}$ there is an immediate predecessor. From this it follows easily that S contains, in addition to a, infinitely many other elements which together with a are ordered like the integers in $\mathscr{M}$:

If we consider the element $a + a$ we are led to further elements of S:



It is clear that $a + a$ lies in a different copy of $\mathbb{Z}$ than a. If they belonged to the same copy, then $a + a = a + n$ for some natural number n. By the cancellation law for addition, $a = n$, which is a contradiction. We can also show that between every two copies of $\mathbb{Z}^<$ in $\mathscr{M}$ there lies another. This is because $\mathbb{N}^<$ satisfies a sentence φ which says that for any two elements m and n, if $m < n$ there exists

a "midpoint" p (i.e. $m+n=2\cdot p$ or $m+n=2\cdot p+1$). The same statement is satisfied by $\mathscr{M}$ as well. If we now consider two elements a and b which lie in different copies of $\mathbb{Z}^<$ in $\mathscr{M}$, they have a midpoint c which has to lie in between a and b but cannot lie in either of their copies $\mathbb{Z}^<$ (since that would imply that a and b lie in the same copy). Thus any nonstandard model of arithmetic looks like the rational line (to the right of and including the point 0) with the point 0 replaced by a copy of $\mathbb{N}^<$ and every other point replaced by a copy of $\mathbb{Z}^<$.

4.10 An Algebraic Characterisation of Elementary Equivalence

In the previous sections we saw that the notion of elementary equivalence was weaker than the notion of isomorphism. This leads us to ask whether there is a purely algebraic notion (not referring to first-order formulas and the like) which is equivalent to elementary equivalence. This would be of much use, since we can now prove two structures elementarily equivalent through means other than showing that the two structures satisfy the same formulas. In this section, we provide such an algebraic characterisation (due to Fraisse) and give examples of its use.

Fraisse's theorem

In the following, we provide a simple proof of Fraisse's theorem. We assume that we are working with the signature of graphs, consisting of a single binary relation symbol R. It is easy to see that what we prove here can be generalised to all signatures containing only relational symbols. Later we will show how to extend the result to arbitrary signatures.

We introduce the following notation to simplify the presentation. We use $\overline{a}$ to denote a tuples of elements. $|\overline{a}|$ denotes the number of elements in the tuple. We also write $\varphi(\overline{x})$ (where $\overline{x}=x_1,\ldots,x_r$) to indicate the fact that $FV(\varphi)\subseteq\{x_1,\ldots,x_r\}$. For a structure $\mathscr{M}$, a tuple $\overline{a}$ of elements from $\mathscr{M}$, and a formula $\varphi(\overline{x})$ with $|\overline{x}|=|\overline{a}|$, we write $(\mathscr{M},\overline{a})\vDash\varphi(\overline{x})$ to mean that $(\mathscr{M},\sigma)\vDash\varphi$, with $\sigma(x_i)=a_i$ for all $i\leq|\overline{x}|$.

Definition 4.36 *Let $G=(V,E)$ and $H=(W,F)$ be two graphs. Let $\overline{a}$ and $\overline{b}$ be finite tuples of elements from V and W respectively, such that $|\overline{a}|=|\overline{b}|$. We say that $(G,\overline{a})$ and $(H,\overline{b})$ are m-equivalent—in symbols $(G,\overline{a})\equiv_m(H,\overline{b})$—if for every formula $\varphi(\overline{x})$ whose quantifier rank (the maximum nesting depth of quantifiers in the formula) is not more than m, $(G,\overline{a})\vDash\varphi(\overline{x})$ if and only if $(H,\overline{b})\vDash\varphi(\overline{x})$.*

Note that for the above definition to make sense, $|\overline{x}|$ should be equal to $|\overline{a}|$. But we will not crib about such minor details here and in what follows.

We now motivate the notion of *m-isomorphism*. The least we require is that any two m-isomorphic graphs are m-equivalent. Consider two graphs $G=(V,E)$ and $H=(W,F)$, $\overline{a}$ from V and $\overline{b}$ from W. Suppose that $(G,\overline{a})\not\equiv_m(H,\overline{b})$. Let us say that there is a formula $\varphi(\overline{x},y)$ with quantifier rank $\leq m-1$ such that $(G,\overline{a})\vDash\exists y\varphi(\overline{x},y)$ and $(H,\overline{b})\nvDash\exists y\varphi(\overline{x},y)$. This means that for some $c\in V$ and for all $d\in W$, $(G,\overline{a}c)\vDash\varphi(\overline{x},y)$ and $(H,\overline{b}d)\nvDash\varphi(\overline{x},y)$. Thus there is $c\in V$ such that for all $d\in W$, $(G,\overline{a}c)\not\equiv_{m-1}(H,\overline{b}d)$. In the symmetric case involving the universal quantifier, we infer that there is $d\in W$ such that for all $c\in V$, $(G,\overline{a}c)\not\equiv_{m-1}(H,\overline{b}d)$. We have proved the following

Lemma 4.37 *Suppose that for every $c \in V$ there is a $d \in W$ such that $(G,\overline{a}c) \equiv_{m-1} (H,\overline{b}d)$ and that for every $d \in W$ there is a $c \in V$ such that $(G,\overline{a}c) \equiv_{m-1} (H,\overline{b}d)$. Then $(G,\overline{a}) \equiv_m (H,\overline{b})$.*

This lemma leads to the following definition.

