Boxes and Diamonds

An Open Introduction to Modal Logic

F19

Boxes and Diamonds

The Open Logic Project

Instigator

Richard Zach, *University of Calgary*

Editorial Board

Aldo Antonelli,† *University of California, Davis*
Andrew Arana, *Université Paris I Panthénon–Sorbonne*
Jeremy Avigad, *Carnegie Mellon University*
Tim Button, *University College London*
Walter Dean, *University of Warwick*
Gillian Russell, *University of North Carolina*
Nicole Wyatt, *University of Calgary*
Audrey Yap, *University of Victoria*

Contributors

Samara Burns, *University of Calgary*
Dana Hägg, *University of Calgary*
Zesen Qian, *Carnegie Mellon University*

Boxes and Diamonds

An Open Introduction to Modal Logic

Remixed by Richard Zach

Fall 2019

The Open Logic Project would like to acknowledge the generous support of the Taylor Institute of Teaching and Learning of the University of Calgary, and the Alberta Open Educational Resources (ABOER) Initiative, which is made possible through an investment from the Alberta government.

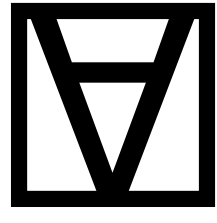
UNIVERSITY OF CALGARY
Taylor Institute for Teaching and Learning

Cover illustrations by Matthew Leadbeater, used under a Creative Commons Attribution-NonCommercial 4.0 International License.

Typeset in Baskervald X and Nimbus Sans by LaTeX.

This version of *Boxes and Diamonds* is revision cfof83d (2020-01-14), with content generated from *Open Logic Text* revision 41f40f7 (2020-01-14). Free download at:

https://github.com/rzach/boxes-and-diamonds



Contents

Preface

This is an introductory textbook on modal logic. I use it as the main text when I teach Philosophy 579.2 (Modal Logic) at the University of Calgary. It is based on material from the Open Logic Project.

The main text assumes familiarity with some elementary set theory and the basics of (propositional) logic. This material is part of a prerequisite for my course, Logic I. The textbook for that course, *Sets, Logic, Computation*, is also based on the OLP, and so is available for free. The required material is included as appendices in this book, however. I assign these appendices for background reading whenever.

Part I is originally based in part on Aldo Antonelli's lecture notes on "Classical Correspondence Theory for Basic Modal Logic," which he contributed to the OLP before his untimely death in 2015. I heavily revised and expanded these notes, e.g., the material on frame definability and tableaux is new.

Introduction

Modal logics are extensions of classical logic by the operators $\Box$ ("box") and $\Diamond$ ("diamond"), which attach to formulas. Intuitively, $\Box$ may be read as "necessarily" and $\Diamond$ as "possibly," so $\Box p$ is "p is necessarily true" and $\Diamond p$ is "p is possibly true." As necessity and possibility are fundamental metaphysical notions, modal logic is obviously of great philosophical interest. It allows the formalization of metaphysical principles such as "$\Box p \to p$" (if p is necessary, it is true) or "$\Diamond p \to \Box \Diamond p$" (if p is possible, it is necessarily possible).

The operators $\Box$ and $\Diamond$ are *intensional*. This means that whether $\Box A$ or $\Diamond A$ holds does not just depend on whether A holds or doesn't. An operator which is not intensional is *extensional*. Negation is extensional: $\neg A$ holds iff A does not; so whether $\neg A$ holds only depends on whether A holds or doesn't. $\Box$ and $\Diamond$ are not like that: whether $\Box A$ or $\Diamond A$ holds depends also on the meaning of A. While ordinary truth-functional semantics is enough to deal with extensional operators, intensional operators like $\Box$ and $\Diamond$ require a different kind of semantics. One such semantics which takes center stage in this book is relational semantics (also called possible-worlds semantics or Kripke semantics).

For the logic which corresponds to the interpretation of $\Box$ as "necessarily," this semantics is relatively simple: instead of assigning truth values to propositional variables, an interpretation $\boldsymbol{M}$ assigns a set of "worlds" to them—intuitively, those worlds w at which p is interpreted as true. On the basis of such an interpre-

tation, we can define a satisfaction relation. The definition of this satisfaction relation for formulas $\Box A$ make A satisfied at a world w iff A is satisfied at *all* worlds: $\boldsymbol{M}, w \Vdash \Box A$ iff $\boldsymbol{M}, v \Vdash A$ for all worlds v. This corresponds to Leibniz's idea that what's necessarily true is what's true in every possible world.

"Necessarily" is not the only way to interpret the $\Box$ operator, but it is the standard one—"necessarily" and "possibly" are the so-called *alethic* modalities. Other interpretations read $\Box$ as "it is known (by some person A) that," as "some person A believes that," "it ought to be the case that," or "it will always be true that." These are epistemic, doxastic, deontic, and temporal modalities, respectively. Different interpretations of $\Box$ will make different formulas as logically true, and different inferences as valid. For instance, everything necessary and everything known is true, so $\Box A \rightarrow A$ is a truth on the alethic and epistemic interpretations. By contrast, not everything believed nor everything that ought to be the case actually is the case, so $\Box A \rightarrow A$ is not a truth on the doxastic or deontic interpretations.

In order to deal with different interpretations of the modal operators, the semantics is extended by a relation between worlds, the so-called accessibility relation. Then $\boldsymbol{M}, w \Vdash A$ if $\boldsymbol{M}, v \Vdash A$ for all worlds v which are accessible from w. The resulting semantics is very versatile and powerful, and the basic idea can be used to provide semantic interpretations for logics based on other intensional operators. One such logic is intuitionistic logic, a constructive logic based on L. E. J. Brouwer's branch of constructive mathematics. Intuitionistic logic is philosophically interesting for this reason—it plays an important role in constructive accounts of mathematics—but was also propsed as a logic superior to classical logic by the influential English philosopher Michael Dummett in the 20th century. Another application of relational models is as a semantics for subjunctive, or counterfactual, conditionals, an approach pioneered by Robert Stalnaker and David K. Lewis.

This book is an introduction to the syntax, semantics, and proof theory of intensional logics. It only deals with propositional logics, although future editions will also treat predicate log-

ics. The material is divided into three parts: The first part deals with normal modal logics. These are logics with the operators $\Box$ and $\Diamond$. We discuss their syntax, relational models and semantic notions based on them (such as validity and consequence) and proof systems (both axiomatic systems and tableaux). We establish some basic results about these logics, such as the soundness and completeness of the proof systems considered, and discuss some model-theoretic constructions such as filtrations. The second part deals with intuitionistic logic. Here we discuss natural deduction and axiomatic derivations, relational and topological semantics, and soundness and completeness of the proof systems. The third part deals with the Lewis-Stalnaker semantics of counterfactual conditionals. The appendix discusses some ideas and results from set theory and the theory of relations that's crucial to the relational semantics, as well as reviews syntax, semantics, and proof theory of classical propositional logic.

PART I

Normal Modal Logics

CHAPTER 1

Syntax and Semantics

1.1 Introduction

Modal Logic deals with *modal propositions* and the entailment relations among them. Examples of modal propositions are the following:

1. It is necessary that $2 + 2 = 4$.
2. It is necessarily possible that it will rain tomorrow.
3. If it is necessarily possible that A then it is possible that A.

Possibility and necessity are not the only modalities: other unary connectives are also classified as modalities, for instance, "it ought to be the case that A," "It will be the case that A," "Dana knows that A," or "Dana believes that A."

Modal logic makes its first appearance in Aristotle's *De Interpretatione*: he was the first to notice that necessity implies possibility, but not vice versa; that possibility and necessity are interdefinable; that If $A \wedge B$ is possibly true then A is possibly true and B is possibly true, but not conversely; and that if $A \rightarrow B$ is necessary, then if A is necessary, so is B.

The first modern approach to modal logic was the work of C. I. Lewis, culminating with Lewis and Langford, *Symbolic Logic* (1932). Lewis & Langford were unhappy with the representation of implication by means of the material conditional: $A \to B$ is a poor substitute for "A implies B." Instead, they proposed to characterize implication as "Necessarily, if A then B," symbolized as $A \dashv B$. In trying to sort out the different properties, Lewis indentified five different modal systems, **S1**, ..., **S4**, **S5**, the last two of which are still in use.

The approach of Lewis and Langford was purely *syntactical*: they identified reasonable axioms and rules and investigated what was provable with those means. A semantic approach remained elusive for a long time, until a first attempt was made by Rudolf Carnap in *Meaning and Necessity* (1947) using the notion of a *state description*, i.e., a collection of atomic sentences (those that are "true" in that state description). After lifting the truth definition to arbitrary sentences A, Carnap defines A to be *necessarily true* if it is true in all state descriptions. Carnap's approach could not handle *iterated* modalities, in that sentences of the form "Possibly necessarily ... possibly A" always reduce to the innermost modality.

The major breakthrough in modal semantics came with Saul Kripke's article "A Completeness Theorem in Modal Logic" (JSL 1959). Kripke based his work on Leibniz's idea that a statement is necessarily true if it is true "at all possible worlds." This idea, though, suffers from the same drawbacks as Carnap's, in that the truth of statement at a world w (or a state description s) does not depend on w at all. So Kripke assumed that worlds are related by an *accessibility relation* R, and that a statement of the form "Necessarily A" is true at a world w if and only if A is true at all worlds w' *accessible from* w. Semantics that provide some version of this approach are called Kripke semantics and made possible the tumultuous development of modal logics (in the plural).

When interpreted by the Kripke semantics, modal logic shows us what *relational structures* look like "from the inside." A relational structure is just a set equipped with a binary relation (for

instance, the set of students in the class ordered by their social security number is a relational structure). But in fact relational structures come in all sorts of domains: besides relative possibility of states of the world, we can have epistemic states of some agent related by epistemic possibility, or states of a dynamical system with their state transitions, etc. Modal logic can be used to model all of these: the first give us ordinary, alethic, modal logic; the others give us epistemic logic, dynamic logic, etc.

We focus on one particular angle, known to modal logicians as "correspondence theory." One of the most significant early discoveries of Kripke's is that many properties of the accessibility relation R (whether it is transitive, symmetric, etc.) can be characterized *in the modal language* itself by means of appropriate "modal schemas." Modal logicians say, for instance, that the reflexivity of R "corresponds" to the schema "If necessarily A, then A". We explore mainly the correspondence theory of a number of classical systems of modal logic (e.g., **S4** and **S5**) obtained by a combination of the schemas D, T, B, 4, and 5.

1.2 The Language of Basic Modal Logic

Definition 1.1. The basic language of modal logic contains

1. The propositional constant for falsity $\bot$.
2. A countably infinite set of propositional variables: p_0, p_1, p_2, ...
3. The propositional connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional).
4. The modal operator $\Box$.
5. The modal operator $\Diamond$.

Definition 1.2. *Formulas* of the basic modal language are inductively defined as follows:

1. $\bot$ is an atomic formula.
2. Every propositional variable p_i is an (atomic) formula.
3. If A and B are formulas, then $(A \wedge B)$ is a formula.
4. If A and B are formulas, then $(A \vee B)$ is a formula.
5. If A and B are formulas, then $(A \to B)$ is a formula.
6. If A is a formula, so is $\Box A$.
7. If A is a formula, then $\Diamond A$ is a formula.
8. Nothing else is a formula.

If a formula A does not contain $\Box$ or $\Diamond$, we say it is *modal-free*.

1.3 Simultaneous Substitution

An *instance* of a formula A is the result of replacing all occurrences of a propositional variable in A by some other formula. We will refer to instances of formulas often, both when discussing validity and when discussing derivability. It therefore is useful to define the notion precisely.

Definition 1.3. Where A is a modal formula all of whose propositional variables are among $p_1, \ldots, p_n$, and $D_1, \ldots, D_n$ are also modal formulas, we define $A[D_1/p_1, \ldots, D_n/p_n]$ as the result of simultaneously substituting each D_i for p_i in A. Formally, this is a definition by induction on A:

1. $A \equiv \bot$: $A[D_1/p_1, \ldots, D_n/p_n]$ is $\bot$.
2. $A \equiv q$: $A[D_1/p_1, \ldots, D_n/p_n]$ is q, provided $q \not\equiv p_i$ for $i = 1$,

$\ldots, n$.

3. $A \equiv p_i$: $A[D_1/p_1, \ldots, D_n/p_n]$ is D_i.
4. $A \equiv \neg B$: $A[D_1/p_1, \ldots, D_n/p_n]$ is $\neg B[D_1/p_1, \ldots, D_n/p_n]$.
5. $A \equiv (B \wedge C)$: $A[D_1/p_1, \ldots, D_n/p_n]$ is
$$(B[D_1/p_1, \ldots, D_n/p_n] \wedge C[D_1/p_1, \ldots, D_n/p_n]).$$
6. $A \equiv (B \vee C)$: $A[D_1/p_1, \ldots, D_n/p_n]$ is
$$(B[D_1/p_1, \ldots, D_n/p_n] \vee C[D_1/p_1, \ldots, D_n/p_n]).$$
7. $A \equiv (B \rightarrow C)$: $A[D_1/p_1, \ldots, D_n/p_n]$ is
$$(B[D_1/p_1, \ldots, D_n/p_n] \rightarrow C[D_1/p_1, \ldots, D_n/p_n]).$$
8. $A \equiv (B \leftrightarrow C)$: $A[D_1/p_1, \ldots, D_n/p_n]$ is
$$(B[D_1/p_1, \ldots, D_n/p_n] \leftrightarrow C[D_1/p_1, \ldots, D_n/p_n]).$$
9. $A \equiv \Box B$: $A[D_1/p_1, \ldots, D_n/p_n]$ is $\Box B[D_1/p_1, \ldots, D_n/p_n]$.
10. $A \equiv \Diamond B$: $A[D_1/p_1, \ldots, D_n/p_n]$ is $\Diamond B[D_1/p_1, \ldots, D_n/p_n]$.

The formula $A[D_1/p_1, \ldots, D_n/p_n]$ is called a *substitution instance* of A.

Example 1.4. Suppose A is $p_1 \rightarrow \Box(p_1 \wedge p_2)$, D_1 is $\Diamond(p_2 \rightarrow p_3)$ and D_2 is $\neg\Box p_1$. Then $A[D_1/p_1, D_2/p_2]$ is

$$\Diamond(p_2 \rightarrow p_3) \rightarrow \Box(\Diamond(p_2 \rightarrow p_3) \wedge \neg\Box p_1)$$

while $A[D_2/p_1, D_1/p_2]$ is

$$\neg\Box p_1 \rightarrow \Box(\neg\Box p_1 \wedge \Diamond(p_2 \rightarrow p_3))$$

Note that simultaneous substitution is in general not the same as iterated substitution, e.g., compare $A[D_1/p_1, D_2/p_2]$ above with $(A[D_1/p_1])[D_2/p_2]$, which is:

$$\Diamond(p_2 \to p_3) \to \Box(\Diamond(p_2 \to p_3) \wedge p_2)[\neg\Box p_1/p_2], \text{ i.e.,}$$
$$\Diamond(\neg\Box p_1 \to p_3) \to \Box(\Diamond(\neg\Box p_1 \to p_3) \wedge \neg\Box p_1)$$

and with $(A[D_2/p_2])[D_1/p_1]$:

$$p_1 \to \Box(p_1 \wedge \neg\Box p_1)[\Diamond(p_2 \to p_3)/p_1], \text{ i.e.,}$$
$$\Diamond(p_2 \to p_3) \to \Box(\Diamond(p_2 \to p_3) \wedge \neg\Box\Diamond(p_2 \to p_3)).$$

1.4 Relational Models

The basic concept of semantics for normal modal logics is that of a *relational model*. It consists of a set of worlds, which are related by a binary "accessibility relation," together with an assignment which determines which propositional variables count as "true" at which worlds.

Definition 1.5. A *model* for the basic modal language is a triple $M = \langle W, R, V\rangle$, where

1. W is a nonempty set of "worlds,"
2. R is a binary accessibility relation on W, and
3. V is a function assigning to each propositional variable p a set $V(p)$ of possible worlds.

When Rww' holds, we say that w' is *accessible from* w. When $w \in V(p)$ we say p is *true at* w.

The great advantage of relational semantics is that models can be represented by means of simple diagrams, such as the one in Figure 1.1. Worlds are represented by nodes, and world w' is accessible from w precisely when there is an arrow

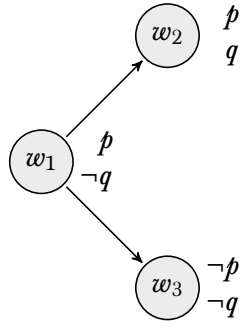


Figure 1.1: A simple model.

from w to w'. Moreover, we label a node (world) by p when $w \in V(p)$, and otherwise by $\neg p$. Figure 1.1 represents the model with $W = \{w_1, w_2, w_3\}$, $R = \{\langle w_1, w_2\rangle, \langle w_1, w_3\rangle\}$, $V(p) = \{w_1, w_2\}$, and $V(q) = \{w_2\}$.

1.5 Truth at a World

Every modal model determines which modal formulas count as true at which worlds in it. The relation "model M makes formula A true at world w" is the basic notion of relational semantics. The relation is defined inductively and coincides with the usual characterization using truth tables for the non-modal operators.

Definition 1.6. *Truth of a formula A at w* in a M, in symbols: $M, w \Vdash A$, is defined inductively as follows:

1. $A \equiv \bot$: Never $M, w \Vdash \bot$.
2. $M, w \Vdash p$ iff $w \in V(p)$
3. $A \equiv \neg B$: $M, w \Vdash A$ iff $M, w \nVdash B$.

4. $A \equiv (B \wedge C)$: $M, w \Vdash A$ iff $M, w \Vdash B$ and $M, w \Vdash C$.
5. $A \equiv (B \vee C)$: $M, w \Vdash A$ iff $M, w \Vdash B$ or $M, w \Vdash C$ (or both).
6. $A \equiv (B \to C)$: $M, w \Vdash A$ iff $M, w \nVdash B$ or $M, w \Vdash C$.
7. $A \equiv \Box B$: $M, w \Vdash A$ iff $M, w' \Vdash B$ for all $w' \in W$ with Rww'
8. $A \equiv \Diamond B$: $M, w \Vdash A$ iff $M, w' \Vdash B$ for at least one $w' \in W$ with Rww'

Note that by clause (7), a formula $\Box B$ is true at w whenever there are no w' with wRw'. In such a case $\Box B$ is *vacuously* true at w. Also, $\Box B$ may be satisfied at w even if B is not. The truth of B at w does not guarantee the truth of $\Diamond B$ at w. This holds, however, if Rww, e.g., if R is reflexive. If there is no w' such that Rww', then $M, w \nVdash \Diamond A$, for any A.

Proposition 1.7. *1.* $M, w \Vdash \Box A$ *iff* $M, w \Vdash \neg\Diamond\neg A$.

2. $M, w \Vdash \Diamond A$ *iff* $M, w \Vdash \neg\Box\neg A$.

Proof. 1. $M, w \Vdash \neg\Diamond\neg A$ iff $M \nVdash \Diamond\neg A$ by definition of $M, w \Vdash$. $M, w \Vdash \Diamond\neg A$ iff for some w' with Rww', $M, w' \Vdash \neg A$. Hence, $M, w \nVdash \Diamond\neg A$ iff for all w' with Rww', $M, w' \nVdash \neg A$. We also have $M, w' \nVdash \neg A$ iff $M, w' \Vdash A$. Together we have $M, w \Vdash \neg\Diamond\neg A$ iff for all w' with Rww', $M, w' \Vdash A$. Again by definition of $M, w \Vdash$, that is the case iff $M, w \Vdash \Box A$.

2. Exercise. □

1.6 Truth in a Model

Sometimes we are interested which formulas are true at every world in a given model. Let's introduce a notation for this.

Definition 1.8. A formula A is *true in a model* $M = \langle W, R, V \rangle$, written $M \Vdash A$, if and only if $M, w \Vdash A$ for every $w \in W$.

Proposition 1.9. *1. If* $M \Vdash A$ *then* $M \nVdash \neg A$, *but* not *vice-versa.*

2. If $M \Vdash A \to B$ *then* $M \Vdash A$ *only if* $M \Vdash B$, *but* not *vice-versa.*

Proof. 1. If $M \Vdash A$ then A is true at all worlds in W, and since $W \neq \emptyset$, it can't be that $M \Vdash \neg A$, or else A would have to be both true and false at some world.

On the other hand, if $M \nVdash \neg A$ then A is true at some world $w \in W$. It does not follow that $M, w \Vdash A$ for *every* $w \in W$. For instance, in the model of Figure 1.1, $M \nVdash \neg p$, and also $M \nVdash p$.

2. Assume $M \Vdash A \to B$ and $M \Vdash A$; to show $M \Vdash B$ let $w \in W$ be an arbitrary world. Then $M, w \Vdash A \to B$ and $M, w \Vdash A$, so $M, w \Vdash B$, and since w was arbitrary, $M \Vdash B$.

To show that the converse fails, we need to find a model M such that $M \Vdash A$ only if $M \Vdash B$, but $M \nVdash A \to B$. Consider again the model of Figure 1.1: $M \nVdash p$ and hence (vacuously) $M \Vdash p$ only if $M \Vdash q$. However, $M \nVdash p \to q$, as p is true but q false at w_1. □

1.7 Validity

Formulas that are true in all models, i.e., true at every world in every model, are particularly interesting. They represent those modal propositions which are true regardless of how $\Box$ and $\Diamond$ are interpreted, as long as the interpretation is "normal" in the sense that it is generated by some accessibility relation on possible worlds. We call such formulas *valid*. For instance, $\Box(p \wedge q) \to \Box p$ is valid. Some formulas one might expect to be valid on the basis

of the alethic interpretation of $\Box$, such as $\Box p \to p$, are not valid, however. Part of the interest of relational models is that different interpretations of $\Box$ and $\Diamond$ can be captured by different kinds of accessibility relations. This suggests that we should define validity not just relative to *all* models, but relative to all models *of a certain kind*. It will turn out, e.g., that $\Box p \to p$ is true in all models where every world is accessible from itself, i.e., R is reflexive. Defining validity relative to classes of models enables us to formulate this succinctly: $\Box p \to p$ is valid in the class of reflexive models.

Definition 1.10. A formula A is *valid* in a class $\mathscr{C}$ of models if it is true in every model in $\mathscr{C}$ (i.e., true at every world in every model in $\mathscr{C}$). If A is valid in $\mathscr{C}$, we write $\mathscr{C} \vDash A$, and we write $\vDash A$ if A is valid in the class of *all* models.

Proposition 1.11. *If A is valid in $\mathscr{C}$ it is also valid in each class $\mathscr{C}' \subseteq \mathscr{C}$.*

Proposition 1.12. *If A is valid, then so is $\Box A$.*

Proof. Assume $\vDash A$. To show $\vDash \Box A$ let $\boldsymbol{M} = \langle W, R, V \rangle$ be a model and $w \in W$. If Rww' then $\boldsymbol{M}, w' \Vdash A$, since A is valid, and so also $\boldsymbol{M}, w \Vdash \Box A$. Since $\boldsymbol{M}$ and w were arbitrary, $\vDash \Box A$. □

1.8 Tautological Instances

A modal-free formula is a tautology if it is true under every truth-value assignment. Clearly, every tautology is true at every world in every model. But for formulas involving $\Box$ and $\Diamond$, the notion of tautology is not defined. Is it the case, e.g., that $\Box p \vee \neg \Box p$—an instance of the principle of excluded middle—is valid? The notion of a *tautological instance* helps: a formula that is a substitution instance of a (non-modal) tautology. It is not surprising, but still requires proof, that every tautological instance is valid.

Definition 1.13. A modal formula B is a *tautological instance* if and only if there is a modal-free tautology A with propositional variables $p_1, \ldots, p_n$ and formulas $D_1, \ldots, D_n$ such that $B \equiv A[D_1/p_1, \ldots, D_n/p_n]$.

Lemma 1.14. *Suppose A is a modal-free formula whose propositional variables are $p_1, \ldots, p_n$, and let $D_1, \ldots, D_n$ be modal formulas. Then for any assignment v, any model $\mathbf{M} = \langle W, R, V \rangle$, and any $w \in W$ such that $v(p_i) = \mathbb{T}$ if and only if $\mathbf{M}, w \Vdash D_i$ we have that $v \vDash A$ if and only if $\mathbf{M}, w \Vdash A[D_1/p_1, \ldots, D_n/p_n]$.*

Proof. By induction on A.

1. $A \equiv \bot$: Both $v \nvDash \bot$ and $\mathbf{M}, w \nVdash \bot$.