Definition 4.38 *Let $G = (V,E)$ and $H = (W,F)$ be two graphs, and let $\overline{a}$ and $\overline{b}$ be tuples of elements from V and W respectively. We say that $(G,\overline{a})$ is 0-isomorphic to $(H,\overline{b})$—in symbols, $(G,\overline{a}) \cong_0 (H,\overline{b})$—if $\overline{a} \longmapsto \overline{b}$ is a* partial isomorphism *from G to H (i.e. for any two $i,j \leq |\overline{a}|$, Ea_ia_j iff Fb_ib_j).*

For $m > 0$, we say that $(G,\overline{a}) \cong_m (H,\overline{b})$ if and only if

- *for all $c \in V$, there is a $d \in W$ such that $(G,\overline{a}c) \cong_{m-1} (H,\overline{b}d)$, and*
- *for all $d \in W$, there is a $c \in V$ such that $(G,\overline{a}c) \cong_{m-1} (H,\overline{b}d)$.*

It is easy to see that $(G,\overline{a}) \cong_0 (H,\overline{b})$ iff $(G,\overline{a}) \equiv_0 (H,\overline{b})$. This can be used as the base case in a proof by induction that for any m, if $(G,\overline{a}) \cong_m (H,\overline{b})$ then $(G,\overline{a}) \equiv_m (H,\overline{b})$. The induction step follows immediately from the above definition and the previous lemma.

Fraisse's theorem says that the other direction also holds. For proving that we need the following lemma.

Lemma 4.39 *There are only finitely many inequivalent formulas of quantifier depth $\leq m$ having at most k free variables.*

Proof: Let $C(m,k)$ denote the number of formulas of quantifier depth $\leq m$ having at most k free variables. (To be precise, $C(m,k)$ is the size of a maximal set of pairwise inequivalent formulas each of which is of quantifier depth $\leq m$ and has at most k free variables.) We prove by induction on m that for all k, $C(m,k)$ is finite.

For any k, there are exactly $p = 2 \cdot k^2$ *atomic* formulas, $x_i \equiv x_j$ and Rx_ix_j where $i,j \leq k$. Thus there are at most 2^{2^p} inequivalent quantifier-free formulas. Thus $C(0,k)$ is finite.

For the case where $m > 0$, we know by the induction hypothesis that $C(m-1,k)$ is finite for all k. A formula of quantifier depth $\leq m$ is a boolean combination of formulas of quantifier depth $\leq m-1$ and formulas of the form $\forall y\varphi(\overline{x},y)$ where φ is of quantifier depth $\leq m-1$. Thus $C(m,k) \leq 2^{2^{2 \cdot C(m-1,k+1)}}$ and is hence finite. ⊣

Theorem 4.40 *If $(G,\overline{a}) \equiv_m (H,\overline{b})$ then $(G,\overline{a}) \cong_m (H,\overline{b})$.*

Proof: When $m = 0$, the theorem is immediate, as has already been noted.

Suppose $m > 0$ and that $(G,\overline{a}) \ncong_m (H,\overline{b})$. Then one of the folowing two cases holds and in both cases we prove that $(G,\overline{a}) \not\equiv_m (H,\overline{b})$.

- There is $c \in V$ such that for all $d \in W$, $(G,\overline{a}c) \not\cong_{m-1} (H,\overline{b}d)$. By the induction hypothesis, $(G,\overline{a}c) \not\equiv_{m-1} (H,\overline{b}d)$. Thus for each $d \in W$, there is a formula $\varphi_d(\overline{x},y)$ of quantifier depth $\leq m-1$ such that $(G,\overline{a}c) \models \varphi_d(\overline{x},y)$ and $(H,\overline{b}d) \not\models \varphi_d(\overline{x},y)$. Since there are only finitely many φ_d's which are inequivalent their conjunction is equivalent to a formula $\psi(\overline{x},y)$ of quantifier depth $\leq m-1$. Now $(G,\overline{a}c) \models \psi(\overline{x},y)$ but for all $d \in W$, $(H,\overline{b}d) \not\models \psi(\overline{x},y)$. Therefore $(G,\overline{a}) \models \exists y\psi(\overline{x},y)$ but $(H,\overline{b}) \not\models \exists y\psi(\overline{x},y)$. This shows that $(G,\overline{a}) \not\equiv_m (H,\overline{b})$.

- There is $d \in W$ such that for all $c \in V$, $(G,\overline{a}c) \not\cong_{m-1} (H,\overline{b}d)$. By the induction hypothesis, $(G,\overline{a}c) \not\equiv_{m-1} (H,\overline{b}d)$. Thus for each $c \in V$, there is a formula $\varphi_c(\overline{x},y)$ of quantifier depth $\leq m-1$ such that $(G,\overline{a}c) \models \varphi_c(\overline{x},y)$ and $(H,\overline{b}d) \not\models \varphi_c(\overline{x},y)$. Since there are only finitely many φ_c's which are inequivalent their disjunction is equivalent to a graph formula $\psi(\overline{x},y)$ of quantifier depth $\leq m-1$. Now for all $c \in V$, $(G,\overline{a}c) \models \psi(\overline{x},y)$ but $(H,\overline{b}d) \not\models \psi(\overline{x},y)$. Therefore $(G,\overline{a}) \models \forall y\psi(\overline{x},y)$ but $(H,\overline{b}) \not\models \forall y\psi(\overline{x},y)$. This shows that $(G,\overline{a}) \not\equiv_m (H,\overline{b})$. $\dashv$

We say that $(G,\overline{a})$ is *finitely isomorphic* to $(H,\overline{b})$ (in symbols $(G,\overline{a}) \cong_f (H,\overline{b})$) iff $(G,\overline{a})$ is m-isomorphic to $(H,\overline{b})$ for all $m \geq 0$. From the definitions and the previous theorem, the following immediately follows, giving us the required algebraic characterisation of elementary equivalence.

Theorem 4.41 (Fraisse's theorem) *For any two graphs G and H, and tuples $\overline{a}$ and $\overline{b}$ of the same length from G and H respectively,*

$$(G,\overline{a}) \cong_f (H,\overline{b}) \text{ iff } (G,\overline{a}) \equiv (H,\overline{b}).$$

Extending the theorem to arbitrary signatures

It is clear that the definitions and proofs in the above section extend to arbitrary (finite) relational signatures (signatures containing only relation symbols) almost verbatim. The definition of partial isomorphism needs to be extended, but that is fairly straightforward. Note also that for any relational signature, there are only finitely many inequivalent formulas with k free variables and quantifier depth $\leq m$. This property does not hold for signatures containing function symbols.

Let L be an arbitrary (finite) signature. For every n-ary function symbol f occurring in L, define a new $(n+1)$-ary relation symbol F and, for each constant symbol c occurring in L, define a new unary relation C. Let L^r consist of the relation symbols from L together with the new relation symbols. L^r is relational. For an L-structure $\mathcal{M}$, let $\mathcal{M}^r$ be the L^r structure obtained from $\mathcal{M}$, with the following interpretation for the new relation symbols: $F^{\mathcal{M}^r}(a_1,\ldots,a_n,a)$ iff $f^{\mathcal{M}}(a_1,\ldots,a_n) = a$, and $C^{\mathcal{M}^r}(a)$ iff $c^{\mathcal{M}} = a$. One can systematically construct an L^r-formula φ^r for every L-formula φ such that $\mathcal{M} \models \varphi$ iff $\mathcal{M}^r \models \varphi^r$. For example, if φ is the formula $f(f(g(c,x))) = y$, then φ^r is the formula $\exists z_1 z_2 z_3\,[C(z_1) \wedge G(z_1,x,z_2) \wedge F(z_2,z_3) \wedge F(z_3,y)]$. We leave it as an exercise to the reader to formally state and prove the result. From the above considerations, it follows that $(\mathcal{M},\overline{a}) \equiv (\mathcal{M}',\overline{b})$ iff $(\mathcal{M}^r,\overline{a}) \equiv ((\mathcal{M}')^r,\overline{b})$. (But note that it is *not* the case that $(\mathcal{M},\overline{a}) \equiv_m (\mathcal{M}',\overline{b})$ iff $(\mathcal{M}^r,\overline{a}) \equiv_m ((\mathcal{M}')^r,\overline{b})$ for all m.)

We also need to extend the definition of partial isomorphism to arbitrary signatures. Here it is.