2. $A \equiv p_i$:

$$\begin{aligned} v \vDash p_i &\Leftrightarrow v(p_i) = \mathbb{T} \\ &\qquad \text{by definition of } v \vDash p_i \\ &\Leftrightarrow \mathbf{M}, w \Vdash D_i \\ &\qquad \text{by assumption} \\ &\Leftrightarrow \mathbf{M}, w \Vdash p_i[D_1/p_1, \ldots, D_n/p_n] \\ &\qquad \text{since } p_i[D_1/p_1, \ldots, D_n/p_n] \equiv D_i. \end{aligned}$$

3. $A \equiv \neg B$:

$$\begin{aligned} v \vDash \neg B &\Leftrightarrow v \nvDash B \\ &\qquad \text{by definition of } v \vDash; \\ &\Leftrightarrow \mathbf{M}, w \nVdash B[D_1/p_1, \ldots, D_n/p_n] \\ &\qquad \text{by induction hypothesis} \\ &\Leftrightarrow \mathbf{M}, w \Vdash \neg B[D_1/p_1, \ldots, D_n/p_n] \\ &\qquad \text{by definition of } v \vDash. \end{aligned}$$

4. $A \equiv (B \wedge C)$:

$$\begin{aligned} v \vDash B \wedge C &\Leftrightarrow v \vDash B \text{ and } v \vDash C \\ &\qquad \text{by definition of } v \vDash \\ &\Leftrightarrow M, w \Vdash B[D_1/p_1, \ldots, D_n/p_n] \text{ and} \\ &\quad\ M, w \Vdash C[D_1/p_1, \ldots, D_n/p_n] \\ &\qquad \text{by induction hypothesis} \\ &\Leftrightarrow M, w \Vdash (B \wedge C)[D_1/p_1, \ldots, D_n/p_n] \\ &\qquad \text{by definition of } M, w \Vdash. \end{aligned}$$

5. $A \equiv (B \vee C)$:

$$\begin{aligned} v \vDash B \vee C &\Leftrightarrow v \vDash B \text{ or } v \vDash C \\ &\qquad \text{by definition of } v \vDash; \\ &\Leftrightarrow M, w \Vdash B[D_1/p_1, \ldots, D_n/p_n] \text{ or} \\ &\quad\ M, w \Vdash C[D_1/p_1, \ldots, D_n/p_n] \\ &\qquad \text{by induction hypothesis} \\ &\Leftrightarrow M, w \Vdash (B \vee C)[D_1/p_1, \ldots, D_n/p_n] \\ &\qquad \text{by definition of } M, w \Vdash. \end{aligned}$$

6. $A \equiv (B \rightarrow C)$:

$$\begin{aligned} v \vDash B \rightarrow C &\Leftrightarrow v \nvDash B \text{ or } v \vDash C \\ &\qquad \text{by definition of } v \vDash \\ &\Leftrightarrow M, w \nVdash B[D_1/p_1, \ldots, D_n/p_n] \text{ or} \\ &\quad\ M, w \Vdash C[D_1/p_1, \ldots, D_n/p_n] \\ &\qquad \text{by induction hypothesis} \\ &\Leftrightarrow M, w \Vdash (B \rightarrow C)[D_1/p_1, \ldots, D_n/p_n] \\ &\qquad \text{by definition of } M, w \Vdash. \end{aligned}$$

□

Proposition 1.15. *All tautological instances are valid.*

Proof. Contrapositively, suppose A is such that $\boldsymbol{M}, w \nVdash A[D_1/p_1, \ldots, D_n/p_n]$, for some model $\boldsymbol{M}$ and world w. Define an assignment v such that $v(p_i) = \mathbb{T}$ if and only if $\boldsymbol{M}, w \Vdash D_i$ (and v assigns arbitrary values to $q \notin \{p_1, \ldots, p_n\}$). Then by Lemma 1.14, $v \nvDash A$, so A is not a tautology. □

1.9 Schemas and Validity

Definition 1.16. A *schema* is a set of formulas comprising all and only the substitution instances of some modal formula C, i.e.,

$$\{B : \exists D_1, \ldots, \exists D_n \left(B = C[D_1/p_1, \ldots, D_n/p_n]\right)\}.$$

The formula C is called the *characteristic* formula of the schema, and it is unique up to a renaming of the propositional variables. A formula A is an *instance* of a schema if it is a member of the set.

It is convenient to denote a schema by the meta-linguistic expression obtained by substituting 'A', 'B', ..., for the atomic components of C. So, for instance, the following denote schemas: 'A', '$A \rightarrow \Box A$', '$A \rightarrow (B \rightarrow A)$'. They correspond to the characteristic formulas p, $p \rightarrow \Box p$, $p \rightarrow (q \rightarrow p)$. The schema '$A$' denotes the set of *all* formulas.

Definition 1.17. A schema is *true* in a model if and only if all of its instances are; and a schema is *valid* if and only if it is true in every model.

Proposition 1.18. *The following schema* K *is valid*

$$\Box(A \to B) \to (\Box A \to \Box B). \tag{K}$$

Proof. We need to show that all instances of the schema are true at every world in every model. So let $\boldsymbol{M} = \langle W, R, V \rangle$ and $w \in W$ be arbitrary. To show that a conditional is true at a world we assume the antecedent is true to show that consequent is true as well. In this case, let $\boldsymbol{M}, w \Vdash \Box(A \to B)$ and $\boldsymbol{M}, w \Vdash \Box A$. We need to show $\boldsymbol{M} \Vdash \Box B$. So let w' be arbitrary such that Rww'. Then by the first assumption $\boldsymbol{M}, w' \Vdash A \to B$ and by the second assumption $\boldsymbol{M}, w' \Vdash A$. It follows that $\boldsymbol{M}, w' \Vdash B$. Since w' was arbitrary, $\boldsymbol{M}, w \Vdash \Box B$. □

Proposition 1.19. *The following schema* DUAL *is valid*

$$\Diamond A \leftrightarrow \neg\Box\neg A. \tag{DUAL}$$

Proof. Exercise. □

Proposition 1.20. *If A and $A \to B$ are true at a world in a model then so is B. Hence, the valid formulas are closed under modus ponens.*

Proposition 1.21. *A formula A is valid iff all its substitution instances are. In other words, a schema is valid iff its characteristic formula is.*

Proof. The "if" direction is obvious, since A is a substitution instance of itself.

To prove the "only if" direction, we show the following: Suppose $\boldsymbol{M} = \langle W, R, V \rangle$ is a modal model, and $B \equiv A[D_1/p_1, \ldots, D_n/p_n]$ is a substitution instance of A. Define $\boldsymbol{M}' = \langle W, R, V' \rangle$ by $V(p_i) = \{w : \boldsymbol{M}, w \Vdash D_i\}$. Then $\boldsymbol{M}, w \Vdash B$ iff $\boldsymbol{M}', w \Vdash A$, for any $w \in W$. (We leave the proof as an exercise.)

Valid Schemas	*Invalid Schemas*
$\Box(A \to B) \to (\Diamond A \to \Diamond B)$	$\Box(A \vee B) \to (\Box A \vee \Box B)$
$\Diamond(A \to B) \to (\Box A \to \Diamond B)$	$(\Diamond A \wedge \Diamond B) \to \Diamond(A \wedge B)$
$\Box(A \wedge B) \leftrightarrow (\Box A \wedge \Box B)$	$A \to \Box A$
$\Box A \to \Box(B \to A)$	$\Box\Diamond A \to B$
$\neg\Diamond A \to \Box(A \to B)$	$\Box\Box A \to \Box A$
$\Diamond(A \vee B) \leftrightarrow (\Diamond A \vee \Diamond B)$	$\Box\Diamond A \to \Diamond\Box A$.

Table 1.1: Valid and (or?) invalid schemas.

Now suppose that A was valid, but some substitution instance B of A was not valid. Then for some $\boldsymbol{M} = \langle W, R, V\rangle$ and some $w \in W$, $\boldsymbol{M}, w \nVdash B$. But then $\boldsymbol{M}', w \nVdash A$ by the claim, and A is not valid, a contradiction. □

Note, however, that it is not true that a schema is true in a model iff its characteristic formula is. Of course, the "only if" direction holds: if every instance of A is true in $\boldsymbol{M}$, A itself is true in $\boldsymbol{M}$. But it may happen that A is true in $\boldsymbol{M}$ but some instance of A is false at some world in $\boldsymbol{M}$. For a very simple counterexample consider p in a model with only one world w and $V(p) = \{w\}$, so that p is true at w. But $\bot$ is an instance of p, and not true at w.

1.10 Entailment

With the definition of truth at a world, we can define an entailment relation between formulas. A formula B entails A iff, whenever B is true, A is true as well. Here, "whenever" means both "whichever model we consider" as well as "whichever world in that model we consider."

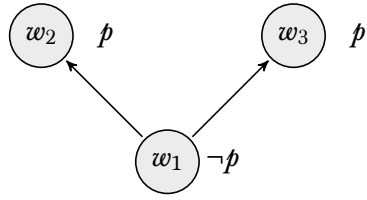


Figure 1.2: Counterexample to $p \to \Diamond p \vDash \Box p \to p$.

Definition 1.22. If Γ is a set of formulas and A a formula, then Γ *entails* A, in symbols: $\Gamma \vDash A$, if and only if for every model $M = \langle W, R, V \rangle$ and world $w \in W$, if $M, w \Vdash B$ for every $B \in \Gamma$, then $M, w \Vdash A$. If Γ contains a single formula B, then we write $B \vDash A$.

Example 1.23. To show that a formula entails another, we have to reason about all models, using the definition of $M, w \Vdash$. For instance, to show $p \to \Diamond p \vDash \Box\neg p \to \neg p$, we might argue as follows: Consider a model $M = \langle W, R, V \rangle$ and $w \in W$, and suppose $M, w \Vdash p \to \Diamond p$. We have to show that $M, w \Vdash \Box\neg p \to \neg p$. Suppose not. Then $M, w \Vdash \Box\neg p$ and $M, w \nVdash \neg p$. Since $M, w \nVdash \neg p$, $M, w \Vdash p$. By assumption, $M, w \Vdash p \to \Diamond p$, hence $M, w \Vdash \Diamond p$. By definition of $M, w \Vdash \Diamond p$, there is some w' with Rww' such that $M, w' \Vdash p$. Since also $M, w \Vdash \Box\neg p$, $M, w' \Vdash \neg p$, a contradiction.

To show that a formula B does not entail another A, we have to give a counterexample, i.e., a model $M = \langle W, R, V \rangle$ where we show that at some world $w \in W$, $M, w \Vdash B$ but $M, w \nVdash A$. Let's show that $p \to \Diamond p \nvDash \Box p \to p$. Consider the model in Figure 1.2. We have $M, w_1 \Vdash \Diamond p$ and hence $M, w_1 \Vdash p \to \Diamond p$. However, since $M, w_1 \Vdash \Box p$ but $M, w_1 \nVdash p$, we have $M, w_1 \nVdash \Box p \to p$.

Often very simple counterexamples suffice. The model $M' = \{W', R', V'\}$ with $W' = \{w\}$, $R' = \emptyset$, and $V'(p) = \emptyset$ is also a counterexample: Since $M', w \nVdash p$, $M', w \Vdash p \to \Diamond p$. As no worlds are accessible from w, we have $M', w \Vdash \Box p$, and so $M', w \nVdash \Box p \to p$.

Problems

Problem 1.1. Consider the model of Figure 1.1. Which of the following hold?

1. $M, w_1 \Vdash q$;
2. $M, w_3 \Vdash \neg q$;
3. $M, w_1 \Vdash p \vee q$;
4. $M, w_1 \Vdash \Box(p \vee q)$;
5. $M, w_3 \Vdash \Box q$;
6. $M, w_3 \Vdash \Box \bot$;
7. $M, w_1 \Vdash \Diamond q$;
8. $M, w_1 \Vdash \Box q$;
9. $M, w_1 \Vdash \neg\Box\Box\neg q$.

Problem 1.2. Complete the proof of Proposition 1.7.

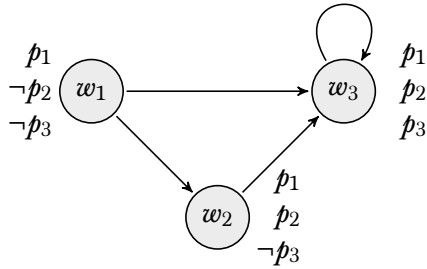
Problem 1.3. Let $M = \langle W, R, V\rangle$ be a model, and suppose $w_1, w_2 \in W$ are such that:

1. $w_1 \in V(p)$ if and only if $w_2 \in V(p)$; and
2. for all $w \in W$: Rw_1w if and only if Rw_2w.

Using induction on formulas, show that for all formulas A: $M, w_1 \Vdash A$ if and only if $M, w_2 \Vdash A$.

Problem 1.4. Let $M = \langle M, R, V\rangle$. Show that $M, w \Vdash \neg\Diamond A$ if and only if $M, w \Vdash \Box\neg A$.

Problem 1.5. Consider the following model M for the language comprising p_1, p_2, p_3 as the only propositional variables:



Are the following formulas and schemas true in the model M, i.e., true at every world in M? Explain.

1. $p \to \Diamond p$ (for p atomic);
2. $A \to \Diamond A$ (for A arbitrary);
3. $\Box p \to p$ (for p atomic);
4. $\neg p \to \Diamond \Box p$ (for p atomic);
5. $\Diamond \Box A$ (for A arbitrary);
6. $\Box \Diamond p$ (for p atomic).

Problem 1.6. Show that the following are valid:

1. $\vDash \Box p \to \Box(q \to p)$;
2. $\vDash \Box \neg \bot$;
3. $\vDash \Box p \to (\Box q \to \Box p)$.

Problem 1.7. Show that $A \to \Box A$ is valid in the class $\mathscr{C}$ of models $M = \langle W, R, V \rangle$ where $W = \{w\}$. Similarly, show that $B \to \Box A$ and $\Diamond A \to B$ are valid in the class of models $M = \langle W, R, V \rangle$ where $R = \emptyset$.

Problem 1.8. Prove Proposition 1.19.

Problem 1.9. Prove the claim in the "only if" part of the proof of Proposition 1.21. (Hint: use induction on A.)

Problem 1.10. Show that none of the following formulas are valid:

D: $\Box p \to \Diamond p$;

T: $\Box p \to p$;

B: $p \to \Box \Diamond p$;

4: $\Box p \to \Box \Box p$;

5: $\Diamond p \to \Box \Diamond p$.

Problem 1.11. Prove that the schemas in the first column of table 1.1 are valid and those in the second column are not valid.

Problem 1.12. Decide whether the following schemas are valid or invalid:

1. $(\Diamond A \to \Box B) \to (\Box A \to \Box B)$;
2. $\Diamond(A \to B) \vee \Box(B \to A)$.

Problem 1.13. For each of the following schemas find a model M such that every instance of the formula is true in M:

1. $p \to \Diamond \Diamond p$;
2. $\Diamond p \to \Box p$.

Problem 1.14. Show that $\Box(A \wedge B) \vDash \Box A$.

Problem 1.15. Show that $\Box(p \to q) \nvDash p \to \Box q$ and $p \to \Box q \nvDash \Box(p \to q)$.

CHAPTER 2

Frame Definability

2.1 Introduction

One question that interests modal logicians is the relationship between the accessibility relation and the truth of certain formulas in models with that accessibility relation. For instance, suppose the accessibility relation is reflexive, i.e., for every $w \in W$, Rww. In other words, every world is accessible from itself. That means that when $\Box A$ is true at a world w, w itself is among the accessible worlds at which A must therefore be true. So, if the accessibility relation R of $\boldsymbol{M}$ is reflexive, then whatever world w and formula A we take, $\Box A \rightarrow A$ will be true there (in other words, the schema $\Box p \rightarrow p$ and all its substitution instances are true in $\boldsymbol{M}$).

The converse, however, is false. It's not the case, e.g., that if $\Box p \rightarrow p$ is true in $\boldsymbol{M}$, then R is reflexive. For we can easily find a non-reflexive model $\boldsymbol{M}$ where $\Box p \rightarrow p$ is true at all worlds: take the model with a single world w, not accessible from itself, but with $w \in V(p)$. By picking the truth value of p suitably, we can make $\Box A \rightarrow A$ true in a model that is not reflexive.

The solution is to remove the variable assignment V from the equation. If we require that $\Box p \rightarrow p$ is true at all worlds in $\boldsymbol{M}$, regardless of which worlds are in $V(p)$, then it is necessary that

R is reflexive. For in any non-reflexive model, there will be at least one world w such that not Rww. If we set $V(p) = W \setminus \{w\}$, then p will be true at all worlds other than w, and so at all worlds accessible from w (since w is guaranteed not to be accessible from w, and w is the only world where p is false). On the other hand, p is false at w, so $\Box p \to p$ is false at w.

This suggests that we should introduce a notation for model structures without a valuation: we call these *frames*. A frame $\boldsymbol{F}$ is simply a pair $\langle W, R\rangle$ consisting of a set of worlds with an accessibility relation. Every model $\langle W, R, V\rangle$ is then, as we say, *based on* the frame $\langle W, R\rangle$. Conversely, a frame determines the class of models based on it; and a class of frames determines the class of models which are based on any frame in the class. And we can define $\boldsymbol{F} \vDash A$, the notion of a formula being *valid* in a frame as: $\boldsymbol{M} \Vdash A$ for all $\boldsymbol{M}$ based on $\boldsymbol{F}$.

With this notation, we can establish correspondence relations between formulas and classes of frames: e.g., $\boldsymbol{F} \vDash \Box p \to p$ if, and only if, $\boldsymbol{F}$ is reflexive.

2.2 Properties of Accessibility Relations

Many modal formulas turn out to be characteristic of simple, and even familiar, properties of the accessibility relation. In one direction, that means that any model that has a given property makes a corresponding formula (and all its substitution instances) true. We begin with five classical examples of kinds of accessibility relations and the formulas the truth of which they guarantee.

Theorem 2.1. *Let $\boldsymbol{M} = \langle W, R, V\rangle$ be a model. If R has the property on the left side of table 2.1, every instance of the formula on the right side is true in $\boldsymbol{M}$.*

Proof. Here is the case for B: to show that the schema is true in a model we need to show that all of its instances are true at all worlds in the model. So let $A \to \Box\Diamond A$ be a given instance of B,

If R is ...	*then ... is true in $\boldsymbol{M}$:*	
serial: $\forall u \exists v Ruv$	$\Box p \to \Diamond p$	(D)
reflexive: $\forall w Rww$	$\Box p \to p$	(T)
symmetric: $\forall u \forall v (Ruv \to Rvu)$	$p \to \Box \Diamond p$	(B)
transitive: $\forall u \forall v \forall w ((Ruv \wedge Rvw) \to Ruw)$	$\Box p \to \Box \Box p$	(4)
euclidean: $\forall w \forall u \forall v ((Rwu \wedge Rwv) \to Ruv)$	$\Diamond p \to \Box \Diamond p$	(5)

Table 2.1: Five correspondence facts.

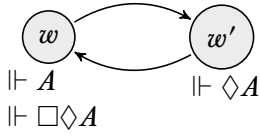


Figure 2.1: The argument from symmetry.

and let $w \in W$ be an arbitrary world. Suppose the antecedent A is true at w, in order to show that $\Box \Diamond A$ is true at w. So we need to show that $\Diamond A$ is true at all w' accessible from w. Now, for any w' such that Rww' we have, using the hypothesis of symmetry, that also $Rw'w$ (see Figure 2.1). Since $\boldsymbol{M}, w \Vdash A$, we have $\boldsymbol{M}, w' \Vdash \Diamond A$. Since w' was an arbitrary world such that Rww', we have $\boldsymbol{M}, w \Vdash \Box \Diamond A$.

We leave the other cases as exercises. □

Notice that the converse implications of Theorem 2.1 do not hold: it's not true that if a model verifies a schema, then the accessibility relation of that model has the corresponding property. In the case of T and reflexive models, it is easy to give an example of a model in which T itself fails: let $W = \{w\}$ and $V(p) = \emptyset$. Then R is not reflexive, but $\boldsymbol{M}, w \Vdash \Box p$ and $\boldsymbol{M}, w \nVdash p$. But here we have just a single instance of T that fails in $\boldsymbol{M}$, other instances,

e.g., $\Box\neg p \rightarrow \neg p$ are true. It is harder to give examples where *every substitution instance* of T is true in $\boldsymbol{M}$ and $\boldsymbol{M}$ is not reflexive. But there are such models, too:

Proposition 2.2. *Let $\boldsymbol{M} = \langle W, R, V\rangle$ be a model such that $W = \{u, v\}$, where worlds u and v are related by R: i.e., both Ruv and Rvu. Suppose that for all p: $u \in V(p) \Leftrightarrow v \in V(p)$. Then:*

1. *For all A: $\boldsymbol{M}, u \Vdash A$ if and only if $\boldsymbol{M}, v \Vdash A$ (use induction on A).*
2. *Every instance of* T *is true in $\boldsymbol{M}$.*

Since $\boldsymbol{M}$ is not reflexive (it is, in fact, irreflexive*), the converse of Theorem 2.1 fails in the case of* T *(similar arguments can be given for some—though not all—the other schemas mentioned in Theorem 2.1).*

Although we will focus on the five classical formulas D, T, B, 4, and 5, we record in table 2.2 a few more properties of accessibility relations. The accessibility relation R is partially functional, if from every world at most one world is accessible. If it is the case that from every world exactly one world is accessible, we call it functional. (Thus the functional relations are precisely those that are both serial and partially functional). They are called "functional" because the accessibility relation operates like a (partial) function. A relation is weakly dense if whenever Ruv, there is a w "between" u and v. So weakly dense relations are in a sense the opposite of transitive relations: in a transitive relation, whenever you can reach v from u by a detour via w, you can reach v from u directly; in a weakly dense relation, whenever you can reach v from u directly, you can also reach it by a detour via some w. A relation is weakly directed if whenever you can reach worlds u and v from some world w, you can reach a single world t from both u and v—this is sometimes called the "diamond property" or "confluence."

If R is ...	*then ... is true in $\boldsymbol{M}$:*
partially functional: $\forall w \forall u \forall v((Rwu \wedge Rwv) \rightarrow u = v)$	$\Diamond p \rightarrow \Box p$
functional: $\forall w \exists v \forall u(Rwu \leftrightarrow u = v)$	$\Diamond p \leftrightarrow \Box p$
weakly dense: $\forall u \forall v(Ruv \rightarrow \exists w(Ruw \wedge Rwv))$	$\Box\Box p \rightarrow \Box p$
weakly connected: $\forall w \forall u \forall v((Rwu \wedge Rwv) \rightarrow (Ruv \vee u = v \vee Rvu))$	$\Box((p \wedge \Box p) \rightarrow q) \vee \Box((q \wedge \Box q) \rightarrow p)$ (L)
weakly directed: $\forall w \forall u \forall v((Rwu \wedge Rwv) \rightarrow \exists t(Rut \wedge Rvt))$	$\Diamond\Box p \rightarrow \Box\Diamond p$ (G)

Table 2.2: Five more correspondence facts.

2.3 Frames

Definition 2.3. A *frame* is a pair $\boldsymbol{F} = \langle W, R \rangle$ where W is a non-empty set of worlds and R a binary relation on W. A model $\boldsymbol{M}$ is *based on* a frame $\boldsymbol{F} = \langle W, R \rangle$ if and only if $\boldsymbol{M} = \langle W, R, V \rangle$ for some valuation V.

Definition 2.4. If $\boldsymbol{F}$ is a frame, we say that A is *valid in* $\boldsymbol{F}$, $\boldsymbol{F} \vDash A$, if $\boldsymbol{M} \Vdash A$ for every model $\boldsymbol{M}$ based on $\boldsymbol{F}$.

If $\mathfrak{F}$ is a class of frames, we say A is *valid in* $\mathfrak{F}$, $\mathfrak{F} \vDash A$, iff $\boldsymbol{F} \vDash A$ for every frame $\boldsymbol{F} \in \mathfrak{F}$.

The reason frames are interesting is that correspondence between schemas and properties of the accessibility relation R is at the level of frames, *not of models.* For instance, although T is true in all reflexive models, not every model in which T is true is reflexive. However, it *is* true that not only is T *valid* on all reflexive *frames*, also every frame in which T is valid is reflexive.

Remark 1. Validity in a class of frames is a special case of the notion of validity in a class of models: $\mathfrak{F} \vDash A$ iff $\mathfrak{C} \vDash A$ where $\mathfrak{C}$ is the class of all models based on a frame in $\mathfrak{F}$.

Obviously, if a formula or a schema is valid, i.e., valid with respect to the class of *all* models, it is also valid with respect to any class $\mathfrak{F}$ of frames.

2.4 Frame Definability

Even though the converse implications of Theorem 2.1 fail, they hold if we replace "model" by "frame": for the properties considered in Theorem 2.1, it *is* true that if a formula is valid in a *frame* then the accessibility relation of that frame has the corresponding property. So, the formulas considered *define* the classes of frames that have the corresponding property.

Definition 2.5. If $\mathfrak{C}$ is a class of frames, we say A *defines* $\mathfrak{C}$ iff $\boldsymbol{F} \vDash A$ for all and only frames $\boldsymbol{F} \in \mathfrak{C}$.

We now proceed to establish the full definability results for frames.