Definition 4.42 *Let $\mathscr{M} = (S, \iota)$ and $\mathscr{M}' = (S', \iota')$ be two L-structures and let p be a partial function from S to S'. We call p a partial isomorphism iff:*

- *p is injective.*
- *p is a homomorphism in the following sense:*
  - *For n-ary relation symbols P in L and $a_1, \ldots, a_n \in \mathrm{dom}(p)$,*

$$P^{\mathscr{M}}(a_1, \ldots, a_n) \text{ iff } P^{\mathscr{M}'}(p(a_1), \ldots, p(a_n)).$$

  - *For n-ary function symbols f in L and $a_1, \ldots, a_n, a \in \mathrm{dom}(p)$,*

$$f^{\mathscr{M}}(a_1, \ldots, a_n) = a \text{ iff } f^{\mathscr{M}'}(p(a_1), \ldots, p(a_n)) = p(a).$$

  - *For constant symbols c in L and $a \in \mathrm{dom}(p)$,*

$$c^{\mathscr{M}} = a \text{ iff } c^{\mathscr{M}'} = p(a).$$

From the above definition it is clear that a given p is a partial isomorphism from $\mathscr{M}$ to $\mathscr{M}'$ iff it is a partial isomorphism from $\mathscr{M}^r$ to $(\mathscr{M}')^r$. Thus it follows that $\mathscr{M} \cong_m \mathscr{M}'$ iff $\mathscr{M}^r \cong_m (\mathscr{M}')^r$, for any given m. We can now easily prove Fraisse's theorem for arbitrary finite signatures. $\mathscr{M} \cong_f \mathscr{M}'$ iff $\mathscr{M}^r \cong_f (\mathscr{M}')^r$ iff $\mathscr{M}^r \equiv (\mathscr{M}')^r$ iff $\mathscr{M} \equiv \mathscr{M}'$.

Examples

We give two examples in this section, which illustrate the use of the easier half of Fraisse's theorem.

***Example** 4.43* Suppose $L = (s, 0)$ where s is a unary "successor" function symbol and 0 is a constant. Let X consist of the "successor axioms":

- $\forall x(\neg(x \equiv 0) \equiv \exists y(s(y) \equiv x))$,
- $\forall x \forall y((s(x) \equiv s(y)) \supset (x \equiv y))$, and
- for every $m \geq 1 : \forall x \neg(s^m(x) \equiv x)$. ($s^0(x) \stackrel{\text{def}}{=} x$ and for all $m \geq 0$, $s^{m+1}(x) = s(s^m(x))$.)

The natural numbers with the usual successor function is a model of X. We want to prove that any two models of X are elementarily equivalent. Towards that end we prove that any two models of X are finitely isomorphic. If any two models of X are elementarily equivalent, then for any sentence φ, either all models of X satisfy φ or all models of X satisfy $\neg\varphi$. Thus for any sentence φ, $X \vDash \varphi$ or $X \vDash \neg\varphi$. Thus X is an example of a so-called *complete theory*, a theory which can decide any statement one way or the other. A further point to note is that X is a recursive set of sentences, and so forms the basis of a procedure to decide the truth or falsity of any L-sentence φ in the structure $\mathbb{N}$. Simply enumerate longer and longer proofs which use formulas from X as additional axioms, apart from the

standard axioms and rules. Since either $X \vdash \varphi$ or $X \vdash \neg\varphi$, eventually a proof of φ or $\neg\varphi$ will turn up. Halt and announce the result at that point.

Let us return to our present concern, which is that of proving any two models of X finitely isomorphic. First, we fix the following notation: For a model $\mathcal{M} = (S, \iota)$ of X and $a \in S$ we set $a^{(m)} \stackrel{\text{def}}{=} f^m(a)$, where $f = s^{\mathcal{M}}$. For every $n \in \mathbb{N}$ we define a "distance function" d_n on $S \times S$ by

$$d_n(a,a') \stackrel{\text{def}}{=} \begin{cases} m & \text{if } a^{(m)} = a' \text{ and } m \leq 2^n \\ -m & \text{if } (a')^{(m)} = a \text{ and } m \leq 2^n \\ \infty & \text{otherwise.} \end{cases}$$

Now suppose $\mathcal{M} = (S, \iota)$ and $\mathcal{M}' = (S', \iota')$ are two models of X. For notational simplicity we will assume that every tuple $\overline{a}$ we mention below contains $0^{\mathcal{M}}$ as the first element and every tuple $\overline{b}$ contains $0^{\mathcal{M}'}$ as the first element. Let $\overline{a}$ and $\overline{b}$ be tuples of elements from $\mathcal{M}$ and $\mathcal{M}'$ respectively, both having the same number of elements. We say that $(\mathcal{M}, \overline{a})$ and $(\mathcal{M}', \overline{b})$ are "d_n-equivalent" if $(\mathcal{M}, \overline{a}) \cong_0 (\mathcal{M}', \overline{b})$ and for all $i, j \leq |\overline{a}|$, $d_n(a_i, a_j) = d_n(b_i, b_j)$. We wish to prove that whenever $(\mathcal{M}, \overline{a})$ and $(\mathcal{M}', \overline{b})$ are d_n-equivalent, $(\mathcal{M}, \overline{a}) \cong_n (\mathcal{M}', \overline{b})$. The base case is quite easy, since whenever $(\mathcal{M}, \overline{a})$ and $(\mathcal{M}, \overline{b})$ are d_0-equivalent, $a_0 = 0^{\mathcal{M}}$, $b_0 = 0^{\mathcal{M}'}$, and $(\mathcal{M}, \overline{a}) \cong_0 (\mathcal{M}', \overline{b})$ by definition. Suppose $(\mathcal{M}, \overline{a})$ and $(\mathcal{M}', \overline{b})$ are d_{n+1}-equivalent. Consider an arbitrary $c \in S$. Now it could be the case that for some $i \leq |\overline{a}|$, $|d_n(a_i, c)| \leq 2^n$. If that is so, choose $d \in S'$ with $d_n(b_i, d) = d_n(a_i, c)$. It is easy to check that $(\mathcal{M}, \overline{a}c)$ and $(\mathcal{M}', \overline{b}d)$ are d_n-equivalent. If $|d_n(a_i, c)| > 2^n$ for all $i \leq |\overline{a}|$ then choose $d \in S'$ such that $|d_n(b_i, d)| > 2^n$ for all i (such an element d must exist since every model of X is infinite!). Now again it is easy to see that $(\mathcal{M}, \overline{a}c)$ and $(\mathcal{M}', \overline{b}d)$ are d_n-equivalent. But by the induction hypothesis (on n) this means that for all $c \in S$ there exists a $d \in S'$ such that $(\mathcal{M}, \overline{a}c) \cong_n (\mathcal{M}', \overline{b}d)$. By symmetric reasoning, we can show that for any $d \in S'$, there exists a c such that $(\mathcal{M}, \overline{a}c) \cong_n (\mathcal{M}', \overline{b}d)$. These two facts imply, by definition, that $(\mathcal{M}, \overline{a}) \cong_{n+1} (\mathcal{M}', \overline{b}d)$.

In earlier sections, we showed that some classes of structures are not Δ-elementary. The arguments involved the Compactness Theorem and used infinite structures. With the techniques at our disposal now, we can show that certain properties cannot be expressed by a first-order sentence, even if we restrict ourselves to *finite* structures. We illustrate this approach by the following example.

Theorem 4.44 *Let L be the language of graphs. There is no L-sentence whose finite models are the finite connected graphs. (Hence, in particular, the class of connected graphs is not elementary.)*

Proof:
For $k \geq 0$ let $G_k = (V_k, E_k)$ be the graph corresponding to the regular $(k+1)$-gon, where

$$V_k = \{0, \ldots, k\}$$

and

$$E_k = \{(i, i+1) \mid i < k\} \cup \{(i, i-1) \mid 1 \leq i \leq k\} \cup \{(0, k), (k, 0)\},$$

and let $H_k = (W_k, F_k)$ consist of two disjoint copies of G_k, say,

$$W_k = \{0, \ldots, k\} \times \{0, 1\}$$

and

$$F_k = \{((i,0),(j,0)) \mid (i,j) \in E_k\} \cup \{((i,1),(j,1)) \mid (i,j) \in E_k\}.$$

We claim that:

$$\text{For all } k \geq 2^m : G_k \cong_m H_k.$$

Then we are done. In fact, let φ be an L-sentence and m be the quantifier rank of φ. Then we have that $G_{2^m} \cong_m H_{2^m}$, i.e. $G_{2^m} \equiv_m H_{2^m}$ and therefore $G_{2^m} \models \varphi$ iff $H_{2^m} \models \varphi$. Since G_{2^m} is connected, but H_{2^m} is not, the class of finite models of φ cannot be identical with the class of all finite connected graphs.

For proving that for all $k \geq 2^m : G_k \cong_m H_k$, we proceed as follows. For fixed $k \geq 2^m$ and $n \geq 0$, we define "distance functions" d on $V_k \times V_k$ and d' on $W_k \times W_k$, as follows:

$$d(a,b) \stackrel{\text{def}}{=} \begin{cases} \text{length of the shortest path connecting } a \text{ and } b \text{ in } G_k, & \text{if this length is} \leq 2^m; \\ \infty, & \text{otherwise;} \end{cases}$$

$$d'((a,i),(b,j)) \stackrel{\text{def}}{=} \begin{cases} d(a,b) & \text{if } i = j; \\ \infty & \text{otherwise.} \end{cases}$$

We say that $(G_k, \overline{a})$ and $(H_k, \overline{b})$ are (d,d')-equivalent iff for all $i, j \leq |\overline{a}|$, $d(a_i, a_j) = d'(b_i, b_j)$. Just like in the previous example, we can prove that whenever $(G_k, \overline{a})$ and $(H_k, \overline{b})$ are (d,d')-equivalent, $(G_k, \overline{a}) \cong_m (H_k, \overline{b})$. ⊣

Ehrenfeucht Games

The algebraic description of elementary equivalence is well-suited for many purposes. However, it lacks the intuitive appeal of a game-theoretical characterisation due to Ehrenfeucht, which we look at in the present section.