Theorem 2.6. *If the formula on the right side of table 2.1 is valid in a frame $\boldsymbol{F}$, then $\boldsymbol{F}$ has the property on the left side.*

Proof. 1. Suppose D is valid in $\boldsymbol{F} = \langle W, R\rangle$, i.e., $\boldsymbol{F} \vDash \Box p \to \Diamond p$. Let $\boldsymbol{M} = \langle W, R, V\rangle$ be a model based on $\boldsymbol{F}$, and $w \in W$. We have to show that there is a v such that Rwv. Suppose not: then both $\boldsymbol{M} \Vdash \Box A$ and $\boldsymbol{M}, w \nVdash \Diamond A$ for any A, including p. But then $\boldsymbol{M}, w \nVdash \Box p \to \Diamond p$, contradicting the assumption that $\boldsymbol{F} \vDash \Box p \to \Diamond p$.

2. Suppose T is valid in $\boldsymbol{F}$, i.e., $\boldsymbol{F} \vDash \Box p \to p$. Let $w \in W$ be an arbitrary world; we need to show Rww. Let $u \in V(p)$ if and only if Rwu (when q is other than p, $V(q)$ is arbitrary, say $V(q) = \emptyset$). Let $\boldsymbol{M} = \langle W, R, V\rangle$. By construction, for all

u such that Rwu: $\boldsymbol{M}, u \Vdash p$, and hence $\boldsymbol{M}, w \Vdash \Box p$. But by hypothesis $\Box p \rightarrow p$ is true at w, so that $\boldsymbol{M}, w \Vdash p$, but by definition of V this is possible only if Rww.

3. We prove the contrapositive: Suppose $\boldsymbol{F}$ is not symmetric, we show that B, i.e., $p \rightarrow \Box\Diamond p$ is not valid in $\boldsymbol{F} = \langle W, R\rangle$. If $\boldsymbol{F}$ is not symmetric, there are $u, v \in W$ such that Ruv but not Rvu. Define V such that $w \in V(p)$ if and only if not Rvw (and V is arbitrary otherwise). Let $\boldsymbol{M} = \langle W, R, V\rangle$. Now, by definition of V, $\boldsymbol{M}, w \Vdash p$ for all w such that not Rvw, in particular, $\boldsymbol{M}, u \Vdash p$ since not Rvu. Also, since Rvw iff $w \notin V(p)$, there is no w such that Rvw and $\boldsymbol{M}, w \Vdash p$, and hence $\boldsymbol{M}, v \nVdash \Diamond p$. Since Ruv, also $\boldsymbol{M}, u \nVdash \Box\Diamond p$. It follows that $\boldsymbol{M}, u \nVdash p \rightarrow \Box\Diamond p$, and so B is not valid in $\boldsymbol{F}$.

4. Suppose 4 is valid in $\boldsymbol{F} = \langle W, R\rangle$, i.e., $\boldsymbol{F} \vDash \Box p \rightarrow \Box\Box p$, and let u, v, $w \in W$ be arbitrary worlds such that Ruv and Rvw; we need to show that Ruw. Define V such that $z \in V(p)$ if and only if Ruz (and V is arbitrary otherwise). Let $\boldsymbol{M} = \langle W, R, V\rangle$. By definition of V, $\boldsymbol{M}, z \Vdash p$ for all z such that Ruz, and hence $\boldsymbol{M}, u \Vdash \Box p$. But by hypothesis 4, $\Box p \rightarrow \Box\Box p$, is true at u, so that $\boldsymbol{M}, u \Vdash \Box\Box p$. Since Ruv and Rvw, we have $\boldsymbol{M}, w \Vdash p$, but by definition of V this is possible only if Ruw, as desired.

5. We proceed contrapositively, assuming that the frame $\boldsymbol{F} = \langle W, R\rangle$ is not euclidean, and show that it falsifies 5, i.e., $\boldsymbol{F} \nVdash \Diamond p \rightarrow \Box\Diamond p$. Suppose there are worlds u, v, $w \in W$ such that Rwu and Rwv but not Ruv. Define V such that for all worlds z, $z \in V(p)$ if and only if it is *not* the case that Ruz. Let $\boldsymbol{M} = \langle W, R, V\rangle$. Then by hypothesis $\boldsymbol{M}, v \Vdash p$ and since Rwv also $\boldsymbol{M}, w \Vdash \Diamond p$. However, there is no world y such that Ruy and $\boldsymbol{M}, y \Vdash p$ so $\boldsymbol{M}, u \nVdash \Diamond p$. Since Rwu, it follows that $\boldsymbol{M}, w \nVdash \Box\Diamond p$, so that 5, $\Diamond p \rightarrow \Box\Diamond p$, fails at w. □

You'll notice a difference between the proof for D and the other cases: no mention was made of the valuation V. In effect,

we proved that if $\boldsymbol{M} \Vdash \mathrm{D}$ then $\boldsymbol{M}$ is serial. So D defines the class of serial *models*, not just frames.

Corollary 2.7. *Any model where* D *is true is serial.*

Corollary 2.8. *Each formula on the right side of table 2.1 defines the class of frames which have the property on the left side.*

Proof. In Theorem 2.1, we proved that if a model has the property on the left, the formula on the right is true in it. Thus, if a frame $\boldsymbol{F}$ has the property on the left, the formula on the right is valid in $\boldsymbol{F}$. In Theorem 2.6, we proved the converse implications: if a formula on the right is valid in $\boldsymbol{F}$, $\boldsymbol{F}$ has the property on the left. □

Theorem 2.6 also shows that the properties can be combined: for instance if both B and 4 are valid in $\boldsymbol{F}$ then the frame is both symmetric and transitive, etc. Many important modal logics are characterized as the set of formulas valid in all frames that combine some frame properties, and so we can characterize them as the set of formulas valid in all frames in which the corresponding defining formulas are valid. For instance, the classical system **S4** is the set of all formulas valid in all reflexive and transitive frames, i.e., in all those where both T and 4 are valid. **S5** is the set of all formulas valid in all reflexive, symmetric, and euclidean frames, i.e., all those where all of T, B, and 5 are valid.

Logical relationships between properties of R in general correspond to relationships between the corresponding defining formulas. For instance, every reflexive relation is serial; hence, whenever T is valid in a frame, so is D. (Note that this relationship is *not* that of entailment. It is not the case that whenever $\boldsymbol{M}, w \Vdash \mathrm{T}$ then $\boldsymbol{M}, w \Vdash \mathrm{D}$.) We record some such relationships.

Proposition 2.9. *Let R be a binary relation on a set W; then:*

1. *If R is reflexive, then it is serial.*

2. *If R is symmetric, then it is transitive if and only if it is euclidean.*
3. *If R is symmetric or euclidean then it is weakly directed (it has the "diamond property").*
4. *If R is euclidean then it is weakly connected.*
5. *If R is functional then it is serial.*

2.5 First-order Definability

We've seen that a number of properties of accessibility relations of frames can be defined by modal formulas. For instance, symmetry of frames can be defined by the formula B, $p \to \Box\Diamond p$. The conditions we've encountered so far can all be expressed by first-order formulas in a language involving a single two-place predicate symbol. For instance, symmetry is defined by $\forall x \forall y (Q(x,y) \to Q(y,x))$ in the sense that a first-order structure $\boldsymbol{M}$ with $|\boldsymbol{M}| = W$ and $Q^{\boldsymbol{M}} = R$ satisfies the preceding formula iff R is symmetric. This suggests the following definition:

Definition 2.10. A class $\mathscr{C}$ of frames is *first-order definable* if there is a sentence A in the first-order language with a single two-place predicate symbol Q such that $\boldsymbol{F} = \langle W, R\rangle \in \mathscr{C}$ iff $\boldsymbol{M} \vDash A$ in the first-order structure $\boldsymbol{M}$ with $|\boldsymbol{M}| = W$ and $Q^{\boldsymbol{M}} = R$.

It turns out that the properties and modal formulas that define them considered so far are exceptional. Not every formula defines a first-order definable class of frames, and not every first-order definable class of frames is definable by a modal formula.

A counterexample to the first is given by the Löb formula:

$$\Box(\Box p \to p) \to \Box p. \tag{W}$$

W defines the class of transitive and converse well-founded frames. A relation is well-founded if there is no infinite sequence

$w_1, w_2, \ldots$ such that Rw_2w_1, Rw_3w_2, ... For instance, the relation $<$ on $\mathbb{N}$ is well-founded, whereas the relation $<$ on $\mathbb{Z}$ is not. A relation is converse well-founded iff its converse is well-founded. So converse well-founded relations are those where there is no infinite sequence $w_1, w_2, \ldots$ such that Rw_1w_2, Rw_2w_3, ...

There is, however, no first-order formula defining transitive converse well-founded relations. For suppose $M \models F$ iff $R = Q^M$ is transitive converse well-founded. Let A_n be the formula

$$(Q(a_1, a_2) \wedge \cdots \wedge Q(a_{n-1}, a_n))$$

Now consider the set of formulas

$$\Gamma = \{F, A_1, A_2, \ldots\}.$$

Every finite subset of Γ is satisfiable: Let k be largest such that A_k is in the subset, $|M_k| = \{1, \ldots, k\}$, $a_i^{M_k} = i$, and $P^{M_k} = <$. Since $<$ on $\{1, \ldots, k\}$ is transitive and converse well-founded, $M_k \models F$. $M_k \models A_i$ by construction, for all $i \leq k$. By the Compactness Theorem for first-order logic, Γ is satisfiable in some structure M. By hypothesis, since $M \models F$, the relation Q^M is converse well-founded. But clearly, $a_1^M, a_2^M, \ldots$ would form an infinite sequence of the kind ruled out by converse well-foundedness.

A counterexample to the second claim is given by the property of universality: for every u and v, Ruv. Universal frames are first-order definable by the formula $\forall x \forall y\, Q(x, y)$. However, no modal formula is valid in all and only the universal frames. This is a consequence of a result that is independently interesting: the formulas valid in universal frames are exactly the same as those valid in reflexive, symmetric, and transitive frames. There are reflexive, symmetric, and transitive frames that are not universal, hence every formula valid in all universal frames is also valid in some non-universal frames.

2.6 Equivalence Relations and S5

The modal logic **S5** is characterized as the set of formulas valid on all universal frames, i.e., every world is accessible from every

world, including itself. In such a scenario, $\Box$ corresponds to necessity and $\Diamond$ to possibility: $\Box A$ is true if A is true at *every* world, and $\Diamond A$ is true if A is true at *some* world. It turns out that **S5** can also be characterized as the formulas valid on all reflexive, symmetric, and transitive frames, i.e., on all *equivalence relations*.

Definition 2.11. A binary relation R on W is an *equivalence relation* if and only if it is reflexive, symmetric and transitive. A relation R on W is *universal* if and only if Ruv for all $u, v \in W$.

Since T, B, and 4 characterize the reflexive, symmetric, and transitive frames, the frames where the accessibility relation is an equivalence relation are exactly those in which all three formulas are valid. It turns out that the equivalence relations can also be characterized by other combinations of formulas, since the conditions with which we've defined equivalence relations are equivalent to combinations of other familiar conditions on R.

Proposition 2.12. *The following are equivalent:*

1. *R is an equivalence relation;*
2. *R is reflexive and euclidean;*
3. *R is serial, symmetric, and euclidean;*
4. *R is serial, symmetric, and transitive.*

Proof. Exercise. □

Proposition 2.12 is the semantic counterpart to Proposition 3.29, in that it gives an equivalent characterization of the modal logic of frames over which R is an equivalence relation (the logic traditionally referred to as **S5**).

What is the relationship between universal and equivalence relations? Although every universal relation is an equivalence

relation, clearly not every equivalence relation is universal. However, the formulas valid on all universal relations are exactly the same as those valid on all equivalence relations.

Proposition 2.13. *Let R be an equivalence relation, and for each $w \in W$ define the* equivalence class *of w as the set $[w] = \{w' \in W : Rww'\}$. Then:*

1. *$w \in [w]$;*
2. *R is universal on each equivalence class $[w]$;*
3. *The collection of equivalence classes partitions W into mutually exclusive and jointly exhaustive subsets.*

Proposition 2.14. *A formula A is valid in all frames $\mathbf{F} = \langle W, R\rangle$ where R is an equivalence relation, if and only if it is valid in all frames $\mathbf{F} = \langle W, R\rangle$ where R is universal. Hence, the logic of universal frames is just* **S5**.

Proof. It's immediate to verify that a universal relation R on W is an equivalence. Hence, if A is valid in all frames where R is an equivalence it is valid in all universal frames. For the other direction, we argue contrapositively: suppose B is a formula that fails at a world w in a model $\boldsymbol{M} = \langle W, R, V\rangle$ based on a frame $\langle W, R\rangle$, where R is an equivalence on W. So $\boldsymbol{M}, w \nVdash B$. Define a model $\boldsymbol{M}' = \langle W', R', V'\rangle$ as follows:

1. $W' = [w]$;
2. R' is universal on W';
3. $V'(p) = V(p) \cap W'$.

(So the set W' of worlds in $\boldsymbol{M}'$ is represented by the shaded area in Figure 2.2.) It is easy to see that R and R' agree on W'. Then one can show by induction on formulas that for all $w' \in W'$: $\boldsymbol{M}', w' \Vdash A$ if and only if $\boldsymbol{M}, w' \Vdash A$ for each A (this makes sense

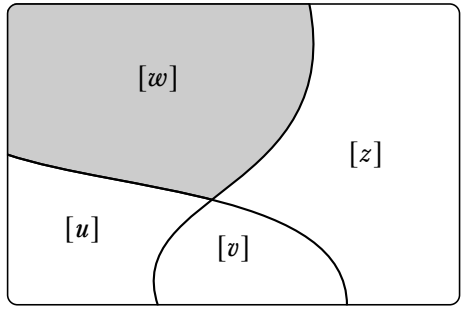


Figure 2.2: A partition of W in equivalence classes.

since $W' \subseteq W$). In particular, $M', w \nVdash B$, and B fails in a model based on a universal frame. □

2.7 Second-order Definability

Not every frame property definable by modal formulas is first-order definable. However, if we allow quantification over one-place predicates (i.e., monadic second-order quantification), we define all modally definable frame properties. The trick is to exploit a systematic way in which the conditions under which a modal formula is true at a world are related to first-order formulas. This is the so-called standard translation of modal formulas into first-order formulas in a language containing not just a two-place predicate symbol Q for the accessibility relation, but also a one-place predicate symbol P_i for the propositional variables p_i occurring in A.

Definition 2.15. The *standard translation* $\mathrm{ST}_x(A)$ is inductively defined as follows:

1. $A \equiv \bot$: $\mathrm{ST}_x(A) = \bot$.
2. $A \equiv p_i$: $\mathrm{ST}_x(A) = P_i(x)$.

3. $A \equiv \neg B$: $\mathrm{ST}_x(A) = \neg \mathrm{ST}_x(B)$.
4. $A \equiv (B \wedge C)$: $\mathrm{ST}_x(A) = (\mathrm{ST}_x(B) \wedge \mathrm{ST}_x(C))$.
5. $A \equiv (B \vee C)$: $\mathrm{ST}_x(A) = (\mathrm{ST}_x(B) \vee \mathrm{ST}_x(C))$.
6. $A \equiv (B \rightarrow C)$: $\mathrm{ST}_x(A) = (\mathrm{ST}_x(B) \rightarrow \mathrm{ST}_x(C))$.
7. $A \equiv \Box B$: $\mathrm{ST}_x(A) = \forall y\,(Q(x,y) \rightarrow \mathrm{ST}_y(B))$.
8. $A \equiv \Diamond B$: $\mathrm{ST}_x(A) = \exists y\,(Q(x,y) \wedge \mathrm{ST}_y(B))$.

For instance, $\mathrm{ST}_x(\Box p \rightarrow p)$ is $\forall y\,(Q(x,y) \rightarrow P(y)) \rightarrow P(x)$. Any structure for the language of $\mathrm{ST}_x(A)$ requires a domain, a two-place relation assigned to Q, and subsets of the domain assigned to the one-place predicate symbols P_i. In other words, the components of such a structure are exactly those of a model for A: the domain is the set of worlds, the two-place relation assigned to Q is the accessibility relation, and the subsets assigned to P_i are just the assignments $V(p_i)$. It won't surprise that satisfaction of A in a modal model and of $\mathrm{ST}_x(A)$ in the corresponding structure agree:

Proposition 2.16. *Let $\boldsymbol{M} = \langle W, R, V\rangle$, $\boldsymbol{M}'$ be the first-order structure with $|\boldsymbol{M}'| = W$, $Q^{\boldsymbol{M}'} = R$, and $P_i^{\boldsymbol{M}'} = V(p_i)$, and $s(x) = w$. Then*

$$\boldsymbol{M}, w \Vdash A \text{ iff } \boldsymbol{M}', s \vDash \mathrm{ST}_x(A)$$

Proof. By induction on A. □

Proposition 2.17. *Suppose A is a modal formula and $\boldsymbol{F} = \langle W, R\rangle$ is a frame. Let $\boldsymbol{F}'$ be the first-order structure with $|\boldsymbol{F}'| = W$ and $Q^{\boldsymbol{F}'} = R$, and let A' be the second-order formula*

$$\forall X_1 \ldots \forall X_n\, \forall x\, \mathrm{ST}_x(A)[X_1/P_1, \ldots, X_n/P_n],$$

where $P_1, \ldots, P_n$ are all one-place predicate symbols in $\mathrm{ST}_x(A)$. *Then*

$$\boldsymbol{F} \vDash A \textit{ iff } \boldsymbol{F}' \vDash A'$$

Proof. $\boldsymbol{F}' \vDash A'$ iff for every structure $\boldsymbol{M}'$ where $P_i^{\boldsymbol{M}'} \subseteq W$ for $i = 1, \ldots, n$, and for every s with $s(x) \in W$, $\boldsymbol{M}', s \vDash \mathrm{ST}_x(A)$. By Proposition 2.16, that is the case iff for all models $\boldsymbol{M}$ based on $\boldsymbol{F}$ and every world $w \in W$, $\boldsymbol{M}, w \Vdash A$, i.e., $\boldsymbol{F} \vDash A$. □

Definition 2.18. A class $\mathscr{C}$ of frames is *second-order definable* if there is a sentence A in the second-order language with a single two-place predicate symbol P and quantifiers only over monadic set variables such that $\boldsymbol{F} = \langle W, R\rangle \in \mathscr{C}$ iff $\boldsymbol{M} \vDash A$ in the structure $\boldsymbol{M}$ with $|\boldsymbol{M}| = W$ and $P^{\boldsymbol{M}} = R$.

Corollary 2.19. *If a class of frames is definable by a formula A, the corresponding class of accessibility relations is definable by a monadic second-order sentence.*

Proof. The monadic second-order sentence A' of the preceding proof has the required property. □

As an example, consider again the formula $\Box p \to p$. It defines reflexivity. Reflexivity is of course first-order definable by the sentence $\forall x\, Q(x, x)$. But it is also definable by the monadic second-order sentence

$$\forall X\, \forall x\, (\forall y\, (Q(x, y) \to X(y)) \to X(x)).$$

This means, of course, that the two sentences are equivalent. Here's how you might convince yourself of this directly: First suppose the second-order sentence is true in a structure $\boldsymbol{M}$. Since x and X are universally quantified, the remainder must hold for any $x \in W$ and set $X \subseteq W$, e.g., the set $\{z : Rxz\}$ where $R = Q^{\boldsymbol{M}}$. So, for any s with $s(x) \in W$ and $s(X) = \{z : Rxz\}$ we have

$\boldsymbol{M} \vDash \forall y\,(Q(x,y) \to X(y)) \to X(x)$. But by the way we've picked $s(X)$ that means $\boldsymbol{M}, s \vDash \forall y\,(Q(x,y) \to Q(x,y)) \to Q(x,x)$, which is equivalent to $Q(x,x)$ since the antecedent is valid. Since $s(x)$ is arbitrary, we have $\boldsymbol{M} \vDash \forall x\, Q(x,x)$.

Now suppose that $\boldsymbol{M} \vDash \forall x\, Q(x,x)$ and show that $\boldsymbol{M} \vDash \forall X\, \forall x\,(\forall y\,(Q(x,y) \to X(y)) \to X(x))$. Pick any assignment s, and assume $\boldsymbol{M}, s \vDash \forall y\,(Q(x,y) \to X(y))$. Let s' be the y-variant of s with $s'(y) = x$; we have $\boldsymbol{M}, s' \vDash Q(x,y) \to X(y)$, i.e., $\boldsymbol{M}, s \vDash Q(x,x) \to X(x)$. Since $\boldsymbol{M} \vDash \forall x\, Q(x,x)$, the antecedent is true, and we have $\boldsymbol{M}, s \vDash X(x)$, which is what we needed to show.

Since some definable classes of frames are not first-order definable, not every monadic second-order sentence of the form A' is equivalent to a first-order sentence. There is no effective method to decide which ones are.

Problems

Problem 2.1. Complete the proof of Theorem 2.1.

Problem 2.2. Prove the claims in Proposition 2.2.

Problem 2.3. Let $\boldsymbol{M} = \langle W, R, V\rangle$ be a model. Show that if R satisfies the left-hand properties of table 2.2, every instance of the corresponding right-hand formula is true in $\boldsymbol{M}$.

Problem 2.4. Show that if the formula on the right side of table 2.2 is valid in a frame $\boldsymbol{F}$, then $\boldsymbol{F}$ has the property on the left side. To do this, consider a frame that does *not* satisfy the property on the left, and define a suitable V such that the formula on the right is false at some world.

Problem 2.5. Prove Proposition 2.9.

Problem 2.6. Prove Proposition 2.12 by showing:

1. If R is symmetric and transitive, it is euclidean.

2. If R is reflexive, it is serial.

3. If R is reflexive and euclidean, it is symmetric.

4. If R is symmetric and euclidean, it is transitive.

5. If R is serial, symmetric, and transitive, it is reflexive.

Explain why this suffices for the proof that the conditions are equivalent.

CHAPTER 3

Axiomatic Derivations

3.1 Introduction

We have a semantics for the basic modal language in terms of modal models, and a notion of a formula being valid—true at all worlds in all models—or valid with respect to some class of models or frames—true at all worlds in all models in the class, or based on the frame. Logic usually connects such semantic characterizations of validity with a proof-theoretic notion of derivability. The aim is to define a notion of derivability in some system such that a formula is derivable iff it is valid.

The simplest and historically oldest derivation systems are so-called Hilbert-type or axiomatic derivation systems. Hilbert-type derivation systems for many modal logics are relatively easy to construct: they are simple as objects of metatheoretical study (e.g., to prove soundness and completeness). However, they are much harder to use to prove formulas in than, say, natural deduction systems.

In Hilbert-type derivation systems, a derivation of a formula is a sequence of formulas leading from certain axioms, via a handful of inference rules, to the formula in question. Since we want the derivation system to match the semantics, we have to guarantee

that the set of derivable formulas are true in all models (or true in all models in which all axioms are true). We'll first isolate some properties of modal logics that are necessary for this to work: the "normal" modal logics. For normal modal logics, there are only two inference rules that need to be assumed: modus ponens and necessitation. As axioms we take all (substitution instances) of tautologies, and, depending on the modal logic we deal with, a number of modal axioms. Even if we are just interested in the class of all models, we must also count all substitution instances of K and Dual as axioms. This alone generates the minimal normal modal logic **K**.

Definition 3.1. The rule of *modus ponens* is the inference schema

$$\frac{A \qquad A \to B}{B}\ \text{MP}$$

We say a formula B *follows from* formulas A, C by modus ponens iff $C \equiv A \to B$.

Definition 3.2. The rule of *necessitation* is the inference schema

$$\frac{A}{\Box A}\ \text{NEC}$$

We say the formula B follows from the formulas A by necessitation iff $B \equiv \Box A$.

Definition 3.3. A *derivation* from a set of axioms Σ is a sequence of formulas $B_1, B_2, \ldots, B_n$, where each B_i is either

1. a substitution instance of a tautology, or
2. a substitution instance of a formula in Σ, or
3. follows from two formulas B_j, B_k with $j, k < i$ by modus

ponens, or

4. follows from a formula B_j with $j < i$ by necessitation.

If there is such a derivation with $B_n \equiv A$, we say that A is *derivable from* Σ, in symbols $\Sigma \vdash A$.

With this definition, it will turn out that the set of derivable formulas forms a normal modal logic, and that any derivable formula is true in every model in which every axiom is true. This property of derivations is called *soundness*. The converse, *completeness*, is harder to prove.

3.2 Normal Modal Logics

Not every set of modal formulas can easily be characterized as those formulas derivable from a set of axioms. We want modal logics to be well-behaved. First of all, everything we can derive in classical propositional logic should still be derivable, of course taking into account that the formulas may now contain also $\Box$ and $\Diamond$. To this end, we require that a modal logic contain all tautological instances and be closed under modus ponens.