Let L be an arbitrary signature and let $\mathcal{M} = (S, \iota)$ and $\mathcal{M}' = (S', \iota')$ be L-structures. To simplify the formulation we assume $S \cap S' = \emptyset$. The *Ehrenfeucht game* $\mathcal{G}(\mathcal{M}, \mathcal{M}')$ corresponding to $\mathcal{M}$ and $\mathcal{M}'$ is played by two players, *Spoiler* and *Duplicator*, according to the following rules:

Each *play* of the game begins with *Spoiler* choosing a natural number $r \geq 1$; r is the number of subsequent moves each player has to make in the course of the play. These subsequent moves are begun by the *Spoiler*, and both players move alternately. Each move consists of choosing an element from $S \cup S'$. If *Spoiler* chooses an element $a_i \in S$ in his i-th move, then *Duplicator* must choose an element $b_i \in S'$ in his i-th move. If *Spoiler* chooses an element $b_i \in S'$ in his i-th move, then *Duplicator* must choose an element $a_i \in S$ in his i-th move. After the r-th move of *Duplicator* the play is completed. Altogether some number $r \geq 1$, elements $a_1, \ldots, a_r \in S$ and $b_1, \ldots, b_r \in S'$ have been chosen. *Duplicator* has won the play iff $(\mathcal{M}, \overline{a}) \cong_0 (\mathcal{M}', \overline{b})$.

We say that *Duplicator* has a winning strategy in $\mathcal{G}(\mathcal{M},\mathcal{M}')$ and write "*Duplicator* wins $\mathcal{G}(\mathcal{M},\mathcal{M}')$" if it is possible for him to win each play. (Following Ebbinghaus, Flum, and Thomas, we omit an exact definition of the notion of "winning strategy".)

Lemma 4.45 *$\mathcal{M} \cong_f \mathcal{M}'$ iff Duplicator wins $\mathcal{G}(\mathcal{M},\mathcal{M}')$.*

Proof: We prove a more general statement: $(\mathcal{M},\overline{a}) \cong_f (\mathcal{M}',\overline{b})$ iff *Duplicator* wins $\mathcal{G}(\mathcal{M},\overline{a},\mathcal{M}',\overline{b})$.

Suppose $(\mathcal{M},\overline{a}) \cong_f (\mathcal{M}',\overline{b})$. We describe a winning strategy for *Duplicator*:

If *Spoiler* chooses the number r at the beginning of a $\mathcal{G}(\mathcal{M},\overline{a},\mathcal{M}',\overline{b})$-play, then for $i = 1,\ldots,r$ *Duplicator* should choose the elements $c_i \in S$ (or respectively $d_i \in S'$) so as to maintain $(\mathcal{M},\overline{a}c_1\ldots c_i) \cong_{r-i} (\mathcal{M}',\overline{b}d_1\cdots d_i)$. That we can always do this follows from the fact that $(\mathcal{M},\overline{a}) \cong_r (\mathcal{M}',\overline{b})$. For $i = r$ it follows that *Duplicator* has a winning strategy for the game.

Conversely, suppose $(\mathcal{M},\overline{a}) \ncong_f (\mathcal{M}',\overline{b})$. Then we give a winning strategy in r moves for *Spoiler* in $\mathcal{G}(\mathcal{M},\mathcal{M}')$. If $(\mathcal{M},\overline{a}) \ncong_0 (\mathcal{M}',\overline{b})$ then it is immediate that *Spoiler* wins all plays in $\mathcal{G}(\mathcal{M},\overline{a},\mathcal{M}',\overline{b})$, even the play with no moves. Suppose $(\mathcal{M},\overline{a}) \ncong_r (\mathcal{M}',\overline{b})$. Then *Spoiler* chooses r at the beginning of the game. Now it is clear that either there is a $c \in S$ such that for all $d \in S'$, $(\mathcal{M},\overline{a}c) \ncong_{r-1} (\mathcal{M}',\overline{b}d)$, or there is a $d \in S'$ such that for all $c \in S$, $(\mathcal{M},\overline{a}c) \ncong_{r-1} (\mathcal{M}',\overline{b}d)$. Suppose the former. Then the *Spoiler* chooses the element c such that for all $d \in S'$, $(\mathcal{M},\overline{a}c) \ncong_{r-1} (\mathcal{M}',\overline{b}d)$. From this and the induction hypothesis it follows that no matter what d *Duplicator* plays, *Spoiler* has a winning strategy in $r-1$ moves in $\mathcal{G}(\mathcal{M},\overline{a}c,\mathcal{M}',\overline{b}d)$. Similarly in the case where there is a $d \in S'$ such that for all $c \in S$, $(\mathcal{M},\overline{a}c) \ncong_{r-1} (\mathcal{M}',\overline{b}d)$. Thus *Spoiler* has a winning strategy in r moves in the original game. ⊣