Definition 3.4. A *modal logic* is a set Σ of modal formulas which

1. contains all tautologies, and
2. is closed under substitution, i.e., if $A \in \Sigma$, and $D_1, \ldots, D_n$ are formulas, then
$$A[D_1/p_1, \ldots, D_n/p_n] \in \Sigma,$$
3. is closed under *modus ponens*, i.e., if A and $A \to B \in \Sigma$, then $B \in \Sigma$.

In order to use the relational semantics for modal logics, we also have to require that all formulas valid in all modal models

are included. It turns out that this requirement is met as soon as all instances of K and DUAL are derivable, and whenever a formula A is derivable, so is $\Box A$. A modal logic that satisfies these conditions is called *normal*. (Of course, there are also non-normal modal logics, but the usual relational models are not adequate for them.)

Definition 3.5. A modal logic Σ is *normal* if it contains

$$\Box(p \to q) \to (\Box p \to \Box q), \tag{K}$$
$$\Diamond p \leftrightarrow \neg\Box\neg p \tag{DUAL}$$

and is closed under *necessitation*, i.e., if $A \in \Sigma$, then $\Box A \in \Sigma$.

Observe that while tautological implication is "fine-grained" enough to preserve *truth at a world*, the rule NEC only preserves *truth in a model* (and hence also validity in a frame or in a class of frames).

Proposition 3.6. *Every normal modal logic is closed under rule RK,*

$$\frac{A_1 \to (A_2 \to \cdots (A_{n-1} \to A_n)\cdots)}{\Box A_1 \to (\Box A_2 \to \cdots (\Box A_{n-1} \to \Box A_n)\cdots).}\ RK$$

Proof. By induction on n: If $n = 1$, then the rule is just NEC, and every normal modal logic is closed under NEC.

Now suppose the result holds for $n-1$; we show it holds for n. Assume

$$A_1 \to (A_2 \to \cdots (A_{n-1} \to A_n)\cdots) \in \Sigma$$

By the induction hypothesis, we have

$$\Box A_1 \to (\Box A_2 \to \cdots \Box(A_{n-1} \to A_n)\cdots) \in \Sigma$$

Since Σ is a normal modal logic, it contains all instances of K, in particular

$$\Box(A_{n-1} \to A_n) \to (\Box A_{n-1} \to \Box A_n) \in \Sigma$$

Using modus ponens and suitable tautological instances we get

$$\Box A_1 \to (\Box A_2 \to \cdots (\Box A_{n-1} \to \Box A_n) \cdots) \in \Sigma. \qquad \Box$$

Proposition 3.7. *Every normal modal logic Σ contains $\neg\Diamond\bot$.*

Proposition 3.8. *Let $A_1, \ldots, A_n$ be formulas. Then there is a smallest modal logic Σ containing all instances of $A_1, \ldots, A_n$.*

Proof. Given $A_1, \ldots, A_n$, define Σ as the intersection of all normal modal logics containing all instances of $A_1, \ldots, A_n$. The intersection is non-empty as $\mathrm{Frm}(\mathscr{L})$, the set of all formulas, is such a modal logic. $\Box$

Definition 3.9. The smallest normal modal logic containing $A_1, \ldots, A_n$ is called a *modal system* and denoted by $\mathbf{K}A_1 \ldots A_n$. The smallest normal modal logic is denoted by $\mathbf{K}$.

3.3 Derivations and Modal Systems

We first define what a derivation is for normal modal logics. Roughly, a derivation is a sequence of formulas in which every element is either (a substitution instance of) one of a number of *axioms*, or follows from previous elements by one of a few inference rules. For normal modal logics, all instances of tautologies, K, and DUAL count as axioms. This results in the modal system $\mathbf{K}$, the smallest normal modal logic. We may wish to add additional axioms to obtain other systems, however. The rules are always modus ponens MP and necessitation NEC.

Definition 3.10. Given a modal system $\mathbf{K}A_1 \dots A_n$ and a formula B we say that B is *derivable* in $\mathbf{K}A_1 \dots A_n$, written $\mathbf{K}A_1 \dots A_n \vdash B$, if and only if there are formulas $C_1, \dots, C_k$ such that $C_k = B$ and each C_i is either a tautological instance, or an instance of one of K, DUAL, $A_1, \dots, A_n$, or it follows from previous formulas by means of the rules MP or NEC.

The following proposition allows us to show that $B \in \Sigma$ by exhibiting a Σ-proof of B.

Proposition 3.11. $\mathbf{K}A_1 \dots A_n = \{B : \mathbf{K}A_1 \dots A_n \vdash B\}$.

Proof. We use induction on the length of derivations to show that $\{B : \mathbf{K}A_1 \dots A_n \vdash B\} \subseteq \mathbf{K}A_1 \dots A_n$.

If the derivation of B has length 1, it contains a single formula. That formula cannot follow from previous formulas by MP or NEC, so must be a tautological instance, an instance of K, DUAL, or an instance of one of $A_1, \dots, A_n$. But $\mathbf{K}A_1 \dots A_n$ contains these as well, so $B \in \mathbf{K}A_1 \dots A_n$.

If the derivation of B has length > 1, then B may in addition be obtained by MP or NEC from formulas not occurring as the last line in the derivation. If B follows from C and $C \to B$ (by MP), then C and $C \to B \in \mathbf{K}A_1 \dots A_n$ by induction hypothesis. But every modal logic is closed under modus ponens, so $B \in \mathbf{K}A_1 \dots A_n$. If $B \equiv \Box C$ follows from C by NEC, then $C \in \mathbf{K}A_1 \dots A_n$ by induction hypothesis. But every normal modal logic is closed under NEC, so $B \in \mathbf{K}A_1 \dots A_n$.

The converse inclusion follows by showing that $\Sigma = \{B : \mathbf{K}A_1 \dots A_n \vdash B\}$ is a normal modal logic containing all the instances of $A_1, \dots, A_n$, and the observation that $\mathbf{K}A_1 \dots A_n$ is, by definition, the smallest such logic.

1. Every tautology B is a tautological instance, so $\mathbf{K}A_1 \dots A_n \vdash B$, so Σ contains all tautologies.

2. If $\mathbf{K}A_1 \dots A_n \vdash C$ and $\mathbf{K}A_1 \dots A_n \vdash C \to B$, then $\mathbf{K}A_1 \dots A_n \vdash B$: Combine the derivation of C with that

of $C \to B$, and add the line B. The last line is justified by MP. So Σ is closed under modus ponens.

3. If B has a derivation, then every substitution instance of B also has a derivation: apply the substitution to every formula in the derivation. (Exercise: prove by induction on the length of derivations that the result is also a correct derivation). So Σ is closed under uniform substitution. (We have now established that Σ satisfies all conditions of a modal logic.)

4. We have $\mathbf{K}A_1 \dots A_n \vdash \mathrm{K}$, so $K \in \Sigma$.

5. We have $\mathbf{K}A_1 \dots A_n \vdash \mathrm{DUAL}$, so $\mathrm{DUAL} \in \Sigma$.

6. If $\mathbf{K}A_1 \dots A_n \vdash C$, the additional line $\Box C$ is justified by NEC. Consequently, Σ is closed under NEC. Thus, Σ is normal. □

3.4 Proofs in K

In order to practice proofs in the smallest modal system, we show the valid formulas on the left-hand side of table 1.1 can all be given **K**-proofs.

Proposition 3.12. $\mathbf{K} \vdash \Box A \to \Box(B \to A)$

Proof.

1.	$A \to (B \to A)$	TAUT
2.	$\Box(A \to (B \to A))$	NEC, 1
3.	$\Box(A \to (B \to A)) \to (\Box A \to \Box(B \to A))$	K
4.	$\Box A \to \Box(B \to A)$	MP, 2, 3

□

Proposition 3.13. $\mathbf{K} \vdash \Box(A \wedge B) \to (\Box A \wedge \Box B)$

Proof.

1.	$(A \wedge B) \to A$	TAUT
2.	$\Box((A \wedge B) \to A)$	NEC
3.	$\Box((A \wedge B) \to A) \to (\Box(A \wedge B) \to \Box A)$	K
4.	$\Box(A \wedge B) \to \Box A$	MP, 2, 3
5.	$(A \wedge B) \to B$	TAUT
6.	$\Box((A \wedge B) \to B)$	NEC
7.	$\Box((A \wedge B) \to B) \to (\Box(A \wedge B) \to \Box B)$	K
8.	$\Box(A \wedge B) \to \Box B$	MP, 6, 7
9.	$(\Box(A \wedge B) \to \Box A) \to$ $((\Box(A \wedge B) \to \Box B) \to$ $(\Box(A \wedge B) \to (\Box A \wedge \Box B)))$	TAUT
10.	$(\Box(A \wedge B) \to \Box B) \to$ $(\Box(A \wedge B) \to (\Box A \wedge \Box B))$	MP, 4, 9
11.	$\Box(A \wedge B) \to (\Box A \wedge \Box B)$	MP, 8, 10.

Note that the formula on line 9 is an instance of the tautology

$$(p \to q) \to ((p \to r) \to (p \to (q \wedge r))).$$

□

Proposition 3.14. $\mathbf{K} \vdash (\Box A \wedge \Box B) \to \Box(A \wedge B)$

Proof.

1.	$A \to (B \to (A \wedge B))$	TAUT
2.	$\Box(A \to (B \to (A \wedge B)))$	NEC, 1
3.	$\Box(A \to (B \to (A \wedge B))) \to (\Box A \to \Box(B \to (A \wedge B)))$	K
4.	$\Box A \to \Box(B \to (A \wedge B))$	MP, 2, 3
5.	$\Box(B \to (A \wedge B)) \to (\Box B \to \Box(A \wedge B))$	K
6.	$(\Box A \to \Box(B \to (A \wedge B))) \to$ $(\Box(B \to (A \wedge B)) \to (\Box B \to \Box(A \wedge B))) \to$ $(\Box A \to (\Box B \to \Box(A \wedge B))))$	TAUT
7.	$(\Box(B \to (A \wedge B)) \to (\Box B \to \Box(A \wedge B))) \to$ $(\Box A \to (\Box B \to \Box(A \wedge B)))$	MP, 4, 6
8.	$\Box A \to (\Box B \to \Box(A \wedge B)))$	MP, 5, 7
9.	$(\Box A \to (\Box B \to \Box(A \wedge B)))) \to$ $((\Box A \wedge \Box B) \to \Box(A \wedge B))$	TAUT
10.	$(\Box A \wedge \Box B) \to \Box(A \wedge B)$	MP, 8, 9

The formulas on lines 6 and 9 are instances of the tautologies

$$(p \to q) \to ((q \to r) \to (p \to r))$$
$$(p \to (q \to r)) \to ((p \wedge q) \to r)$$

□

Proposition 3.15. $\mathbf{K} \vdash \neg\Box p \to \Diamond\neg p$

Proof.

1.	$\Diamond\neg p \leftrightarrow \neg\Box\neg\neg p$	DUAL
2.	$(\Diamond\neg p \leftrightarrow \neg\Box\neg\neg p) \to$ $(\neg\Box\neg\neg p \to \Diamond\neg p)$	TAUT
3.	$\neg\Box\neg\neg p \to \Diamond\neg p$	MP, 1, 2
4.	$\neg\neg p \to p$	TAUT
5.	$\Box(\neg\neg p \to p)$	NEC, 4
6.	$\Box(\neg\neg p \to p) \to (\Box\neg\neg p \to \Box p)$	K
7.	$(\Box\neg\neg p \to \Box p)$	MP, 5, 6
8.	$(\Box\neg\neg p \to \Box p) \to (\neg\Box p \to \neg\Box\neg\neg p)$	TAUT
9.	$\neg\Box p \to \neg\Box\neg\neg p$	MP, 7, 8
10.	$(\neg\Box p \to \neg\Box\neg\neg p) \to$ $((\neg\Box\neg\neg p \to \Diamond\neg p) \to (\neg\Box p \to \Diamond\neg p))$	TAUT
11.	$(\neg\Box\neg\neg p \to \Diamond\neg p) \to (\neg\Box p \to \Diamond\neg p)$	MP, 9, 10
12.	$\neg\Box p \to \Diamond\neg p$	MP, 3, 11

The formulas on lines 8 and 10 are instances of the tautologies

$$(p \to q) \to (\neg q \to \neg p)$$
$$(p \to q) \to ((q \to r) \to (p \to r)).$$ □

3.5 Derived Rules

Finding and writing derivations is obviously difficult, cumbersome, and repetitive. For instance, very often we want to pass from $A \to B$ to $\Box A \to \Box B$, i.e., apply rule RK. That requires an application of NEC, then recording the proper instance of K, then applying MP. Passing from $A \to B$ and $B \to C$ to $A \to C$ requires recording the (long) tautological instance

$$(A \to B) \to ((B \to C) \to (A \to C))$$

and applying MP twice. Often we want to replace a sub-formula by a formula we know to be equivalent, e.g., $\Diamond A$ by $\neg\Box\neg A$, or

$\neg\neg A$ by A. So rather than write out the actual derivation, it is more convenient to simply record why the intermediate steps are derivable. For this purpose, let us collect some facts about derivability.

Proposition 3.16. *If* $\mathbf{K} \vdash A_1, \ldots, \mathbf{K} \vdash A_n$*, and* B *follows from* $A_1, \ldots, A_n$ *by propositional logic, then* $\mathbf{K} \vdash B$.

Proof. If B follows from $A_1, \ldots, A_n$ by propositional logic, then

$$A_1 \to (A_2 \to \cdots (A_n \to B) \ldots)$$

is a tautological instance. Applying MP n times gives a derivation of B. □

We will indicate use of this proposition by PL.

Proposition 3.17. *If* $\mathbf{K} \vdash A_1 \to (A_2 \to \cdots (A_{n-1} \to A_n) \ldots)$ *then* $\mathbf{K} \vdash \Box A_1 \to (\Box A_2 \to \cdots (\Box A_{n-1} \to \Box A_n) \ldots)$.

Proof. By induction on n, just as in the proof of Proposition 3.6.□

We will indicate use of this proposition by RK. Let's illustrate how these results help establishing derivability results more easily.

Proposition 3.18. $\mathbf{K} \vdash (\Box A \wedge \Box B) \to \Box(A \wedge B)$

Proof.

1.	$\mathbf{K} \vdash A \to (B \to (A \wedge B))$	TAUT
2.	$\mathbf{K} \vdash \Box A \to (\Box B \to \Box(A \wedge B)))$	RK, 1
3.	$\mathbf{K} \vdash (\Box A \wedge \Box B) \to \Box(A \wedge B)$	PL, 2

□

Proposition 3.19. *If* $\mathbf{K} \vdash A \leftrightarrow B$ *and* $\mathbf{K} \vdash C[A/q]$ *then* $\mathbf{K} \vdash C[B/q]$

Proof. Exercise. □

This proposition comes in handy especially when we want to convert $\Diamond$ into $\Box$ (or vice versa), or remove double negations inside a formula. In what follows, we will mark applications of Proposition 3.19 by "A for B" whenever we re-write a formula $C(B)$ for $C(A)$. In other words, "A for B" abbreviates:

$\vdash C(A)$
$\vdash A \leftrightarrow B$
$\vdash C(B)$ by Proposition 3.19

For instance:

Proposition 3.20. $\mathbf{K} \vdash \neg\Box p \rightarrow \Diamond\neg p$

Proof.

1. $\mathbf{K} \vdash \Diamond\neg p \leftrightarrow \neg\Box\neg\neg p$ DUAL
2. $\mathbf{K} \vdash \neg\Box\neg\neg p \rightarrow \Diamond\neg p$ PL, 1
3. $\mathbf{K} \vdash \neg\Box p \rightarrow \Diamond\neg p$ p for $\neg\neg p$

□

In the above derivation, the final step "p for $\neg\neg p$" is short for

$\mathbf{K} \vdash \neg\Box\neg\neg p \rightarrow \Diamond\neg p$
$\mathbf{K} \vdash \neg\neg p \leftrightarrow p$ TAUT
$\mathbf{K} \vdash \neg\Box p \rightarrow \Diamond\neg p$ by Proposition 3.19

The roles of $C(q)$, A, and B in Proposition 3.19 are played here, respectively, by $\neg\Box q \rightarrow \Diamond\neg p$, $\neg\neg p$, and p.

When a formula contains a sub-formula $\neg\Diamond A$, we can replace it by $\Box\neg A$ using Proposition 3.19, since $\mathbf{K} \vdash \neg\Diamond A \leftrightarrow \Box\neg A$. We'll indicate this and similar replacements simply by "$\Box\neg$ for $\neg\Diamond$."

The following proposition justifies that we can establish derivability results schematically. E.g., the previous proposition does not just establish that $\mathbf{K} \vdash \neg\Box p \rightarrow \Diamond\neg p$, but $\mathbf{K} \vdash \neg\Box A \rightarrow \Diamond\neg A$ for arbitrary A.

Proposition 3.21. *If A is a substitution instance of B and $\mathbf{K} \vdash B$, then $\mathbf{K} \vdash A$.*

Proof. It is tedious but routine to verify (by induction on the length of the derivation of B) that applying a substitution to an entire derivation also results in a correct derivation. Specifically, substitution instances of tautological instances are themselves tautological instances, substitution instances of instances of DUAL and K are themselves instances of DUAL and K, and applications of MP and NEC remain correct when substituting formulas for propositional variables in both premise(s) and conclusion. □

3.6 More Proofs in K

Let's see some more examples of derivability in **K**, now using the simplified method introduced in section 3.5.

Proposition 3.22. $\mathbf{K} \vdash \Box(A \to B) \to (\Diamond A \to \Diamond B)$

Proof.

1.	$\mathbf{K} \vdash (A \to B) \to (\neg B \to \neg A)$	PL
2.	$\mathbf{K} \vdash \Box(A \to B) \to (\Box\neg B \to \Box\neg A)$	RK, 1
3.	$\mathbf{K} \vdash (\Box\neg B \to \Box\neg A) \to (\neg\Box\neg A \to \neg\Box\neg B)$	TAUT
4.	$\mathbf{K} \vdash (\Box\neg B \to \Box\neg A) \to (\neg\Box\neg A \to \neg\Box\neg B)$	PL, 2, 3
5.	$\mathbf{K} \vdash \Box(A \to B) \to (\Diamond A \to \Diamond B)$	$\Diamond$ for $\neg\Box\neg$. □

Proposition 3.23. $\mathbf{K} \vdash \Box A \to (\Diamond(A \to B) \to \Diamond B)$

Proof.

1.	$\mathbf{K} \vdash A \to (\neg B \to \neg(A \to B))$	TAUT
2.	$\mathbf{K} \vdash \Box A \to (\Box\neg B \to \Box\neg(A \to B))$	RK, 1
3.	$\mathbf{K} \vdash \Box A \to (\neg\Box\neg(A \to B) \to \neg\Box\neg B)$	PL, 2
4.	$\mathbf{K} \vdash \Box A \to (\Diamond(A \to B) \to \Diamond B)$	$\Diamond$ for $\neg\Box\neg$. □

Proposition 3.24. $\mathbf{K} \vdash (\Diamond A \vee \Diamond B) \to \Diamond(A \vee B)$

Proof.

1.	$\mathbf{K} \vdash \neg(A \vee B) \to \neg A$	TAUT
2.	$\mathbf{K} \vdash \Box\neg(A \vee B) \to \Box\neg A$	RK, 1
3.	$\mathbf{K} \vdash \neg\Box\neg A \to \neg\Box\neg(A \vee B)$	PL, 2
4.	$\mathbf{K} \vdash \Diamond A \to \Diamond(A \vee B)$	$\Diamond$ for $\neg\Box\neg$
5.	$\mathbf{K} \vdash \Diamond B \to \Diamond(A \vee B)$	similarly
6.	$\mathbf{K} \vdash (\Diamond A \vee \Diamond B) \to \Diamond(A \vee B)$	PL, 4, 5. □

Proposition 3.25. $\mathbf{K} \vdash \Diamond(A \vee B) \to (\Diamond A \vee \Diamond B)$

Proof.

1.	$\mathbf{K} \vdash \neg A \to (\neg B \to \neg(A \vee B)$	TAUT
2.	$\mathbf{K} \vdash \Box\neg A \to (\Box\neg B \to \Box\neg(A \vee B)$	RK
3.	$\mathbf{K} \vdash \Box\neg A \to (\neg\Box\neg(A \vee B) \to \neg\Box\neg B))$	PL, 2
4.	$\mathbf{K} \vdash \neg\Box\neg(A \vee B) \to (\Box\neg A \to \neg\Box\neg B)$	PL, 3
5.	$\mathbf{K} \vdash \neg\Box\neg(A \vee B) \to (\neg\neg\Box\neg B \to \neg\Box\neg A)$	PL, 4
6.	$\mathbf{K} \vdash \Diamond(A \vee B) \to (\neg\Diamond B \to \Diamond A)$	$\Diamond$ for $\neg\Box\neg$
7.	$\mathbf{K} \vdash \Diamond(A \vee B) \to (\Diamond B \vee \Diamond A)$	PL, 6. □

3.7 Dual Formulas

Definition 3.26. Each of the formulas T, B, 4, and 5 has a *dual*, denoted by a subscripted diamond, as follows:

$$p \to \Diamond p \qquad (\mathrm{T}_\Diamond)$$
$$\Diamond\Box p \to p \qquad (\mathrm{B}_\Diamond)$$
$$\Diamond\Diamond p \to \Diamond p \qquad (4_\Diamond)$$
$$\Diamond\Box p \to \Box p \qquad (5_\Diamond)$$

Each of the above dual formulas is obtained from the corresponding formula by substituting $\neg p$ for p, contraposing, replacing $\neg\Box\neg$ by $\Diamond$, and replacing $\neg\Diamond\neg$ by $\Box$. D, i.e., $\Box A \to \Diamond A$ is its own dual in that sense.

3.8 Proofs in Modal Systems

We now come to proofs in systems of modal logic other than **K**.

Proposition 3.27. *The following provability results obtain:*

1. $\mathbf{KT5} \vdash$ B;
2. $\mathbf{KT5} \vdash 4$;
3. $\mathbf{KDB4} \vdash$ T;
4. $\mathbf{KB4} \vdash 5$;
5. $\mathbf{KB5} \vdash 4$;
6. $\mathbf{KT} \vdash$ D.

Proof. We exhibit proofs for each.

1. $\mathbf{KT5} \vdash$ B:

1.	$\mathbf{KT5} \vdash \Diamond A \to \Box\Diamond A$	5
2.	$\mathbf{KT5} \vdash A \to \Diamond A$	$\mathrm{T}_\Diamond$
3.	$\mathbf{KT5} \vdash A \to \Box\Diamond A$	PL.

2. $\mathbf{KT5} \vdash 4$:

1.	$\mathbf{KT5} \vdash \Diamond\Box A \to \Box\Diamond\Box A$	5 with $\Box A$ for p
2.	$\mathbf{KT5} \vdash \Box A \to \Diamond\Box A$	$\mathrm{T}_\Diamond$ with $\Box A$ for p
3.	$\mathbf{KT5} \vdash \Box A \to \Box\Diamond\Box A$	PL, 1, 2
4.	$\mathbf{KT5} \vdash \Diamond\Box A \to \Box A$	$5_\Diamond$
5.	$\mathbf{KT5} \vdash \Box\Diamond\Box A \to \Box\Box A$	RK, 4
6.	$\mathbf{KT5} \vdash \Box A \to \Box\Box A$	PL, 3, 5.

3. **KDB4** $\vdash$ T:

1.	**KDB4** $\vdash \Diamond\Box A \to A$	$B_\Diamond$
2.	**KDB4** $\vdash \Box\Box A \to \Diamond\Box A$	D with $\Box A$ for p
3.	**KDB4** $\vdash \Box\Box A \to A$	PL1, 2
4.	**KDB4** $\vdash \Box A \to \Box\Box A$	4
5.	**KDB4** $\vdash \Box A \to A$	PL, 1, 4.

4. **KB4** $\vdash$ 5:

1.	**KB4** $\vdash \Diamond A \to \Box\Diamond\Diamond A$	B with $\Diamond A$ for p
2.	**KB4** $\vdash \Diamond\Diamond A \to \Diamond A$	$4_\Diamond$
3.	**KB4** $\vdash \Box\Diamond\Diamond A \to \Box\Diamond A$	RK, 2
4.	**KB4** $\vdash \Diamond A \to \Box\Diamond A$	PL, 1, 3.

5. **KB5** $\vdash$ 4:

1.	**KB5** $\vdash \Box A \to \Box\Diamond\Box A$	B with $\Box A$ for p
2.	**KB5** $\vdash \Diamond\Box A \to \Box A$	$5_\Diamond$
3.	**KB5** $\vdash \Box\Diamond\Box A \to \Box\Box A$	RK, 2
4.	**KB5** $\vdash \Box A \to \Box\Box A$	PL, 1, 3.

6. **KT** $\vdash$ D:

1.	**KT** $\vdash \Box A \to A$	T
2.	**KT** $\vdash A \to \Diamond A$	$T_\Diamond$
3.	**KT** $\vdash \Box A \to \Diamond A$	PL, 1, 2

□

Definition 3.28. Following tradition, we define **S4** to be the system **KT4**, and **S5** the system **KTB4**.

The following proposition shows that the classical system **S5** has several equivalent axiomatizations. This should not surprise, as the various combinations of axioms all characterize equivalence relations (see Proposition 2.12).

Proposition 3.29. **KTB4 = KT5 = KDB4 = KDB5.**

Proof. Exercise. □

3.9 Soundness

A derivation system is called sound if everything that can be derived is valid. When considering modal systems, i.e., derivations where in addition to K we can use instances of some formulas A_1, ..., A_n, we want every derivable formula to be true in any model in which A_1, ..., A_n are true.

Theorem 3.30 (Soundness Theorem). *If every instance of A_1, ..., A_n is valid in the classes of models $\mathscr{C}_1$, ..., $\mathscr{C}_n$, respectively, then $\mathbf{K}A_1 \dots A_n \vdash B$ implies that B is valid in the class of models $\mathscr{C}_1 \cap \dots \cap \mathscr{C}_n$.*

Proof. By induction on length of proofs. For brevity, put $\mathscr{C} = \mathscr{C}_n \cap \dots \cap \mathscr{C}_n$.

1. Induction Basis: If B has a proof of length 1, then it is either a tautological instance, an instance of K, or of DUAL, or an instance of one of A_1, ..., A_n. In the first case, B is valid in $\mathscr{C}$, since tautological instance are valid in *any* class of models, by Proposition 1.15. Similarly in the second case, by Proposition 1.18 and Proposition 1.19. Finally in the third case, since B is valid in $\mathscr{C}_i$ and $\mathscr{C} \subseteq \mathscr{C}_i$, we have that B is valid in $\mathscr{C}$ as well.

2. Inductive step: Suppose B has a proof of length $k > 1$. If B is a tautological instance or an instance of one of A_1, ..., A_n, we proceed as in the previous step. So suppose B is obtained by MP from previous formulas $C \to B$ and C. Then $C \to B$ and C have proofs of length $< k$, and by inductive hypothesis they are valid in $\mathscr{C}$. By Proposition 1.20, B is valid in $\mathscr{C}$ as well. Finally suppose B is obtained by NEC

from C (so that $B = \Box C$). By inductive hypothesis, C is valid in $\mathscr{C}$, and by Proposition 1.12 so is B. □

3.10 Showing Systems are Distinct

In section 3.8 we saw how to prove that two systems of modal logic are in fact the same system. Theorem 3.30 allows us to show that two modal systems Σ and Σ' are distinct, by finding a formula A such that $\Sigma' \vdash A$ that fails in a model of Σ.

Proposition 3.31. KD $\subsetneq$ KT

Proof. This is the syntactic counterpart to the semantic fact that all reflexive relations are serial. To show **KD** $\subseteq$ **KT** we need to see that **KD** $\vdash B$ implies **KT** $\vdash B$, which follows from **KT** $\vdash$ **D**, as shown in Proposition 3.27(6). To show that the inclusion is proper, by Soundness (Theorem 3.30), it suffices to exhibit a model of **KD** where T, i.e., $\Box p \to p$, fails (an easy task left as an exercise), for then by Soundness **KD** $\nvdash \Box p \to p$. □

Proposition 3.32. KB $\neq$ K4.

Proof. We construct a symmetric model where some instance of 4 fails; since obviously the instance is derivable for **K4** but not in **KB**, it will follow **K4** $\not\subseteq$ **KB**. Consider the symmetric model M of Figure 3.1. Since the model is symmetric, K and B are true in M (by Proposition 1.18 and Theorem 2.1, respectively). However, $M, w_1 \nVdash \Box p \to \Box\Box p$. □

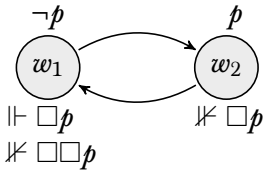


Figure 3.1: A symmetric model falsifying an instance of 4.

Theorem 3.33. **KTB $\nvdash$ 4** *and* **KTB $\nvdash$ 5.**

Proof. By Theorem 2.1 we know that all instances of T and B are true in every reflexive symmetric model (respectively). So by soundness, it suffices to find a reflexive symmetric model containing a world at which some instance of 4 fails, and similarly for 5. We use the same model for both claims. Consider the symmetric, reflexive model in Figure 3.2. Then $M, w_1 \nVdash \Box p \to \Box\Box p$, so 4 fails at w_1. Similarly, $M, w_2 \nVdash \Diamond\neg p \to \Box\Diamond\neg p$, so the instance of 5 with $A = \neg p$ fails at w_2. □

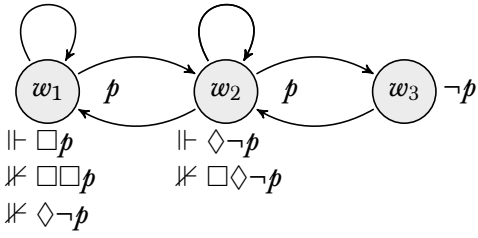


Figure 3.2: The model for Theorem 3.33.

Theorem 3.34. **KD5 $\neq$ KT4 = S4.**

Proof. By Theorem 2.1 we know that all instances of D and 5 are true in all serial euclidean models. So it suffices to find a serial euclidean model containing a world at which some instance of 4 fails. Consider the model of Figure 3.3, and notice that $M, w_1 \nVdash \Box p \to \Box\Box p$. □

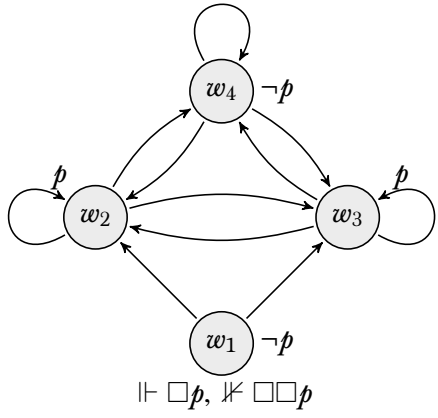


Figure 3.3: The model for Theorem 3.34.

3.11 Derivability from a Set of Formulas

In section 3.8 we defined a notion of provability of a formula in a system Σ. We now extend this notion to provability in Σ from formulas in a set Γ.

Definition 3.35. A formula A is derivable in a system Σ from a set of formulas Γ, written $\Gamma \vdash_\Sigma A$ if and only if there are $B_1, \ldots, B_n \in \Gamma$ such that $\Sigma \vdash B_1 \to (B_2 \to \cdots (B_n \to A) \cdots)$.

3.12 Properties of Derivability

Proposition 3.36. *Let Σ be a modal system and Γ a set of modal formulas. The following properties hold:*

1. Monotony*: If $\Gamma \vdash_\Sigma A$ and $\Gamma \subseteq \Delta$ then $\Delta \vdash_\Sigma A$;*
2. Reflexivity*: If $A \in \Gamma$ then $\Gamma \vdash_\Sigma A$;*
3. Cut*: If $\Gamma \vdash_\Sigma A$ and $\Delta \cup \{A\} \vdash_\Sigma B$ then $\Gamma \cup \Delta \vdash_\Sigma B$;*

4. Deduction theorem: *$\Gamma \cup \{B\} \vdash_\Sigma A$ if and only if $\Gamma \vdash_\Sigma B \to A$;*

5. *$\Gamma \vdash_\Sigma A_1$ and ... and $\Gamma \vdash_\Sigma A_n$ and $A_1 \to (A_2 \to \cdots (A_n \to B) \cdots)$ is a tautological instance, then $\Gamma \vdash_\Sigma B$.*

The proof is an easy exercise. Part (5) of Proposition 3.36 gives us that, for instance, if $\Gamma \vdash_\Sigma A \vee B$ and $\Gamma \vdash_\Sigma \neg A$, then $\Gamma \vdash_\Sigma B$. Also, in what follows, we write $\Gamma, A \vdash_\Sigma B$ instead of $\Gamma \cup \{A\} \vdash_\Sigma B$.

Definition 3.37. A set Γ is *deductively closed* relatively to a system Σ if and only if $\Gamma \vdash_\Sigma A$ implies $A \in \Gamma$.

3.13 Consistency

Consistency is an important property of sets of formulas. A set of formulas is inconsistent if a contradiction, such as $\bot$, is derivable from it; and otherwise consistent. If a set is inconsistent, its formulas cannot all be true in a model at a world. For the completeness theorem we prove the converse: every consistent set is true at a world in a model, namely in the "canonical model."

Definition 3.38. A set Γ is *consistent* relatively to a system Σ or, as we will say, Σ-consistent, if and only if $\Gamma \nvdash_\Sigma \bot$.

So for instance, the set $\{\Box(p \to q), \Box p, \neg\Box q\}$ is consistent relatively to propositional logic, but not **K**-consistent. Similarly, the set $\{\Diamond p, \Box\Diamond p \to q, \neg q\}$ is not **K5**-consistent.

Proposition 3.39. *Let Γ be a set of formulas. Then:*

1. *A set Γ is Σ-consistent if and only if there is some formula A such that $\Gamma \nvdash_\Sigma A$.*

2. *$\Gamma \vdash_\Sigma A$ if and only if $\Gamma \cup \{\neg A\}$ is not Σ-consistent.*

3. If Γ is Σ-consistent, then for any formula A, either $\Gamma \cup \{A\}$ is Σ-consistent or $\Gamma \cup \{\neg A\}$ is Σ-consistent.

Proof. These facts follow easily using classical propositional logic. We give the argument for (3). Proceed contrapositively and suppose neither $\Gamma \cup \{A\}$ nor $\Gamma \cup \{\neg A\}$ is Σ-consistent. Then by (2), both $\Gamma, A \vdash_\Sigma \bot$ and $\Gamma, \neg A \vdash_\Sigma \bot$. By the deduction theorem $\Gamma \vdash_\Sigma A \to \bot$ and $\Gamma \vdash_\Sigma \neg A \to \bot$. But $(A \to \bot) \to ((\neg A \to \bot) \to \bot)$ is a tautological instance, hence by Proposition 3.36(5), $\Gamma \vdash_\Sigma \bot$. □

Problems

Problem 3.1. Prove Proposition 3.7.

Problem 3.2. Find derivations in **K** for the following formulas:

1. $\Box\neg p \to \Box(p \to q)$
2. $(\Box p \vee \Box q) \to \Box(p \vee q)$
3. $\Diamond p \to \Diamond(p \vee q)$

Problem 3.3. Prove Proposition 3.19 by proving, by induction on the complexity of C, that if $\mathbf{K} \vdash A \leftrightarrow B$ then $\mathbf{K} \vdash C[A/q] \leftrightarrow C[B/q]$.

Problem 3.4. Show that the following derivability claims hold:

1. $\mathbf{K} \vdash \Diamond\neg\bot \to (\Box A \to \Diamond A)$;
2. $\mathbf{K} \vdash \Box(A \vee B) \to (\Diamond A \vee \Box B)$;
3. $\mathbf{K} \vdash (\Diamond A \to \Box B) \to \Box(A \to B)$.

Problem 3.5. Show that for each formula A in Definition 3.26: $\mathbf{K} \vdash A \leftrightarrow A_\Diamond$.

Problem 3.6. Prove Proposition 3.29.

Problem 3.7. Give an alternative proof of Theorem 3.34 using a model with 3 worlds.

Problem 3.8. Provide a single reflexive transitive model showing that both **KT4** $\nvdash$ B and **KT4** $\nvdash$ 5.

CHAPTER 4

Completeness and Canonical Models

4.1 Introduction

If Σ is a modal system, then the soundness theorem establishes that if $\Sigma \vdash A$, then A is valid in any class $\mathscr{C}$ of models in which all instances of all formulas in Σ are valid. In particular that means that if $\mathbf{K} \vdash A$ then A is true in all models; if $\mathbf{KT} \vdash A$ then A is true in all reflexive models; if $\mathbf{KD} \vdash A$ then A is true in all serial models, etc.

Completeness is the converse of soundness: that $\mathbf{K}$ is complete means that if a formula A is valid, $\vdash A$, for instance. Proving completeness is a lot harder to do than proving soundness. It is useful, first, to consider the contrapositive: $\mathbf{K}$ is complete iff whenever $\nvdash A$, there is a countermodel, i.e., a model $\mathfrak{M}$ such that $\mathfrak{M} \nVdash A$. Equivalently (negating A), we could prove that whenever $\nvdash \neg A$, there is a model of A. In the construction of such a model, we can use information contained in A. When we find models for specific formulas we often do the same: E.g., if we want to

find a countermodel to $p \rightarrow \Box q$, we know that it has to contain a world where p is true and $\Box q$ is false. And a world where $\Box q$ is false means there has to be a world accessible from it where q is false. And that's all we need to know: which worlds make the propositional variables true, and which worlds are accessible from which worlds.

In the case of proving completeness, however, we don't have a specific formula A for which we are constructing a model. We want to establish that a model exists for every A such that $\nvdash_\Sigma \neg A$. This is a minimal requirement, since *if* $\vdash_\Sigma \neg A$, by soundness, there is no model for A (in which Σ is true). Now note that $\nvdash_\Sigma \neg A$ iff A is Σ-consistent. (Recall that $\Sigma \nvdash_\Sigma \neg A$ and $A \nvdash_\Sigma \bot$ are equivalent.) So our task is to construct a model for every Σ-consistent formula.

The trick we'll use is to find a Σ-consistent set of formulas that contains A, but also other formulas which tell us what the world that makes A true has to look like. Such sets are *complete* Σ-consistent sets. It's not enough to construct a model with a single world to make A true, it will have to contain multiple worlds and an accessibility relation. The complete Σ-consistent set containing A will also contain other formulas of the form $\Box B$ and $\Diamond C$. In all accessible worlds, B has to be true; in at least one, C has to be true. In order to accomplish this, we'll simply take *all* possible complete Σ-consistent sets as the basis for the set of worlds. A tricky part will be to figure out when a complete Σ-consistent set should count as being accessible from another in our model.

We'll show that in the model so defined, A is true at a world—which is also a complete Σ-consistent set—iff A is an element of that set. If A is Σ-consistent, it will be an element of at least one complete Σ-consistent set (a fact we'll prove), and so there will be a world where A is true. So we will have a single model where every Σ-consistent formula A is true at some world. This single model is the *canonical* model for Σ.

4.2 Complete Σ-Consistent Sets

Suppose Σ is a set of modal formulas—think of them as the axioms or defining principles of a normal modal logic. A set Γ is Σ-consistent iff $\Gamma \nvdash_\Sigma \bot$, i.e., if there is no derivation of $A_1 \to (A_2 \to \cdots (A_n \to \bot)\dots)$ from Σ, where each $A_i \in \Gamma$. We will construct a "canonical" model in which each world is taken to be a special kind of Σ-consistent set: one which is not just Σ-consistent, but maximally so, in the sense that it settles the truth value of every modal formula: for every A, either $A \in \Gamma$ or $\neg A \in \Gamma$:

Definition 4.1. A set Γ is *complete Σ-consistent* if and only if it is Σ-consistent and for every A, either $A \in \Gamma$ or $\neg A \in \Gamma$.

Complete Σ-consistent sets Γ have a number of useful properties. For one, they are deductively closed, i.e., if $\Gamma \vdash_\Sigma A$ then $A \in \Gamma$. This means in particular that every instance of a formula $A \in \Sigma$ is also $\in \Gamma$. Moreover, membership in Γ mirrors the truth conditions for the propositional connectives. This will be important when we define the "canonical model."

Proposition 4.2. *Suppose Γ is complete Σ-consistent. Then:*

1. *Γ is deductively closed in Σ.*
2. *$\Sigma \subseteq \Gamma$.*
3. *$\bot \notin \Gamma$*
4. *$\neg A \in \Gamma$ if and only if $A \notin \Gamma$.*
5. *$A \wedge B \in \Gamma$ iff $A \in \Gamma$ and $B \in \Gamma$*
6. *$A \vee B \in \Gamma$ iff $A \in \Gamma$ or $B \in \Gamma$*
7. *$A \to B \in \Gamma$ iff $A \notin \Gamma$ or $B \in \Gamma$*

Proof. 1. Suppose $\Gamma \vdash_\Sigma A$ but $A \notin \Gamma$. Then since Γ is complete Σ-consistent, $\neg A \in \Gamma$. This would make Γ inconsistent, since $A, \neg A \vdash_\Sigma \bot$.

2. If $A \in \Sigma$ then $\Gamma \vdash_\Sigma A$, and $A \in \Gamma$ by deductive closure, i.e., case (1).

3. If $\bot \in \Gamma$, then $\Gamma \vdash_\Sigma \bot$, so Γ would be Σ-inconsistent.

4. If $\neg A \in \Gamma$, then by consistency $A \notin \Gamma$; and if $A \notin \Gamma$ then $A \in \Gamma$ since Γ is complete Σ-consistent.

5. Exercise.

6. Suppose $A \vee B \in \Gamma$, and $A \notin \Gamma$ and $B \notin \Gamma$. Since Γ is complete Σ-consistent, $\neg A \in \Gamma$ and $\neg B \in \Gamma$. Then $\neg(A \vee B) \in \Gamma$ since $\neg A \rightarrow (\neg B \rightarrow \neg(A \vee B))$ is a tautological instance. This would mean that Γ is Σ-inconsistent, a contradiction.

7. Exercise.

4.3 Lindenbaum's Lemma

Lindenbaum's Lemma establishes that every Σ-consistent set of formulas is contained in at least one *complete* Σ-consistent set. Our construction of the canonical model will show that for each complete Σ-consistent set Δ, there is a world in the canonical model where all and only the formulas in Δ are true. So Lindenbaum's Lemma guarantees that every Σ-consistent set is true at some world in the canonical model.

Theorem 4.3 (Lindenbaum's Lemma). *If Γ is Σ-consistent then there is a complete Σ-consistent set Δ extending Γ.*

Proof. Let A_0, A_1, ... be an exhaustive listing of all formulas of the language (repetitions are allowed). For instance, start by listing p_0, and at each stage $n \geq 1$ list the finitely many formulas of length n using only variables among $p_0, \ldots, p_n$. We define sets of formulas Δ_n by induction on n, and we then set $\Delta = \bigcup_n \Delta_n$. We first put $\Delta_0 = \Gamma$. Supposing that Δ_n has been defined, we define Δ_{n+1} by:

$$\Delta_{n+1} = \begin{cases} \Delta_n \cup \{A_n\}, & \text{if } \Delta_n \cup \{A_n\} \text{ is consistent;} \\ \Delta_n \cup \{\neg A_n\}, & \text{otherwise.} \end{cases}$$

If we now let $\Delta = \bigcup_{n=0}^{\infty} \Delta_n$.

We have to show that this definition actually yields a set Δ with the required properties, i.e., $\Gamma \subseteq \Delta$ and Δ is complete Σ-consistent.

It's obvious that $\Gamma \subseteq \Delta$, since $\Delta_0 \subseteq \Delta$ by construction, and $\Delta_0 = \Gamma$. In fact, $\Delta_n \subseteq \Delta$ for all n, since Δ is the union of all Δ_n. (Since in each step of the construction, we add a formula to the set already constructed, $\Delta_n \subseteq \Delta_{n+1}$, so since $\subseteq$ is transitive, $\Delta_n \subseteq \Delta_m$ whenever $n \leq m$.) At each stage of the construction, we either add A_n or $\neg A_n$, and every formula appears (at least once) in the list of all A_n. So, for every A either $A \in \Delta$ or $\neg A \in \Delta$, so Δ is complete by definition.

Finally, we have to show, that Δ is Σ-consistent. To do this, we show that (a) if Δ were Σ-inconsistent, then some Δ_n would be Σ-inconsistent, and (b) all Δ_n are Σ-consistent.

So suppose Δ were Σ-inconsistent. Then $\Delta \vdash_\Sigma \bot$, i.e., there are $A_1, \ldots, A_k \in \Delta$ such that $\Sigma \vdash A_1 \to (A_2 \to \cdots (A_k \to \bot) \ldots)$. Since $\Delta = \bigcap_{n=0}^{\infty}$, each $A_i \in \Delta_{n_i}$ for some n_i. Let n be the largest of these. Since $n_i \leq n$, $\Delta_{n_i} \subseteq \Delta_n$. So, all A_i are in some Δ_n. This would mean $\Delta_n \vdash_\Sigma \bot$, i.e., Δ_n is Σ-inconsistent.

To show that each Δ_n is Σ-consistent, we use a simple induction on n. $\Delta_0 = \Gamma$, and we assumed Γ was Σ-consistent. So

the claim holds for $n = 0$. Now suppose it holds for n, i.e., Δ_n is Σ-consistent. Δ_{n+1} is either $\Delta_n \cup \{A_n\}$ is that is Σ-consistent, otherwise it is $\Delta_n \cup \{\neg A_n\}$. In the first case, Δ_{n+1} is clearly Σ-consistent. However, by Proposition 3.39(3), either $\Delta_n \cup \{A_n\}$ or $\Delta_n \cup \{\neg A_n\}$ is consistent, so Δ_{n+1} is consistent in the other case as well. □

Corollary 4.4. *$\Gamma \vdash_\Sigma A$ if and only if $A \in \Delta$ for each complete Σ-consistent set Δ extending Γ (including when $\Gamma = \emptyset$, in which case we get another characterization of the modal system Σ.)*

Proof. Suppose $\Gamma \vdash_\Sigma A$, and let Δ be any complete Σ-consistent set extending Γ. If $A \notin \Delta$ then by maximality $\neg A \in \Delta$ and so $\Delta \vdash_\Sigma A$ (by monotony) and $\Delta \vdash_\Sigma \neg A$ (by reflexivity), and so Δ is inconsistent. Conversely if $\Gamma \nvdash_\Sigma A$, then $\Gamma \cup \{\neg A\}$ is Σ-consistent, and by Lindenbaum's Lemma there is a complete consistent set Δ extending $\Gamma \cup \{\neg A\}$. By consistency, $A \notin \Delta$. □

4.4 Modalities and Complete Consistent Sets

When we construct a model M^Σ whose set of worlds is given by the complete Σ-consistent sets Δ in some normal modal logic Σ, we will also need to define an accessibility relation R^Σ between such "worlds." We want it to be the case that the accessibility relation (and the assignment V^Σ) are defined in such a way that $M^\Sigma, \Delta \Vdash A$ iff $A \in \Delta$. How should we do this?

Once the accessibility relation is defined, the definition of truth at a world ensures that $M^\Sigma, \Delta \Vdash \Box A$ iff $M^\Sigma, \Delta' \Vdash A$ for all Δ' such that $R^\Sigma \Delta\Delta'$. The proof that $M^\Sigma, \Delta \Vdash A$ iff $A \in \Delta$ requires that this is true in particular for formulas starting with a modal operator, i.e., $M^\Sigma, \Delta \Vdash \Box A$ iff $\Box A \in \Delta$. Combining this requirement with the definition of truth at a world for $\Box A$ yields:

$$\Box A \in \Delta \text{ iff } A \in \Delta' \text{ for all } \Delta' \text{ with } R^\Sigma \Delta\Delta'$$

Consider the left-to-right direction: it says that if $\Box A \in \Delta$, then $A \in \Delta'$ for any A and any Δ' with $R^\Sigma \Delta\Delta'$. If we stipulate that $R^\Sigma \Delta\Delta'$ iff $A \in \Delta'$ for all $\Box A \in \Delta$, then this holds. We can write the condition on the right of the "iff" more compactly as: $\{A : \Box A \in \Delta\} \subseteq \Delta'$.

So the question is: does this definition of R^Σ in fact guarantee that $\Box A \in \Delta$ iff $M^\Sigma, \Delta \Vdash \Box A$? Does it also guarantee that $\Diamond A \in \Delta$ iff $M^\Sigma, \Delta \Vdash \Diamond A$? The next few results will establish this.

Definition 4.5. If Γ is a set of formulas, let

$$\Box\Gamma = \{\Box B : B \in \Gamma\}$$
$$\Diamond\Gamma = \{\Diamond B : B \in \Gamma\}$$

and

$$\Box^{-1}\Gamma = \{B : \Box B \in \Gamma\}$$
$$\Diamond^{-1}\Gamma = \{B : \Diamond B \in \Gamma\}$$

In other words, $\Box\Gamma$ is Γ with $\Box$ in front of every formula in Γ; $\Box^{-1}\Gamma$ is all the $\Box$'ed formulas of Γ with the initial $\Box$'s removed. This definition is not terribly important on its own, but will simplify the notation considerably.

Note that $\Box\Box^{-1}\Gamma \subseteq \Gamma$:

$$\Box\Box^{-1}\Gamma = \{\Box B : \Box B \in \Gamma\}$$

i.e., it's just the set of all those formulas of Γ that start with $\Box$.