The above lemma and Fraisse's theorem together yield the following:

Theorem 4.46 (Ehrenfeucht's Theorem) *Let L be a finite signature. Then for any L-structures $\mathcal{M}$ and $\mathcal{M}'$:*

$$\mathcal{M} \equiv \mathcal{M}' \text{ iff Duplicator wins } \mathcal{G}(\mathcal{M},\mathcal{M}').$$

4.11 Decidability

We consider in this section the *satisfiability problem* for first-order logic. This is the problem of determining whether a given first-order formula is satisfiable. We saw earlier that the corresponding problem for propositional logic, many modal logics, and dynamic logic is decidable. In contrast, the problem is undecidable for first-order logic. We present a particularly simple proof of this result here. Our undecidability proof proceeds by reducing the reachability problem for *two-counter machines* to the satisfiability problem.

A *two-counter machine* is a finite-state automaton equipped with two counters which can contain arbitrary natural numbers. Formally it is a tuple $M = (Q, q_0, \Delta, F)$ where:

- Q is a finite set of states,

- $q_0 \in Q$ is the initial state,
- $F \subseteq Q$ is the set of final states, and
- $\Delta \subseteq Q \times \{0,1\}^2 \times Q \times \{-1,0,1\}^2$ is the transition relation satisfying the following condition:
  - for all $(q, z_1, z_2, q', \delta_1, \delta_2) \in \Delta$ and $i \in \{1,2\}$, if $\delta_i = -1$ then $z_i = 1$.

In a transition $(q, z_1, z_2, q', \delta_1, \delta_2)$, for $i \in \{1,2\}$, $z_i = 0$ denotes the fact that the value of the i-th counter is zero, and $z_i = 1$ denotes the fact that the value of the i-th counter is nonzero. δ_i specifies the value to be added to the i-th counter. The condition on transitions reflect the fact that we can decrement only positive counters.

A configuration of a two-counter machine M is a triple $(q, m_1, m_2) \in Q \times \mathbb{N} \times \mathbb{N}$. For a transition $t = (r, z_1, z_2, r', \delta_1, \delta_2)$ and configurations (q, m_1, m_2) and (q', m_1', m_2'), $(q, m_1, m_2) \xrightarrow{t} (q', m_1', m_2')$ exactly when $q = r$, $q' = r'$, and for $i \in \{1,2\}$, (i) $z_i = 0$ iff $m_i = 0$ and (ii) $m_i' = m_i + \delta_i$.

We say that $(q, m_1, m_2) \xrightarrow{*} (q', m_1', m_2')$ iff there is a sequence of transitions leading from (q, m_1, m_2) to (q', m_1', m_2'). The configuration $(q_0, 0, 0)$ is called the *initial configuration*. (q, m_1, m_2) is called a *final configuration* if $q \in F$. The reachability problem for two-counter machines is the problem of determining whether a final configuration is reachable from the initial configuration. We assume that the reader is familiar with the fact that this problem is undecidable.

Given a two-counter machine M one can define a first-order language L_M and a first-order formula φ_M over L_M such that a final configuration is reachable in M iff φ_M is valid. It is easy to see now that the satisfiability problem for first-order logic is undecidable. Suppose, on the contrary, that it is decidable. Then we could decide the reachability problem for two-counter machines as follows: given any two-counter machine M, construct φ_M and declare a final configuration to be reachable exactly when $\neg\varphi_M$ is not satisfiable.

The reduction

Let $M = (Q, q_0, \Delta, F)$ be a given two-counter machine. Then L_M is defined to be (C_M, F_M, R_M) where:

- $C_M = \{\mathbf{q} \mid q \in Q\} \cup \{\mathbf{o}\}$,
- $F_M = \{\mathbf{s}\}$ with $\#(\mathbf{s}) = 1$, and
- $R_M = \{\mathbf{conf}\}$ with $\#(\mathbf{conf}) = 3$.

For each $t = (q, z_1, z_2, q', \delta_1, \delta_2) \in \Delta$ we define a formula φ_t. Rather than giving the most general definition, we show the construction for two representative examples:
Let $t = (q, 0, 1, q', 1, -1)$. Then φ_t is the following formula:

$$\forall x\ [(\mathbf{conf}(\mathbf{q}, \mathbf{o}, x) \wedge \exists y\ (x = \mathbf{s}(y))) \supset \mathbf{conf}(\mathbf{q}', \mathbf{s}(\mathbf{o}), y)].$$