Lemma 4.6. *If $\Gamma \vdash_\Sigma A$ then $\Box\Gamma \vdash_\Sigma \Box A$.*

Proof. If $\Gamma \vdash_\Sigma A$ then there are $B_1, \ldots, B_k \in \Gamma$ such that $\Sigma \vdash B_1 \to (B_2 \to \cdots (B_n \to A)\cdots)$. Since Σ is normal, by rule RK, $\Sigma \vdash \Box B_1 \to (\Box B_2 \to \cdots (\Box B_n \to \Box A)\cdots)$, where obviously $\Box B_1, \ldots, \Box B_k \in \Box\Gamma$. Hence, by definition, $\Box\Gamma \vdash_\Sigma \Box A$. □

Lemma 4.7. *If $\Box^{-1}\Gamma \vdash_\Sigma A$ then $\Gamma \vdash_\Sigma \Box A$.*

Proof. Suppose $\Box^{-1}\Gamma \vdash_\Sigma A$; then by Lemma 4.6, $\Box\Box^{-1}\Gamma \vdash \Box A$. But since $\Box\Box^{-1}\Gamma \subseteq \Gamma$, also $\Gamma \vdash_\Sigma \Box A$ by Monotony. □

Proposition 4.8. *If Γ is complete Σ-consistent, then $\Box A \in \Gamma$ if and only if for every complete Σ-consistent Δ such that $\Box^{-1}\Gamma \subseteq \Delta$, it holds that $A \in \Delta$.*

Proof. Suppose Γ is complete Σ-consistent. The "only if" direction is easy: Suppose $\Box A \in \Gamma$ and that $\Box^{-1}\Gamma \subseteq \Delta$. Since $\Box A \in \Gamma$, $A \in \Box^{-1}\Gamma \subseteq \Delta$, so $A \in \Delta$.

For the "if" direction, we prove the contrapositive: Suppose $\Box A \notin \Gamma$. Since Γ is complete Σ-consistent, it is deductively closed, and hence $\Gamma \nvdash_\Sigma \Box A$. By Lemma 4.7, $\Box^{-1}\Gamma \nvdash_\Sigma A$. By Proposition 3.39(2), $\Box^{-1}\Gamma \cup \{\neg A\}$ is Σ-consistent. By Lindenbaum's Lemma, there is a complete Σ-consistent set Δ such that $\Box^{-1}\Gamma \cup \{\neg A\} \subseteq \Delta$. By consistency, $A \notin \Delta$. □

Lemma 4.9. *Suppose Γ and Δ are complete Σ-consistent. Then: $\Box^{-1}\Gamma \subseteq \Delta$ if and only if $\Diamond\Delta \subseteq \Gamma$.*

Proof. "Only if" direction: Assume $\Box^{-1}\Gamma \subseteq \Delta$ and suppose $\Diamond A \in \Diamond\Delta$ (i.e., $A \in \Delta$). In order to show $\Diamond A \in \Gamma$ it suffices to show $\Box\neg A \notin \Gamma$ for then by maximality $\neg\Box\neg A \in \Gamma$. Now, if $\Box\neg A \in \Gamma$ then by hypothesis $\neg A \in \Delta$, against the consistency of Δ (since $A \in \Delta$). Hence $\Box\neg A \notin \Gamma$, as required.

"If" direction: Assume $\Diamond\Delta \subseteq \Gamma$. We argue contrapositively: suppose $A \notin \Delta$ in order to show $\Box A \notin \Gamma$. If $A \notin \Delta$ then by maximality $\neg A \in \Delta$ and so by hypothesis $\Diamond\neg A \in \Gamma$. But in a normal modal logic $\Diamond\neg A$ is equivalent to $\neg\Box A$, and if the latter is in Γ, by consistency $\Box A \notin \Gamma$, as required. □

Proposition 4.10. *If Γ is complete Σ-consistent, then $\Diamond A \in \Gamma$ if and only if for some complete Σ-consistent Δ such that $\Diamond\Delta \subseteq \Gamma$, it holds that $A \in \Delta$.*

Proof. Suppose Γ is complete Σ-consistent. $\Diamond A \in \Gamma$ iff $\neg\Box\neg A \in \Gamma$ by DUAL and closure. $\neg\Box\neg A \in \Gamma$ iff $\Box\neg A \notin \Gamma$ by Proposition 4.2(4) since Γ is complete Σ-consistent. By Proposition 4.8, $\Box\neg A \notin \Gamma$ iff, for some complete Σ-consistent Δ with $\Box^{-1}\Gamma \subseteq \Delta$, $\neg A \notin \Delta$. Now consider any such Δ. By Lemma 4.9, $\Box^{-1}\Gamma \subseteq \Delta$ iff $\Diamond\Delta \subseteq \Gamma$. Also, $\neg A \notin \Delta$ iff $A \in \Delta$ by Proposition 4.2(4). So $\Diamond A \in \Gamma$ iff, for some complete Σ-consistent Δ with $\Diamond\Delta \subseteq \Gamma$, $A \in \Delta$. □

4.5 Canonical Models

The *canonical model* for a modal system Σ is a specific model M^Σ in which the worlds are all complete Σ-consistent sets. Its accessibility relation R^Σ and valuation V^Σ are defined so as to guarantee that the formulas true at a world Δ are exactly the formulas making up Δ.

Definition 4.11. Let Σ be a normal modal logic. The *canonical model* for Σ is $M^\Sigma = \langle W^\Sigma, R^\Sigma, V^\Sigma\rangle$, where:

1. $M^\Sigma = \{\Delta : \Delta \text{ is complete } \Sigma\text{-consistent}\}$.
2. $R^\Sigma\Delta\Delta'$ holds if and only if $\Box^{-1}\Delta \subseteq \Delta'$.
3. $V^\Sigma(p) = \{\Delta : p \in \Delta\}$.

4.6 The Truth Lemma

The canonical model M^Σ is defined in such a way that $M^\Sigma, \Delta \Vdash A$ iff $A \in \Delta$. For propositional variables, the definition of V^Σ yields this directly. We have to verify that the equivalence holds for all formulas, however. We do this by induction. The inductive step

involves proving the equivalence for formulas involving propositional operators (where we have to use Proposition 4.2) and the modal operators (where we invoke the results of section 4.4).

Proposition 4.12 (Truth Lemma). *For every formula A, $M^\Sigma, \Delta \Vdash A$ if and only if $A \in \Delta$.*

Proof. By induction on A.

1. $A \equiv \bot$: $M^\Sigma, \Delta \nVdash \bot$ by Definition 1.6, and $\bot \notin \Delta$ by Proposition 4.2(3).

2. $A \equiv p$: $M^\Sigma, \Delta \Vdash p$ iff $\Delta \in V^\Sigma(p)$ by Definition 1.6. Also, $\Delta \in V^\Sigma(p)$ iff $p \in \Delta$ by definition of V^Σ.

3. $A \equiv \neg B$: $M^\Sigma, \Delta \Vdash \neg B$ iff $M^\Sigma, \Delta \nVdash B$ (Definition 1.6) iff $B \notin \Delta$ (by inductive hypothesis) iff $\neg B \in \Delta$ (by Proposition 4.2(4)).

4. $A \equiv B \wedge C$: Exercise.

5. $A \equiv B \vee C$: $M^\Sigma, \Delta \Vdash B \vee C$ iff $M^\Sigma, \Delta \Vdash B$ or $M^\Sigma, \Delta \Vdash C$ (by Definition 1.6) iff $B \in \Delta$ or $C \in \Delta$ (by inductive hypothesis) iff $B \vee C \in \Delta$ (by Proposition 4.2(6)).

6. $A \equiv B \to C$: Exercise.

7. $A \equiv \Box B$: First suppose that $M^\Sigma, \Delta \Vdash \Box B$. By Definition 1.6, for every Δ' such that $R^\Sigma \Delta\Delta'$, $M^\Sigma, \Delta' \Vdash B$. By inductive hypothesis, for every Δ' such that $R^\Sigma \Delta\Delta'$, $B \in \Delta'$. By definition of R^Σ, for every Δ' such that $\Box^{-1}\Delta \subseteq \Delta'$, $B \in \Delta'$. By Proposition 4.8, $\Box B \in \Delta$.

   Now assume $\Box B \in \Delta$. Let $\Delta' \in W^\Sigma$ be such that $R^\Sigma \Delta\Delta'$, i.e., $\Box^{-1}\Delta \subseteq \Delta'$. Since $\Box B \in \Delta$, $B \in \Box^{-1}\Delta$. Consequently, $B \in \Delta'$. By inductive hypothesis, $M^\Sigma, \Delta' \Vdash B$. Since Δ' is arbitrary with $R^\Sigma \Delta\Delta'$, for all $\Delta' \in W^\Sigma$ such that $R^\Sigma \Delta\Delta'$, $M^\Sigma, \Delta' \Vdash B$. By Definition 1.6, $M^\Sigma, \Delta \Vdash \Box B$.

8. $A \equiv \Diamond B$: Exercise. □

4.7 Determination and Completeness for K

We are now prepared to use the canonical model to establish completeness. Completeness follows from the fact that the formulas true in the canonical for Σ are exactly the Σ-derivable ones. Models with this property are said to *determine* Σ.

Definition 4.13. A model M *determines* a normal modal logic Σ precisely when $M \Vdash A$ if and only if $\Sigma \vdash A$, for all formulas A.

Theorem 4.14 (Determination). *$M^\Sigma \Vdash A$ if and only if $\Sigma \vdash A$.*

Proof. If $M^\Sigma \Vdash A$, then for every complete Σ-consistent Δ, we have $M^\Sigma, \Delta \Vdash A$. Hence, by the Truth Lemma, $A \in \Delta$ for every complete Σ-consistent Δ, whence by Corollary 4.4 (with $\Gamma = \emptyset$), $\Sigma \vdash A$.

Conversely, if $\Sigma \vdash A$ then by Proposition 4.2(1), every complete Σ-consistent Δ contains A, and hence by the Truth Lemma, $M^\Sigma, \Delta \Vdash A$ for every $\Delta \in W^\Sigma$, i.e., $M^\Sigma \Vdash A$. □

Since the canonical model for **K** determines **K**, we immediately have completeness of **K** as a corollary:

Corollary 4.15. *The basic modal logic* **K** *is complete with respect to the class of all models, i.e., if $\vDash A$ then* $\mathbf{K} \vdash A$.

Proof. Contrapositively, if $\mathbf{K} \nvdash A$ then by Determination $M^\mathbf{K} \nVdash A$ and hence A is not valid. □

For the general case of completeness of a system Σ with respect to a class of models, e.g., of **KTB4** with respect to the class of reflexive, symmetric, transitive models, determination alone is not enough. We must also show that the canonical model for the system Σ is a member of the class, which does not follow obviously from the canonical model construction—nor is it always true!

4.8 Frame Completeness

The completeness theorem for **K** can be extended to other modal systems, once we show that the canonical model for a given logic has the corresponding frame property.

Theorem 4.16. *If a normal modal logic Σ contains one of the formulas on the left-hand side of table 4.1, then the canonical model for Σ has the corresponding property on the right-hand side.*

If Σ contains ...	*... the canonical model for Σ is:*
D: $\Box A \to \Diamond A$	serial;
T: $\Box A \to A$	reflexive;
B: $A \to \Box\Diamond A$	symmetric;
4: $\Box A \to \Box\Box A$	transitive;
5: $\Diamond A \to \Box\Diamond A$	euclidean.

Table 4.1: Basic correspondence facts.

Proof. We take each of these up in turn.

Suppose Σ contains D, and let $\Delta \in W^\Sigma$; we need to show that there is a Δ' such that $R^\Sigma \Delta\Delta'$. It suffices to show that $\Box^{-1}\Delta$ is Σ-consistent, for then by Lindenbaum's Lemma, there is a complete Σ-consistent set $\Delta' \supseteq \Box^{-1}\Delta$, and by definition of R^Σ we have $R^\Sigma \Delta\Delta'$. So, suppose for contradiction that $\Box^{-1}\Delta$ is *not* Σ-consistent, i.e., $\Box^{-1}\Delta \vdash_\Sigma \bot$. By Lemma 4.7, $\Delta \vdash_\Sigma \Box\bot$, and since Σ contains D, also $\Delta \vdash_\Sigma \Diamond\bot$. But Σ is normal, so $\Sigma \vdash \neg\Diamond\bot$ (Proposition 3.7), whence also $\Delta \vdash_\Sigma \neg\Diamond\bot$, against the consistency of Δ.

Now suppose Σ contains T, and let $\Delta \in W^\Sigma$. We want to show $R^\Sigma \Delta\Delta$, i.e., $\Box^{-1}\Delta \subseteq \Delta$. But if $\Box A \in \Delta$ then by T also $A \in \Delta$, as desired.

Now suppose Σ contains B, and suppose $R^\Sigma \Delta\Delta'$ for Δ, $\Delta' \in W^\Sigma$. We need to show that $R^\Sigma \Delta'\Delta$, i.e., $\Box^{-1}\Delta' \subseteq \Delta$. By Lemma 4.9, this is equivalent to $\Diamond\Delta \subseteq \Delta'$. So suppose $A \in \Delta$. By B, also $\Box\Diamond A \in \Delta$. By the hypothesis that $R^\Sigma \Delta\Delta'$, we have that $\Box^{-1}\Delta \subseteq \Delta'$, and hence $\Diamond A \in \Delta'$, as required.

Now suppose Σ contains 4, and suppose $R^\Sigma \Delta_1 \Delta_2$ and $R^\Sigma \Delta_2 \Delta_3$. We need to show $R^\Sigma \Delta_1 \Delta_3$. From the hypothesis we have both $\Box^{-1}\Delta_1 \subseteq \Delta_2$ and $\Box^{-1}\Delta_2 \subseteq \Delta_3$. In order to show $R^\Sigma \Delta_1 \Delta_3$ it suffices to show $\Box^{-1}\Delta_1 \subseteq \Delta_3$. So let $B \in \Box^{-1}\Delta_1$, i.e., $\Box B \in \Delta_1$. By 4, also $\Box\Box B \in \Delta_1$ and by hypothesis we get, first, that $\Box B \in \Delta_2$ and, second, that $B \in \Delta_3$, as desired.

Now suppose Σ contains 5, suppose $R^\Sigma \Delta_1 \Delta_2$ and $R^\Sigma \Delta_1 \Delta_3$. We need to show $R^\Sigma \Delta_2 \Delta_3$. The first hypothesis gives $\Box^{-1}\Delta_1 \subseteq \Delta_2$, and the second hypothesis is equivalent to $\Diamond\Delta_3 \subseteq \Delta_2$, by Lemma 4.9. To show $R^\Sigma \Delta_2 \Delta_3$, by Lemma 4.9, it suffices to show $\Diamond\Delta_3 \subseteq \Delta_2$. So let $\Diamond A \in \Diamond\Delta_3$, i.e., $A \in \Delta_3$. By the second hypothesis $\Diamond A \in \Delta_1$ and by 5, $\Box\Diamond A \in \Delta_1$ as well. But now the first hypothesis gives $\Diamond A \in \Delta_2$, as desired. □

As a corollary we obtain completeness results for a number of systems. For instance, we know that **S5** = **KT5** = **KTB4** is complete with respect to the class of all reflexive euclidean models, which is the same as the class of all reflexive, symmetric and transitive models.

Theorem 4.17. *Let $\mathcal{C}_{\mathrm{D}}$, $\mathcal{C}_{\mathrm{T}}$, $\mathcal{C}_{\mathrm{B}}$, $\mathcal{C}_4$, and $\mathcal{C}_5$ be the class of all serial, reflexive, symmetric, transitive, and euclidean models (respectively). Then for any schemas $A_1, \ldots, A_n$ among* D, T, B, 4, *and* 5, *the system* $\mathbf{K}A_1 \ldots A_n$ *is determined by the class of models* $\mathcal{C} = \mathcal{C}_{A_1} \cap \cdots \cap \mathcal{C}_{A_n}$.

Proposition 4.18. *Let Σ be a normal modal logic; then:*

1. *If Σ contains the schema $\Diamond A \to \Box A$ then the canonical model for Σ is partially functional.*
2. *If Σ contains the schema $\Diamond A \leftrightarrow \Box A$ then the canonical model for Σ is functional.*
3. *If Σ contains the schema $\Box\Box A \to \Box A$ then the canonical model for Σ is weakly dense.*

(see table 2.2 for definitions of these frame properties).

Proof. 1. Suppose that Σ contains the schema $\Diamond A \to \Box A$, to show that R^Σ is partially functional we need to prove that for any $\Delta_1, \Delta_2, \Delta_3 \in W^\Sigma$, if $R^\Sigma \Delta_1 \Delta_2$ and $R^\Sigma \Delta_1 \Delta_3$ then $\Delta_2 = \Delta_3$. Since $R^\Sigma \Delta_1 \Delta_2$ we have $\Box^{-1}\Delta_1 \subseteq \Delta_2$ and since $R^\Sigma \Delta_1 \Delta_3$ also $\Box^{-1}\Delta_1 \subseteq \Delta_3$. The identity $\Delta_2 = \Delta_3$ will follow if we can establish the two inclusions $\Delta_2 \subseteq \Delta_3$ and $\Delta_3 \subseteq \Delta_2$. For the first inclusion, let $A \in \Delta_2$; then $\Diamond A \in \Delta_1$, and by the schema and deductive closure of Δ_1 also $\Box A \in \Delta_1$, whence by the hypothesis that $R^\Sigma \Delta_1 \Delta_3$, $A \in \Delta_3$. The second inclusion is similar.

2. This follows immediately from part (1) and the seriality proof in Theorem 4.16.

3. Suppose Σ contains the schema $\Box\Box A \to \Box A$ and to show that R^Σ is weakly dense, let $R^\Sigma \Delta_1 \Delta_2$. We need to show that there is a complete Σ-consistent set Δ_3 such that $R^\Sigma \Delta_1 \Delta_3$ and $R^\Sigma \Delta_3 \Delta_2$. Let:

$$\Gamma = \Box^{-1}\Delta_1 \cup \Diamond\Delta_2.$$

It suffices to show that Γ is Σ-consistent, for then by Lindenbaum's Lemma it can be extended to a complete Σ-consistent set Δ_3 such that $\Box^{-1}\Delta_1 \subseteq \Delta_3$ and $\Diamond\Delta_2 \subseteq \Delta_3$, i.e., $R^\Sigma \Delta_1 \Delta_3$ and $R^\Sigma \Delta_3 \Delta_2$ (by Lemma 4.9).

Suppose for contradiction that Γ is not consistent. Then there are formulas $\Box A_1, \ldots, \Box A_n \in \Delta_1$ and $B_1, \ldots, B_m \in \Delta_2$ such that

$$A_1, \ldots, A_n, \Diamond B_1, \ldots, \Diamond B_m \vdash_\Sigma \bot.$$

Since $\Diamond(B_1 \wedge \cdots \wedge B_m) \to (\Diamond B_1 \wedge \cdots \wedge \Diamond B_m)$ is derivable in every normal modal logic, we argue as follows, contradicting the consistency of Δ_2:

$$A_1, \ldots, A_n, \Diamond B_1, \ldots, \Diamond B_m \vdash_\Sigma \bot$$

$$A_1, \ldots, A_n \vdash_\Sigma (\Diamond B_1 \wedge \cdots \wedge \Diamond B_m) \to \bot$$
by the deduction theorem
Proposition 3.36(4), and TAUT
$$A_1, \ldots, A_n \vdash_\Sigma \Diamond(B_1 \wedge \cdots \wedge B_m) \to \bot$$
since Σ is normal
$$A_1, \ldots, A_n \vdash_\Sigma \neg\Diamond(B_1 \wedge \cdots \wedge B_m)$$
by PL
$$A_1, \ldots, A_n \vdash_\Sigma \Box\neg(B_1 \wedge \cdots \wedge B_m)$$
$\Box\neg$ for $\neg\Diamond$
$$\Box A_1, \ldots, \Box A_n \vdash_\Sigma \Box\Box\neg(B_1 \wedge \cdots \wedge B_m)$$
by Lemma 4.6
$$\Box A_1, \ldots, \Box A_n \vdash_\Sigma \Box\neg(B_1 \wedge \cdots \wedge B_m)$$
by schema $\Box\Box A \to \Box A$
$$\Delta_1 \vdash_\Sigma \Box\neg(B_1 \wedge \cdots \wedge B_m)$$
by monotony, Proposition 3.36(1)
$$\Box\neg(B_1 \wedge \cdots \wedge B_m) \in \Delta_1$$
by deductive closure;
$$\neg(B_1 \wedge \cdots \wedge B_m) \in \Delta_2$$
since $R^\Sigma \Delta_1 \Delta_2$. □

On the strength of these examples, one might think that every system Σ of modal logic is *complete*, in the sense that it proves every formula which is valid in every frame in which every theorem of Σ is valid. Unfortunately, there are many systems that are not complete in this sense.

Problems

Problem 4.1. Complete the proof of Proposition 4.2.

Problem 4.2. Show that if Γ is complete Σ-consistent, then $\Diamond A \in \Gamma$ if and only if there is a complete Σ-consistent Δ such that $\Box^{-1}\Gamma \subseteq \Delta$ and $A \in \Delta$. *Do this without using Lemma 4.9.*

Problem 4.3. Complete the proof of Proposition 4.12.

CHAPTER 5

Filtrations and Decidability

5.1 Introduction

One important question about a logic is always whether it is decidable, i.e., if there is an effective procedure which will answer the question "is this formula valid." Propositional logic is decidable: we can effectively test if a formula is a tautology by constructing a truth table, and for a given formula, the truth table is finite. But we can't obviously test if a modal formula is true in all models, for there are infinitely many of them. We can list all the finite models relevant to a given formula, since only the assignment of subsets of worlds to propositional variables which actually occur in the formula are relevant. If the accessibility relation is fixed, the possible different assignments $V(p)$ are just all the subsets of W, and if $|W| = n$ there are 2^n of those. If our formula A contains m propositional variables there are then 2^{nm} different models with n worlds. For each one, we can test if A is true at all worlds, simply by computing the truth value of A in each. Of course, we also have to check all possible accessibility relations, but there are only finitely many relations on n worlds as well (specifically, the number of subsets of $W \times W$, i.e., 2^{n^2}.

If we are not interested in the logic **K**, but a logic defined by some class of models (e.g., the reflexive transitive models), we also have to be able to test if the accessibility relation is of the right kind. We can do that whenever the frames we are interested in are definable by modal formulas (e.g., by testing if T and 4 valid in the frame). So, the idea would be to run through all the finite frames, test each one if it is a frame in the class we're interested in, then list all the possible models on that frame and test if A is true in each. If not, stop: A is not valid in the class of models of interest.

There is a problem with this idea: we don't know when, if ever, we can stop looking. If the formula has a finite countermodel, our procedure will find it. But if it has no finite countermodel, we won't get an answer. The formula may be valid (no countermodels at all), or it have only an infinite countermodel, which we'll never look at. This problem can be overcome if we can show that every formula that has a countermodel has a finite countermodel. If this is the case we say the logic has the *finite model property*.

But how would we show that a logic has the finite model property? One way of doing this would be to find a way to turn an infinite (counter)model of A into a finite one. If that can be done, then whenever there is a model in which A is not true, then the resulting finite model also makes A not true. That finite model will show up on our list of all finite models, and we will eventually determine, for every formula that is not valid, that it isn't. Our procedure won't terminate if the formula is valid. If we can show in addition that there is some maximum size that the finite model our procedure provides can have, and that this maximum size depends only on the formula A, we will have a size up to which we have to test finite models in our search for countermodels. If we haven't found a countermodel by then, there are none. Then our procedure will, in fact, decide the question "is A valid?" for any formula A.

A strategy that often works for turning infinite structures into finite structures is that of "identifying" elements of the structure

which behave the same way in relevant respects. If there are infinitely many worlds in M that behave the same in relevant respects, then we might hope that there are only finitely many "classes" of such worlds. In other words, we partition the set of worlds in the right way. Each partition contains infinitely many worlds, but there are only finitely many partitions. Then we define a new model M^* where the worlds are the partitions. Finitely many partitions in the old model give us finitely many worlds in the new model, i.e., a finite model. Let's call the partition a world w is in $[w]$. We'll want it to be the case that $M, w \Vdash A$ iff $M^*, [w] \Vdash A$, since we want the new model to be a countermodel to A if the old one was. This requires that we define the partition, as well as the accessibility relation of M^* in the right way.

To see how this would go, first imagine we have no accessibility relation. $M, w \Vdash \Box B$ iff for some $v \in W$, $M, v \Vdash \Box B$, and the same for M^*, except with $[w]$ and $[v]$. As a first idea, let's say that two worlds u and v are equivalent (belong to the same partition) if they agree on all propositional variables in M, i.e., $M, u \Vdash p$ iff $M, v \Vdash p$. Let $V^*(p) = \{[w] : M, w \Vdash p\}$. Our aim is to show that $M, w \Vdash A$ iff $M^*, [w] \Vdash A$. Obviously, we'd prove this by induction: The base case would be $A \equiv p$. First suppose $M, w \Vdash p$. Then $[w] \in V^*$ by definition, so $M^*, [w] \Vdash p$. Now suppose that $M^*, [w] \Vdash p$. That means that $[w] \in V^*(p)$, i.e., for some v equivalent to w, $M, v \Vdash p$. But "w equivalent to v" means "w and v make all the same propositional variables true," so $M, w \Vdash p$. Now for the inductive step, e.g., $A \equiv \neg B$. Then $M, w \Vdash \neg B$ iff $M, w \nVdash B$ iff $M^*, [w] \nVdash B$ (by inductive hypothesis) iff $M^*, [w] \Vdash \neg B$. Similarly for the other non-modal operators. It also works for $\Box$: suppose $M^*, [w] \Vdash \Box B$. That means that for every $[u]$, $M^*, [u] \Vdash B$. By inductive hypothesis, for every u, $M, u \Vdash B$. Consequently, $M, w \Vdash \Box B$.