Let $t = (q, 1, 1, q', 0, 1)$. Then φ_t is the following formula:

$$\forall x\ y\ [(\mathbf{conf}(\mathbf{q}, x, y) \wedge \exists x'\ y'\ (x = \mathbf{s}(x') \wedge y = \mathbf{s}(y'))) \supset \mathbf{conf}(\mathbf{q}', x, \mathbf{s}(y))].$$

Now we define the following sequence of formulas:

- $\mathbf{init} \stackrel{\text{def}}{=} \mathbf{conf}(\mathbf{q_0}, \mathbf{o}, \mathbf{o})$.
- $\mathbf{final} \stackrel{\text{def}}{=} \exists x\ \exists y \bigwedge_{q \in F} \mathbf{conf}(\mathbf{q}, x, y)$.
- $\varphi_\Delta \stackrel{\text{def}}{=} \bigvee_{t \in \Delta} \varphi_t$.
- $\varphi_M \stackrel{\text{def}}{=} (\varphi_\Delta \wedge \mathbf{init}) \supset \mathbf{final}$.

The following two lemmas prove that the reduction is correct. (We use $\mathbf{m}$ as an abbreviation for $\mathbf{s}^m(\mathbf{o})$ in the formulas, in what follows.)

Lemma 4.47 *For every configuration (q, m_1, m_2) of M,*

$$(q_0, 0, 0) \xrightarrow{*} (q, m_1, m_2) \Longrightarrow \models (\varphi_\Delta \wedge \mathit{init}) \supset \mathit{conf}(\mathbf{q}, \mathbf{m_1}, \mathbf{m_2}).$$

In particular, if a final configuration is reachable in M then φ_M is valid.

Proof: We prove that whenever $(q_0, 0, 0) \xrightarrow{*} (q, m_1, m_2)$, it is also the case that $\models (\varphi_\Delta \wedge \mathbf{init}) \supset \mathbf{conf}(\mathbf{q}, \mathbf{m_1}, \mathbf{m_2})$. We do this by induction on the number of steps it takes to reach (q, m_1, m_2).

Basis: The only configuration reachable in zero steps is $(q_0, 0, 0)$ itself, and sure enough, $\models (\varphi_\Delta \wedge \mathbf{init}) \supset \mathbf{conf}(\mathbf{q_0}, \mathbf{o}, \mathbf{o})$.

Induction step: Suppose $(q_0, 0, 0) \xrightarrow{*} (q', m_1', m_2') \xrightarrow{t} (q, m_1, m_2)$ and $\models (\varphi_\Delta \wedge \mathbf{init}) \supset \mathbf{conf}(\mathbf{q'}, \mathbf{m_1'}, \mathbf{m_2'})$. Now it is an easy exercise to check that $\models (\mathbf{conf}(\mathbf{q'}, \mathbf{m_1'}, \mathbf{m_2'}) \wedge \varphi_t) \supset \mathbf{conf}(\mathbf{q}, \mathbf{m_1}, \mathbf{m_2})$. It follows that $\models (\varphi_\Delta \wedge \mathbf{init}) \supset \mathbf{conf}(\mathbf{q}, \mathbf{m_1}, \mathbf{m_2})$, as desired. ⊣

Lemma 4.48 *If φ_M is valid, then a final configuration is reachable in M.*

Proof: We prove the desired statement in the contrapositive form. Suppose no final state is reachable from the initial configuration. Now we define an L_M structure $\mathscr{M} = (S, \iota)$ as follows: $S = \mathbb{N}$; for each $q \in Q$, $\iota(\mathbf{q})$ is an arbitrary distinct natural number; $\iota(\mathbf{o}) = 0$; $\iota(\mathbf{s})$ is the successor function on $\mathbb{N}$; and $\iota(\mathbf{conf}) \stackrel{\text{def}}{=} \{(\iota(\mathbf{q'}), m_1', m_2') \mid (q_0, 0, 0) \xrightarrow{*} (q', m_1', m_2')\}$. It is again an easy exercise to check that $\mathscr{M} \models \varphi_t$ for all $t \in \Delta$, and of course, $\mathscr{M} \models \mathbf{init} \wedge \neg \mathbf{final}$. Thus we see that $\mathscr{M} \not\models (\varphi_\Delta \wedge \mathbf{init}) \supset \mathbf{final}$. It follows that φ_M is not valid.

Thus we see that if φ_M is valid, then a final configuration is reachable in M. ⊣

The above two lemmas, in conjunction with the fact that the reachability problem for two-counter machines is undecidable, immediately yields the following theorem.

Theorem 4.49 *The satisfiability problem (as also the validity problem) for first-order logic is undecidable.*

The above reduction uses a language with a unary function symbol, a ternary relation symbol, and some constants. Using some coding tricks, we can get by with using just the ternary relation symbol and constants. Working out the minimal expressive power which leads to undecidability is an interesting problem, which has generated a lot of research over the years. In fact, there are books solely devoted to the study of the status of decidability of various fragments of first-order logic.