In the general case, where we have to also define the accessibility relation for M^*, things are more complicated. We'll call a model M^* a *filtration* if its accessibility relation R^* satisfies the conditions required to make the inductive proof above go through. Then any filtration M^* will make A true at $[w]$ iff M

makes A true at w. However, now we also have to show that there *are* filtrations, i.e., we can define R^* so that it satisfies the required conditions. In order for this to work, however, we have to require that worlds u, v count as equivalent not just when they agree on all propositional variables, but on all sub-formulas of A. Since A has only finitely many sub-formulas, this will still guarantee that the filtration is finite. There is not just one way to define a filtration, and in order to make sure that the accessibility relation of the filtration satisfies the required properties (e.g., reflexive, transitive, etc.) we have to be inventive with the definition of R^*.

5.2 Preliminaries

Filtrations allow us to establish the decidability of our systems of modal logic by showing that they have the *finite model property*, i.e., that any formula that is true (false) in a model is also true (false) in a *finite* model. Filtrations are defined relative to sets of formulas which are closed under subformulas.

Definition 5.1. A set Γ of formulas is *closed under subformulas* if it contains every subformula of a formula in Γ. Further, Γ is *modally closed* if it is closed under subformulas and moreover $A \in \Gamma$ implies $\Box A, \Diamond A \in \Gamma$.

For instance, given a formula A, the set of all its sub-formulas is closed under sub-formulas. When we're defining a filtration of a model through the set of sub-formulas of A, it will have the property we're after: it makes A true (false) iff the original model does.

The set of worlds of a filtration of M through Γ is defined as the set of all equivalence classes of the following equivalence relation.

Definition 5.2. Let $M = \langle W, R, V \rangle$ and suppose Γ is closed under sub-formulas. Define a relation $\equiv$ on W to hold of any two worlds that make the same formulas from Γ true, i.e.:

$$u \equiv v \quad \text{if and only if} \quad \forall A \in \Gamma : \boldsymbol{M}, u \Vdash A \Leftrightarrow \boldsymbol{N}, v \Vdash A.$$

The equivalence class $[w]_{\equiv}$ of a world w, or $[w]$ for short, is the set of all worlds $\equiv$-equivalent to w:

$$[w] = \{v : v \equiv w\}.$$

Proposition 5.3. *Given $\boldsymbol{M}$ and Γ, $\equiv$ as defined above is an equivalence relation, i.e., it is reflexive, symmetric, and transitive.*

Proof. The relation $\equiv$ is reflexive, since w makes exactly the same formulas from Γ true as itself. It is symmetric since if u makes the same formulas from Γ true as v, the same holds for v and u. It is also transitive, since if u makes the same formulas from Γ true as v, and v as w, then u makes the same formulas from Γ true as w. □

The relation $\equiv$, like any equivalence relation, divides W into *partitions*, i.e., subsets of W which are pairwise disjoint, and together cover all of W. Every $w \in W$ is an element of one of the partitions, namely of $[w]$, since $w \equiv w$. So the partitions $[w]$ cover all of W. They are pairwise disjoint, for if $u \in [w]$ and $u \in [v]$, then $u \equiv w$ and $u \equiv v$, and by symmetry and transitivity, $w \equiv v$, and so $[w] = [v]$.

5.3 Filtrations

Rather than define "the" filtration of $\boldsymbol{M}$ through Γ, we define when a model $\boldsymbol{M}^*$ counts as a filtration of $\boldsymbol{M}$. All filtrations have the same set of worlds W^* and the same valuation V^*. But different filtrations may have different accessibility relations R^*. To

count as a filtration, R^* has to satisfy a number of conditions, however. These conditions are exactly what we'll require to prove the main result, namely that $M, w \Vdash A$ iff $M^*, [w] \Vdash A$, provided $A \in \Gamma$.

Definition 5.4. Let Γ be closed under subformulas and $M = \langle W, R, V\rangle$. A *filtration of M through Γ* is any model $M^* = \langle W^*, R^*, V^*\rangle$, where:

1. $W^* = \{[w] : w \in W\}$;
2. For any $u, v \in W$:
   a) If Ruv then $R^*[u][v]$;
   b) If $R^*[u][v]$ then for any $\Box A \in \Gamma$, if $M, u \Vdash \Box A$ then $M, v \Vdash A$;
   c) If $R^*[u][v]$ then for any $\Diamond A \in \Gamma$, if $M, v \Vdash A$ then $M, u \Vdash \Diamond A$.
3. $V^*(p) = \{[u] : u \in V(p)\}$.

It's worthwhile thinking about what $V^*(p)$ is: the set consisting of the equivalence classes $[w]$ of all worlds w where p is true in M. On the one hand, if $w \in V(p)$, then $[w] \in V^*(p)$ by that definition. However, it is not necessarily the case that if $[w] \in V^*(p)$, then $w \in V(p)$. If $[w] \in V^*(p)$ we are only guaranteed that $[w] = [u]$ for *some* $u \in V(p)$. Of course, $[w] = [u]$ means that $w \equiv u$. So, when $[w] \in V^*(p)$ we can (only) conclude that $w \equiv u$ for some $u \in V(p)$.

Theorem 5.5. *If M^* is a filtration of M through Γ, then for every $A \in \Gamma$ and $w \in W$, we have $M, w \Vdash A$ if and only if $M^*, [w] \Vdash A$.*

Proof. By induction on A, using the fact that Γ is closed under subformulas. Since $A \in \Gamma$ and Γ is closed under sub-formulas, all sub-formulas of A are also $\in \Gamma$. Hence in each inductive step, the induction hypothesis applies to the sub-formulas of A.

1. $A \equiv \bot$: Neither $M, w \Vdash A$ nor $M^*, w \Vdash A$.

2. $A \equiv p$: The left-to-right direction is immediate, as $M, w \Vdash A$ only if $w \in V(p)$, which implies $[w] \in V^*(p)$, i.e., $M^*, [w] \Vdash A$. Conversely, suppose $M^*, [w] \Vdash A$, i.e., $[w] \in V^*(p)$. Then for some $v \in V(p)$, $w \equiv v$. Of course then also $M, v \Vdash p$. Since $w \equiv v$, w and v make the same formulas from Γ true. Since by assumption $p \in \Gamma$ and $M, v \Vdash p$, $M, w \Vdash A$.

3. $A \equiv \neg B$: $M, w \Vdash A$ iff $M, w \nVdash B$. By induction hypothesis, $M, w \nVdash B$ iff $M^*, [w] \nVdash B$. Finally, $M^*, [w] \nVdash B$ iff $M^*, [w] \Vdash A$.

4. Exercise.

5. $A \equiv (B \vee C)$: $M, w \Vdash A$ iff $M, w \Vdash B$ or $M, w \Vdash C$. By induction hypothesis, $M, w \Vdash B$ iff $M^*, [w] \Vdash B$, and $M, w \Vdash C$ iff $M^*, [w] \Vdash C$. And $M^*, [w] \Vdash A$ iff $M^*, [w] \Vdash B$ or $M^*, [w] \Vdash C$.

6. Exercise.

7. $A \equiv \Box B$: Suppose $M, w \Vdash A$; to show that $M^*, [w] \Vdash A$, let v be such that $R^*[w][v]$. From Definition 5.4(2b), we have that $M, v \Vdash B$, and by inductive hypothesis $M^*, [v] \Vdash B$. Since v was arbitrary, $M^*, [w] \Vdash A$ follows.

   Conversely, suppose $M^*, [w] \Vdash A$ and let v be arbitrary such that Rwv. From Definition 5.4(2a), we have $R^*[w][v]$, so that $M^*, [v] \Vdash B$; by inductive hypothesis $M, v \Vdash B$, and since v was arbitrary, $M, u \Vdash A$.

8. Exercise. □

What holds for truth at worlds in a model also holds for truth in a model and validity in a class of models.

Corollary 5.6. *Let Γ be closed under subformulas. Then:*

1. *If M^* is a filtration of M through Γ then for any $A \in \Gamma$: $M \Vdash A$ if and only if $M^* \Vdash A$.*
2. *If $\mathscr{C}$ is a class of models and $\Gamma(\mathscr{C})$ is the class of Γ-filtrations of models in $\mathscr{C}$, then any formula $A \in \Gamma$ is valid in $\mathscr{C}$ if and only if it is valid in $\Gamma(\mathscr{C})$.*

5.4 Examples of Filtrations

We have not yet shown that there are any filtrations. But indeed, for any model M, there are many filtrations of M through Γ. We identify two, in particular: the finest and coarsest filtrations. Filtrations of the same models will differ in their accessibility relation (as Definition 5.4 stipulates directly what W^* and V^* should be). The finest filtration will have as few related worlds as possible, whereas the coarsest will have as many as possible.

Definition 5.7. Where Γ is closed under subformulas, the *finest* filtration M^* of a model M is defined by putting:

$$R^*[u][v] \quad \text{if and only if} \quad \exists u' \in [u]\ \exists v' \in [v] : Ru'v'.$$

Proposition 5.8. *The finest filtration M^* is indeed a filtration.*

Proof. We need to check that R^*, so defined, satisfies Definition 5.4(2). We check the three conditions in turn.

If Ruv then since $u \in [u]$ and $v \in [v]$, also $R^*[u][v]$, so (2a) is satisfied.

For (2b), suppose $\Box A \in \Gamma$, $R^*[u][v]$, and $M, u \Vdash \Box A$. By definition of R^*, there are $u' \equiv u$ and $v' \equiv v$ such that $Ru'v'$. Since u and u' agree on Γ, also $M, u' \Vdash \Box A$, so that $M, v' \Vdash A$. By closure of Γ under sub-formulas, v and v' agree on A, so $M, v \Vdash A$, as desired.

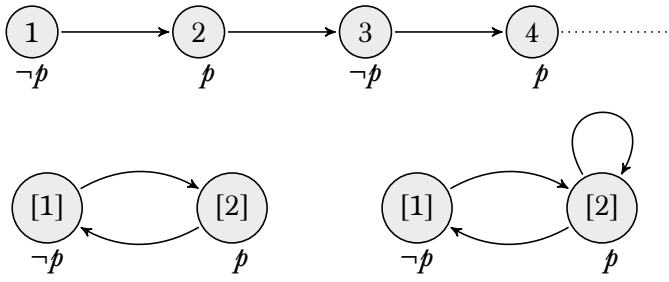


Figure 5.1: An infinite model and its filtrations.

We leave the verification of (2c) as an exercise. □

Definition 5.9. Where Γ is closed under subformulas, the *coarsest* filtration M^* of a model M is defined by putting $R^*[u][v]$ if and only if *both* of the following conditions are met:

1. If $\Box A \in \Gamma$ and $M, u \Vdash \Box A$ then $M, v \Vdash A$;
2. If $\Diamond A \in \Gamma$ and $M, v \Vdash A$ then $M, u \Vdash \Diamond A$.

Proposition 5.10. *The coarsest filtration M^* is indeed a filtration.*

Proof. Given the definition of R^*, the only condition that is left to verify is the implication from Ruv to $R^*[u][v]$. So assume Ruv. Suppose $\Box A \in \Gamma$ and $M, u \Vdash \Box A$; then obviously $M, v \Vdash A$, and (1) is satisfied. Suppose $\Diamond A \in \Gamma$ and $M, v \Vdash A$. Then $M, u \Vdash \Diamond A$ since Ruv, and (2) is satisfied. □

Example 5.11. Let $W = \mathbb{Z}^+$, Rnm iff $m = n + 1$, and $V(p) = \{2n : n \in \mathbb{N}\}$. The model $M = \langle W, R, V\rangle$ is depicted in Figure 5.1. The worlds are 1, 2, etc.; each world can access exactly one other world—its successor, and p is true at all and only the even numbers.

Now let Γ be the set of sub-formulas of $\Box p \to p$, i.e., $\{p, \Box p, \Box p \to p\}$. p is true at all and only the even numbers,

$\Box p$ is true at all and only the odd numbers, so $\Box p \to p$ is true at all and only the even numbers. In other words, every odd number makes $\Box p$ true and p and $\Box p \to p$ false; every even number makes p and $\Box p \to p$ true, but $\Box p$ false. So $W^* = \{[1], [2]\}$, where $[1] = \{1, 3, 5, \dots\}$ and $[2] = \{2, 4, 6, \dots\}$. Since $2 \in V(p)$, $[2] \in V^*(p)$; since $1 \notin V(p)$, $[1] \notin V^*(p)$. So $V^*(p) = \{[2]\}$.

Any filtration based on W^* must have an accessibility relation that includes $\langle [1], [2] \rangle, \langle [2], [1] \rangle$: since $R12$, we must have $R^*[1][2]$ by Definition 5.4(2a), and since $R23$ we must have $R^*[2][3]$, and $[3] = [1]$. It cannot include $\langle [1], [1] \rangle$: if it did, we'd have $R^*[1][1]$, $M, 1 \Vdash \Box p$ but $M, 1 \nVdash p$, contradicting (2a). Nothing requires or rules out that $R^*[2][2]$. So, there are two possible filtrations of M, corresponding to the two accessibility relations

$$\{\langle [1], [2] \rangle, \langle [2], [1] \rangle\} \text{ and } \{\langle [1], [2] \rangle, \langle [2], [1] \rangle, \langle [2], [2] \rangle\}.$$

In either case, p and $\Box p \to p$ are false and $\Box p$ is true at $[1]$; p and $\Box p \to p$ are true and $\Box p$ is false at $[2]$.

5.5 Filtrations are Finite

We've defined filtrations for any set Γ that is closed under sub-formulas. Nothing in the definition itself guarantees that filtrations are finite. In fact, when Γ is infinite (e.g., is the set of all formulas), it may well be infinite. However, if Γ is finite (e.g., when it is the set of sub-formulas of a given formula A), so is any filtration through Γ.

Proposition 5.12. *If Γ is finite then any filtration M^* of a model M through Γ is also finite.*

Proof. The size of W^* is the number of different classes $[w]$ under the equivalence relation $\equiv$. Any two worlds u, v in such class—that is, any u and v such that $u \equiv v$—agree on all formulas A in Γ, $A \in \Gamma$ either A is true at both u and v, or at neither. So each class $[w]$ corresponds to subset of Γ, namely the set of all

$A \in \Gamma$ such that A is true at the worlds in $[w]$. No two different classes $[u]$ and $[v]$ correspond to the same subset of Γ. For if the set of formulas true at u and that of formulas true at v are the same, then u and v agree on all formulas in Γ, i.e., $u \equiv v$. But then $[u] = [v]$. So, there is an injective function from W^* to $\wp(\Gamma)$, and hence $|W^*| \leq |\wp(\Gamma)|$. Hence if Γ contains n sentences, the cardinality of W^* is no greater than 2^n. □

5.6 K and S5 have the Finite Model Property

Definition 5.13. A system Σ of modal logic is said to have the *finite model property* if whenever a formula A is true at a world in a model of Σ then A is true at a world in a *finite* model of Σ.

Proposition 5.14. **K** *has the finite model property.*

Proof. **K** is the set of valid formulas, i.e., any model is a model of **K**. By Theorem 5.5, if $M, w \Vdash A$, then $M^*, w \Vdash A$ for any filtration of M through the set Γ of sub-formulas of A. Any formula only has finitely many sub-formulas, so Γ is finite. By Proposition 5.12, $|W^*| \leq 2^n$, where n is the number of formulas in Γ. And since **K** imposes no restriction on models, M^* is a **K**-model. □

To show that a logic **L** has the finite model property via filtrations it is essential that the filtration of an **L**-model is itself a **L**-model. Often this requires a fair bit of work, and not any filtration yields a **L**-model. However, for universal models, this still holds.

Proposition 5.15. *Let $\mathcal{U}$ be the class of universal models (see Proposition 2.14) and $\mathcal{U}_{\text{Fin}}$ the class of all finite universal models. Then any formula A is valid in $\mathcal{U}$ if and only if it is valid in $\mathcal{U}_{\text{Fin}}$.*

Proof. Finite universal models are universal models, so the left-to-right direction is trivial. For the right-to left direction, suppose that A is false at some world w in a universal model M. Let Γ contain A as well as all of its subformulas; clearly Γ is finite. Take a filtration M^* of M; then M^* is finite by Proposition 5.12, and by Theorem 5.5, A is false at $[w]$ in M^*. It remains to observe that M^* is also universal: given u and v, by hypothesis Ruv and by Definition 5.4(2), also $R^*[u][v]$. □

Corollary 5.16. **S5** *has the finite model property.*

Proof. By Proposition 2.14, if A is true at a world in some reflexive and euclidean model then it is true at a world in a universal model. By Proposition 5.15, it is true at a world in a finite universal model (namely the filtration of the model through the set of sub-formulas of A). Every universal model is also reflexive and euclidean; so A is true at a world in a finite reflexive euclidean model. □

5.7 S5 is Decidable

The finite model property gives us an easy way to show that systems of modal logic given by schemas are *decidable* (i.e., that there is a computable procedure to determine whether a formulas is derivable in the system or not).

Theorem 5.17. **S5** *is decidable.*

Proof. Let A be given, and suppose the propositional variables occurring in A are among $p_1, \ldots, p_k$. Since for each n there are only finitely many models with n worlds assigning a value to $p_1, \ldots, p_k$, we can enumerate, *in parallel*, all the theorems of **S5** by generating proofs in some systematic way; and all the models containing $1, 2, \ldots$ worlds and checking whether A fails at a world in some such model. Eventually one of the two parallel processes will give an answer, as by Theorem 4.17 and Corollary 5.16, either A is derivable or it fails in a finite universal model. □

The above proof works for **S5** because filtrations of universal models are automatically universal. The same holds for reflexivity and seriality, but more work is needed for other properties.

5.8 Filtrations and Properties of Accessibility

As noted, filtrations of universal, serial, and reflexive models are always also universal, serial, or reflexive. But not every filtration of a symmetric or transitive model is symmetric or transitive, respectively. In some cases, however, it is possible to define filtrations so that this does hold. In order to do so, we proceed as in the definition of the coarsest filtration, but add additional conditions to the definition of R^*. Let Γ be closed under sub-formulas. Consider the relations $C_i(u, v)$ in table 5.1 between worlds u, v in a model $M = \langle W, R, V \rangle$. We can define $R^*[u][v]$ on the basis of combinations of these conditions. For instance, if we stipulate that $R^*[u][v]$ iff the condition $C_1(u, v)$ holds, we get exactly the coarsest filtration. If we stipulate $R^*[u][v]$ iff both $C_1(u, v)$ and $C_2(u, v)$ hold, we get a different filtration. It is "finer" than the coarsest since fewer pairs of worlds satisfy $C_1(u, v)$ and $C_2(u, v)$ than $C_1(u, v)$ alone.

$C_1(u,v)$:	if $\Box A \in \Gamma$ and $\boldsymbol{M}, u \Vdash \Box A$ then $\boldsymbol{M}, v \Vdash A$; and if $\Diamond A \in \Gamma$ and $\boldsymbol{M}, v \Vdash A$ then $\boldsymbol{M}, u \Vdash \Diamond A$;
$C_2(u,v)$:	if $\Box A \in \Gamma$ and $\boldsymbol{M}, v \Vdash \Box A$ then $\boldsymbol{M}, u \Vdash A$; and if $\Diamond A \in \Gamma$ and $\boldsymbol{M}, u \Vdash A$ then $\boldsymbol{M}, v \Vdash \Diamond A$;
$C_3(u,v)$:	if $\Box A \in \Gamma$ and $\boldsymbol{M}, u \Vdash \Box A$ then $\boldsymbol{M}, v \Vdash \Box A$; and if $\Diamond A \in \Gamma$ and $\boldsymbol{M}, v \Vdash \Diamond A$ then $\boldsymbol{M}, u \Vdash \Diamond A$;
$C_4(u,v)$:	if $\Box A \in \Gamma$ and $\boldsymbol{M}, v \Vdash \Box A$ then $\boldsymbol{M}, u \Vdash \Box A$; and if $\Diamond A \in \Gamma$ and $\boldsymbol{M}, u \Vdash \Diamond A$ then $\boldsymbol{M}, v \Vdash \Diamond A$;

Table 5.1: Conditions on possible worlds for defining filtrations.

Theorem 5.18. *Let $\boldsymbol{M} = \langle W, R, P\rangle$ be a model, Γ closed under subformulas. Let W^* and V^* be defined as in Definition 5.4. Then:*

1. *Suppose $R^*[u][v]$ if and only if $C_1(u,v) \wedge C_2(u,v)$. Then R^* is symmetric, and $\boldsymbol{M}^* = \langle W^*, R^*, V^*\rangle$ is a filtration if $\boldsymbol{M}$ is symmetric.*

2. *Suppose $R^*[u][v]$ if and only if $C_1(u,v) \wedge C_3(u,v)$. Then R^* is transitive, and $\boldsymbol{M}^* = \langle W^*, R^*, V^*\rangle$ is a filtration if $\boldsymbol{M}$ is transitive.*

3. *Suppose $R^*[u][v]$ if and only if $C_1(u,v) \wedge C_2(u,v) \wedge C_3(u,v) \wedge C_4(u,v)$. Then R^* is symmetric and transitive, and $\boldsymbol{M}^* = \langle W^*, R^*, V^*\rangle$ is a filtration if $\boldsymbol{M}$ is symmetric and transitive.*

4. *Suppose R^* is defined as $R^*[u][v]$ if and only if $C_1(u,v) \wedge C_3(u,v) \wedge C_4(u,v)$. Then R^* is transitive and euclidean, and $\boldsymbol{M}^* = \langle W^*, R^*, V^*\rangle$ is a filtration if $\boldsymbol{M}$ is transitive and euclidean.*

Proof. 1. It's immediate that R^* is symmetric, since $C_1(u,v) \Leftrightarrow C_2(v,u)$ and $C_2(u,v) \Leftrightarrow C_1(v,u)$. So it's left to show that if $\boldsymbol{M}$ is symmetric then $\boldsymbol{M}^*$ is a filtration through Γ. Condition $C_1(u,v)$ guarantees that (2b) and (2c) of Definition 5.4 are satisfied. So we just have to verify Definition 5.4(2a), i.e., that Ruv implies $R^*[u][v]$.

So suppose Ruv. To show $R^*[u][v]$ we need to establish that $C_1(u,v)$ and $C_2(u,v)$. For C_1: if $\Box A \in \Gamma$ and $M, u \Vdash \Box A$ then also $M, v \Vdash A$ (since Ruv). Similarly, if $\Diamond A \in \Gamma$ and $M, v \Vdash A$ then $M, u \Vdash \Diamond A$ since Ruv. For C_2: if $\Box A \in \Gamma$ and $M, v \Vdash \Box A$ then Ruv implies Rvu by symmetry, so that $M, u \Vdash A$. Similarly, if $\Diamond A \in \Gamma$ and $M, u \Vdash A$ then $M, v \Vdash \Diamond A$ (since Rvu by symmetry).

2. Exercise.

3. Exercise.

4. Exercise. □

5.9 Filtrations of Euclidean Models

The approach of section 5.8 does not work in the case of models that are euclidean or serial and euclidean. Consider the model at the top of Figure 5.2, which is both euclidean and serial. Let $\Gamma = \{p, \Box p\}$. When taking a filtration through Γ, then $[w_1] = [w_3]$ since w_1 and w_3 are the only worlds that agree on Γ. Any filtration will also have the arrow inherited from M, as depicted in Figure 5.3. That model isn't euclidean. Moreover, we cannot add arrows to that model in order to make it euclidean. We would have to add double arrows between $[w_2]$ and $[w_4]$, and then also between w_2 and w_5. But $\Box p$ is supposed to be true at w_2, while p is false at w_5.

In particular, to obtain a euclidean flitration it is not enough to consider filtrations through arbitrary Γ's closed under subformulas. Instead we need to consider sets Γ that are *modally closed* (see Definition 5.1). Such sets of sentences are infinite, and therefore do not immediately yield a finite model property or the decidability of the corresponding system.

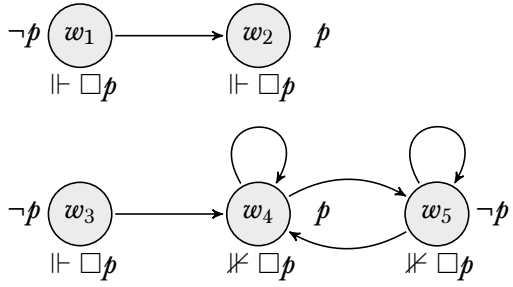


Figure 5.2: A serial and euclidean model.

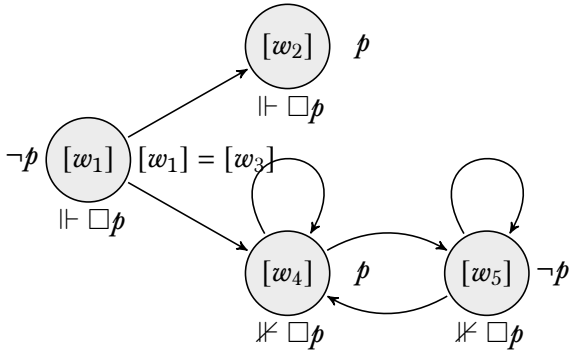


Figure 5.3: The filtration of the model in Figure 5.2.

Theorem 5.19. *Let Γ be modally closed, $\mathbf{M} = \langle W, R, V\rangle$, and $\mathbf{M}^* = \langle W^*, R^*, V^*\rangle$ be a coarsest filtration of $\mathbf{M}$.*

1. *If $\mathbf{M}$ is symmetric, so is $\mathbf{M}^*$.*
2. *If $\mathbf{M}$ is transitive, so is $\mathbf{M}^*$.*
3. *If $\mathbf{M}$ is euclidean, so is $\mathbf{M}^*$.*

Proof. 1. If $\mathbf{M}^*$ is a coarsest filtration, then by definition $R^*[u][v]$ holds if and only if $C_1(u, v)$. For transitivity, suppose $C_1(u, v)$ and $C_1(v, w)$; we have to show $C_1(u, w)$. Suppose $\mathbf{M}, u \Vdash \Box A$; then $\mathbf{M}, u \Vdash \Box\Box A$ since 4 is valid in

all transitive models; since $\Box\Box A \in \Gamma$ by closure, also by $C_1(u,v)$, $M, v \Vdash \Box A$ and by $C_1(v,w)$, also $M, w \Vdash A$. Suppose $M, w \Vdash A$; then $M, v \Vdash \Diamond A$ by $C_1(v,w)$, since $\Diamond A \in \Gamma$ by modal closure. By $C_1(u,v)$, we get $M, u \Vdash \Diamond\Diamond A$ since $\Diamond\Diamond A \in \Gamma$ by modal closure. Since $4_\Diamond$ is valid in all transitive models, $M, u \Vdash \Diamond A$.

2. Exercise. Use the fact that both 5 and $5_\Diamond$ are valid in all euclidean models.

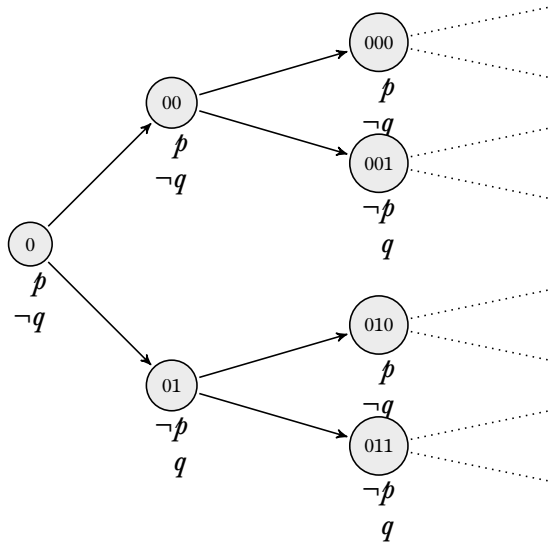
3. Exercise. Use the fact that B and $B_\Diamond$ are valid in all symmetric models. □

Problems

Problem 5.1. Complete the proof of Theorem 5.5

Problem 5.2. Complete the proof of Proposition 5.8.

Problem 5.3. Consider the following model $M = \langle W, R, V\rangle$ where $W = \{0\sigma : \sigma \in \mathbb{B}^*\}$, the set of sequences of 0s and 1s starting with 0, with $R\sigma\sigma'$ iff $\sigma' = \sigma 0$ or $\sigma' = \sigma 1$, and $V(p) = \{\sigma 0 : \sigma \in \mathbb{B}^*\}$ and $V(q) = \{\sigma 1 : \sigma \in \mathbb{B}^* \setminus \{1\}\}$. Here's a picture:



We have $M, w \nVdash \Box(p \vee q) \to (\Box p \vee \Box q)$ for every w.

Let Γ be the set of sub-formulas of $\Box(p \vee q) \to (\Box p \vee \Box q)$. What are W^* and V^*? What is the accessibility relation of the finest filtration of M? Of the coarsest?

Problem 5.4. Show that any filtration of a serial or reflexive model is also serial or reflexive (respectively).

Problem 5.5. Find a non-symmetric (non-transitive, non-euclidean) filtration of a symmetric (transitive, euclidean) model.

Problem 5.6. Show that any filtration of a serial or reflexive model is also serial or reflexive (respectively).

Problem 5.7. Find a non-symmetric (non-transitive, non-euclidean) filtration of a symmetric (transitive, euclidean) model.

Problem 5.8. Complete the proof of Theorem 5.18.

Problem 5.9. Complete the proof of Theorem 5.19.

CHAPTER 6

Modal Tableaux

6.1 Introduction

Tableaux are certain (downward-branching) trees of signed formulas, i.e., pairs consisting of a truth value sign ($\mathbb{T}$ or $\mathbb{F}$) and a sentence

$$\mathbb{T}\,A \text{ or } \mathbb{F}\,A.$$

A tableau begins with a number of *assumptions*. Each further signed formula is generated by applying one of the inference rules. Some inference rules add one or more signed formulas to a tip of the tree; others add two new tips, resulting in two branches. Rules result in signed formulas where the formula is less complex than that of the signed formula to which it was applied. When a branch contains both $\mathbb{T}\,A$ and $\mathbb{F}\,A$, we say the branch is *closed*. If every branch in a tableau is closed, the entire tableau is closed. A closed tableau consititues a derivation that shows that the set of signed formulas which were used to begin the tableau are unsatisfiable. This can be used to define a $\vdash$ relation: $\Gamma \vdash A$ iff there is some finite set $\Gamma_0 = \{B_1, \ldots, B_n\} \subseteq \Gamma$ such that there is a closed tableau for the assumptions

$$\{\mathbb{F}\,A, \mathbb{T}\,B_1, \ldots, \mathbb{T}\,B_n\}.$$

For modal logics, we have to both extend the notion of signed formula and add rules that cover $\Box$ and $\Diamond$ In addition to a sign($\mathbb{T}$ or $\mathbb{F}$), formulas in modal tableaux also have *prefixes* σ. The prefixes are non-empty sequences of positive integers, i.e., $\sigma \in (\mathbb{Z}^+)^* \setminus \{\Lambda\}$. When we write such prefixes without the surrounding $\langle\ \rangle$, and separate the individual elements by .'s instead of ,'s. If σ is a prefix, then $\sigma.n$ is $\sigma \frown \langle n \rangle$; e.g., if $\sigma = 1.2.1$, then $\sigma.3$ is 1.2.1.3. So for instance,

$$1.2\,\mathbb{T}\,\Box A \to A$$

is a *prefixed signed formula* (or just a *prefixed formula* for short).

Intuitively, the prefix names a world in a model that might satisfy the formulas on a branch of a tableau, and if σ names some world, then $\sigma.n$ names a world accessible from (the world named by) σ.

6.2 Rules for K

The rules for the regular propositional connectives are the same as for regular propositional signed tableaux, just with prefixes added. In each case, the rule applied to a signed formula $\sigma\, S\, A$ produces new formulas that are also prefixed by σ. This should be intuitively clear: e.g., if $A \wedge B$ is true at (a world named by) σ, then A and B are true at σ (and not at any other world). We collect the propositional rules in table 6.1.

The closure condition is the same as for ordinary tableaux, although we require that not just the formulas but also the prefixes must match. So a branch is closed if it contains both

$$\sigma\,\mathbb{T}\,A \quad \text{and} \quad \sigma\,\mathbb{F}\,A$$

for some prefix σ and formula A.

The rules for setting up assumptions is also as for ordinary tableaux, except that for asusmptions we always use the prefix 1. (It does not matter which prefix we use, as long as it's the same

$\dfrac{\sigma\,\mathbb{T}\,\neg A}{\sigma\,\mathbb{F}\,A}\ \neg\mathbb{T}$	$\dfrac{\sigma\,\mathbb{F}\,\neg A}{\sigma\,\mathbb{T}\,A}\ \neg\mathbb{F}$
$\dfrac{\sigma\,\mathbb{T}\,A \wedge B}{\begin{array}{l}\sigma\,\mathbb{T}\,A\\ \sigma\,\mathbb{T}\,B\end{array}}\ \wedge\mathbb{T}$	$\dfrac{\sigma\,\mathbb{F}\,A \wedge B}{\sigma\,\mathbb{F}\,A \quad \mid \quad \sigma\,\mathbb{F}\,B}\ \wedge\mathbb{F}$
$\dfrac{\sigma\,\mathbb{T}\,A \vee B}{\sigma\,\mathbb{T}\,A \quad \mid \quad \sigma\,\mathbb{T}\,B}\ \vee\mathbb{T}$	$\dfrac{\sigma\,\mathbb{F}\,A \vee B}{\begin{array}{l}\sigma\,\mathbb{F}\,A\\ \sigma\,\mathbb{F}\,B\end{array}}\ \vee\mathbb{F}$
$\dfrac{\sigma\,\mathbb{T}\,A \to B}{\sigma\,\mathbb{F}\,A \quad \mid \quad \sigma\,\mathbb{T}\,B}\ \to\mathbb{T}$	$\dfrac{\sigma\,\mathbb{F}\,A \to B}{\begin{array}{l}\sigma\,\mathbb{T}\,A\\ \sigma\,\mathbb{F}\,B\end{array}}\ \to\mathbb{F}$

Table 6.1: Prefixed tableau rules for the propositional connectives

for all assumptions.) So, e.g., we say that

$$B_1, \ldots, B_n \vdash A$$

iff there is a closed tableau for the assumptions

$$1\,\mathbb{T}\,B_1, \ldots, 1\,\mathbb{T}\,B_n, 1\,\mathbb{F}\,A.$$

For the modal operators $\Box$ and $\Diamond$, the prefix of the conclusion of the rule applied to a formula with prefix σ is $\sigma.n$. However, which n is allowed depends on whether the sign is $\mathbb{T}$ or $\mathbb{F}$.

The $\mathbb{T}\Box$ rule extends a branch containing $\sigma\,\mathbb{T}\,\Box A$ by $\sigma.n\,\mathbb{T}\,A$. Similarly, the $\mathbb{F}\Diamond$ rule extends a branch containing $\sigma\,\mathbb{F}\,\Diamond A$ by $\sigma.n\,\mathbb{F}\,A$. They can only be applied for a prefix $\sigma.n$ which *already* occurs on the branch in which it is applied. Let's call such a prefix "used" (on the branch).

The $\mathbb{F}\Box$ rule extends a branch containing $\sigma\,\mathbb{F}\,\Box A$ by $\sigma.n\,\mathbb{F}\,A$. Similarly, the $\mathbb{T}\Diamond$ rule extends a branch containing $\sigma\,\mathbb{T}\,\Diamond A$ by

$\dfrac{\sigma\,\mathbb{T}\,\Box A}{\sigma.n\,\mathbb{T}\,A}\,\Box\mathbb{T}$ $\sigma.n$ is used	$\dfrac{\sigma\,\mathbb{F}\,\Box A}{\sigma.n\,\mathbb{F}\,A}\,\Box\mathbb{F}$ $\sigma.n$ is new
$\dfrac{\sigma\,\mathbb{T}\,\Diamond A}{\sigma.n\,\mathbb{T}\,A}\,\Diamond\mathbb{T}$ $\sigma.n$ is new	$\dfrac{\sigma\,\mathbb{F}\,\Diamond A}{\sigma.n\,\mathbb{F}\,A}\,\Diamond\mathbb{F}$ $\sigma.n$ is used

Table 6.2: The modal rules for K.

$\sigma.n\,\mathbb{T}\,A$. These rules, however, can only be applied for a prefix $\sigma.n$ which *does not* already occur on the branch in which it is applied. We call such prefixes "new" (to the branch).

The rules are given in table 6.2.

The requirements that the restriction that the prefix for $\Box\mathbb{T}$ must be used is necessary as otherwise we would count the following as a closed tableau:

1.	$1\,\mathbb{T}\ \ \Box A$	Assumption
2.	$1\,\mathbb{F}\ \ \Diamond A$	Assumption
3.	$1.1\,\mathbb{T}\ \ A$	$\Box\mathbb{T}\,1$
4.	$1.1\,\mathbb{F}\ \ A$	$\Diamond\mathbb{F}\,2$
	$\otimes$	

But $\Box A \not\models \Diamond A$, so our proof system would be unsound. Likewise, $\Diamond A \not\models \Box A$, but without the restriction that the prefix for $\Box\mathbb{F}$ must be new, this would be a closed tableau:

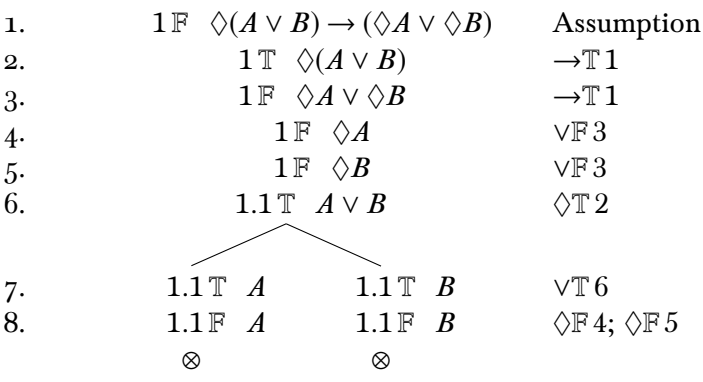
1.	$1\,\mathbb{T}\ \ \Diamond A$	Assumption
2.	$1\,\mathbb{F}\ \ \Box A$	Assumption
3.	$1.1\,\mathbb{T}\ \ A$	$\Diamond\mathbb{T}\,1$
4.	$1.1\,\mathbb{F}\ \ A$	$\Box\mathbb{F}\,2$
	$\otimes$	

6.3 Tableaux for K

Example 6.1. We give a closed tableau that shows $\vdash (\Box A \wedge \Box B) \rightarrow \Box(A \wedge B)$.

1.	$1\,\mathbb{F}\ \ (\Box A \wedge \Box B) \rightarrow \Box(A \wedge B)$		Assumption
2.	$1\,\mathbb{T}\ \ \Box A \wedge \Box B$		$\rightarrow\mathbb{T}\,1$
3.	$1\,\mathbb{F}\ \ \Box(A \wedge B)$		$\rightarrow\mathbb{T}\,1$
4.	$1\,\mathbb{T}\ \ \Box A$		$\wedge\mathbb{T}\,2$
5.	$1\,\mathbb{T}\ \ \Box B$		$\wedge\mathbb{T}\,2$
6.	$1.1\,\mathbb{F}\ \ A \wedge B$		$\Box\mathbb{F}\,3$
7.	$1.1\,\mathbb{F}\ \ A$	$1.1\,\mathbb{F}\ \ B$	$\wedge\mathbb{F}\,6$
8.	$1.1\,\mathbb{T}\ \ A$	$1.1\,\mathbb{T}\ \ B$	$\Box\mathbb{T}\,4$; $\Box\mathbb{T}\,5$
	$\otimes$	$\otimes$	

Example 6.2. We give a closed tableau that shows $\vdash \Diamond(A \vee B) \rightarrow (\Diamond A \vee \Diamond B)$:



6.4 Soundness for K

In order to show that prefixed tableaux are sound, we have to show that if

$$1\,\mathbb{T}\,B_1, \ldots, 1\,\mathbb{T}\,B_n, 1\,\mathbb{F}\,A$$

has a closed tableau then $B_1, \ldots, B_n \vDash A$. It is easier to prove the contrapositive: if for some M and world w, $M, w \Vdash B_i$ for all $i = 1, \ldots, n$ but $M, w \nVdash A$, then no tableau can close. Such a countermodel shows that the initial assumptions of the tableau are satisfiable. The strategy of the proof is to show that whenever all the prefixed formulas on a tableau branch are satisfiable, any application of a rule results in at least one extended branch that is also satisfiable. Since closed branches are unsatisfiable, any tableau for a satisfiable set of prefixed formulas must have at least one open branch.

In order to apply this strategy in the modal case, we have to extend our definition of "satisfiable" to modal modals and prefixes. With that in hand, however, the proof is straightforward.

Definition 6.3. Let P be some set of prefixes, i.e., $P \subseteq (\mathbb{Z}^+)^* \setminus \{\Lambda\}$ and let $\boldsymbol{M}$ be a model. A function $f : P \to W$ is an *interpretation of* P in $\boldsymbol{M}$ if, whenever σ and $\sigma.n$ are both in P, then $Rf(\sigma)f(\sigma.n)$.

Relative to an interpretation of prefixes P we can define:

1. $\boldsymbol{M}$ satisfies $\sigma\,\mathbb{T}\,A$ iff $\boldsymbol{M}, f(\sigma) \Vdash A$.
2. $\boldsymbol{M}$ satisfies $\sigma\,\mathbb{F}\,A$ iff $\boldsymbol{M}, f(\sigma) \nVdash A$.

Definition 6.4. Let Γ be a set of prefixed formulas, and let $P(\Gamma)$ be the set of prefixes that occur in it. If f is an interpretation of $P(\Gamma)$ in $\boldsymbol{M}$, we say that $\boldsymbol{M}$ satisfies Γ with respect to f, $\boldsymbol{M}, f \Vdash \Gamma$, if $\boldsymbol{M}$ satisfies every prefixed formula in Γ with respect to f. Γ is *satisfiable* iff there is a model $\boldsymbol{M}$ and interpretation f of $P(\Gamma)$ such that $\boldsymbol{M}, f \Vdash \Gamma$.

Proposition 6.5. *If Γ contains both $\sigma\,\mathbb{T}\,A$ and $\sigma\,\mathbb{F}\,A$, for some formula A and prefix σ, then Γ is unsatisfiable.*

Proof. There cannot be a model $\boldsymbol{M}$ and interpretation f of $P(\Gamma)$ such that both $\boldsymbol{M}, f(\sigma) \Vdash A$ and $\boldsymbol{M}, f(\sigma) \nVdash A$. □

Theorem 6.6 (Soundness). *If Γ has a closed tableau, Γ is unsatisfiable.*

Proof. We call a branch of a tableau satisfiable iff the set of signed formulas on it is satisfiable, and let's call a tableau satisfiable if it contains at least one satisfiable branch.

We show the following: Extending a satisfiable tableau by one of the rules of inference always results in a satisfiable tableau. This will prove the theorem: any closed tableau results by applying rules of inference to the tableau consisting only of assumptions from Γ. So if Γ were satisfiable, any tableau for it would

be satisfiable. A closed tableau, however, is clearly not satisfiable, since all its branches are closed and closed branches are unsatisfiable.

Suppose we have a satisfiable tableau, i.e., a tableau with at least one satisfiable branch. Applying a rule of inference either adds signed formulas to a branch, or splits a branch in two. If the tableau has a satisfiable branch which is not extended by the rule application in question, it remains a satisfiable branch in the extended tableau, so the extended tableau is satisfiable. So we only have to consider the case where a rule is applied to a satisfiable branch.

Let Γ be the set of signed formulas on that branch, and let $\sigma\, S\, A \in \Gamma$ be the signed formula to which the rule is applied. If the rule does not result in a split branch, we have to show that the extended branch, i.e., Γ together with the conclusions of the rule, is still satisfiable. If the rule results in split branch, we have to show that at least one of the two resulting branches is satisfiable. First, we consider the possible inferences with only one premise.

1. The branch is expanded by applying $\neg\mathbb{T}$ to $\sigma\, \mathbb{T}\, \neg B \in \Gamma$. Then the extended branch contains the signed formulas $\Gamma \cup \{\sigma\, \mathbb{F}\, B\}$. Suppose $\boldsymbol{M}, f \Vdash \Gamma$. In particular, $\boldsymbol{M}, f(\sigma) \Vdash \neg B$. Thus, $\boldsymbol{M}, f(\sigma) \nVdash B$, i.e., $\boldsymbol{M}$ satisfies $\sigma\, \mathbb{F}\, B$ with respect to f.

2. The branch is expanded by applying $\neg\mathbb{F}$ to $\sigma\, \mathbb{F}\, \neg B \in \Gamma$: Exercise.

3. The branch is expanded by applying $\wedge\mathbb{T}$ to $\sigma\, \mathbb{T}\, B \wedge C \in \Gamma$, which results in two new signed formulas on the branch: $\sigma\, \mathbb{T}\, B$ and $\sigma\, \mathbb{T}\, C$. Suppose $\boldsymbol{M}, f \Vdash \Gamma$, in particular $\boldsymbol{M}, f(\sigma) \Vdash B \wedge C$. Then $\boldsymbol{M}, f(\sigma) \Vdash B$ and $\boldsymbol{M}, f(\sigma) \Vdash C$. This means that $\boldsymbol{M}$ satisfies both $\sigma\, \mathbb{T}\, B$ and $\sigma\, \mathbb{T}\, C$ with respect to f.

4. The branch is expanded by applying $\vee\mathbb{F}$ to $\mathbb{T}\, B \vee C \in \Gamma$: Exercise.

5. The branch is expanded by applying $\to\mathbb{F}$ to $\sigma\,\mathbb{F}\,B\to C \in \Gamma$: This results in two new signed formulas on the branch: $\sigma\,\mathbb{T}\,B$ and $\sigma\,\mathbb{F}\,C$. Suppose $M, f \Vdash \Gamma$, in particular $M, f(\sigma) \nVdash B \to C$. Then $M, f(\sigma) \Vdash B$ and $M, f(\sigma) \nVdash C$. This means that M, f satisfies both $\sigma\,\mathbb{T}\,B$ and $\sigma\,\mathbb{F}\,C$.

6. The branch is expanded by applying $\Box\mathbb{T}$ to $\sigma\,\mathbb{T}\,\Box B \in \Gamma$: This results in a new signed formula $\sigma.n\,\mathbb{T}\,B$ on the branch, for some $\sigma.n \in P(\Gamma)$ (since $\sigma.n$ must be used). Suppose $M, f \Vdash \Gamma$, in particular, $M, f(\sigma) \Vdash \Box B$. Since f is an interpretation of prefixes and both σ, $\sigma.n \in P(\Gamma)$, we know that $Rf(\sigma)f(\sigma.n)$. Hence, $M, f(\sigma.n) \Vdash B$, i.e., M, f satisfies $\sigma.n\,\mathbb{T}\,B$.

7. The branch is expanded by applying $\Box\mathbb{F}$ to $\sigma\,\mathbb{F}\,\Box B \in \Gamma$: This results in a new signed formula $\sigma.n\,\mathbb{F}\,A$, where $\sigma.n$ is a new prefix on the branch, i.e., $\sigma.n \notin P(\Gamma)$. Since Γ is satisfiable, there is a M and interpretation f of $P(\Gamma)$ such that $M, f \vDash \Gamma$, in particular $M, f(\sigma) \nVdash \Box B$. We have to show that $\Gamma \cup \{\sigma.n\,\mathbb{F}\,B\}$ is satisfiable. To do this, we define an interpretation of $P(\Gamma) \cup \{\sigma.n\}$ as follows:

   Since $M, f(\sigma) \nVdash \Box B$, there is a $w \in W$ such that $Rf(\sigma)w$ and $M, w \nVdash B$. Let f' be like f, except that $f'(\sigma.n) = w$. Since $f'(\sigma) = f(\sigma)$ and $Rf(\sigma)w$, we have $Rf'(\sigma)f'(\sigma.n)$, so f' is an interpretation of $P(\Gamma) \cup \{\sigma.n\}$. Obviously $M, f'(\sigma.n) \nVdash B$. Since $f(\sigma') = f'(\sigma')$ for all prefixes $\sigma' \in P(\Gamma)$, $M, f' \Vdash \Gamma$. So, M, f' satisfies $\Gamma \cup \{\sigma.n\,\mathbb{F}\,B\}$.

Now let's consider the possible inferences with two premises.

1. The branch is expanded by applying $\wedge\mathbb{F}$ to $\sigma\,\mathbb{F}\,B \wedge C \in \Gamma$, which results in two branches, a left one continuing through $\sigma\,\mathbb{F}\,B$ and a right one through $\sigma\,\mathbb{F}\,C$. Suppose $M, f \Vdash \Gamma$, in particular $M, f(\sigma) \nVdash B \wedge C$. Then $M, f(\sigma) \nVdash B$ or $M, f(\sigma) \nVdash C$. In the former case, M, f satisfies $\sigma\,\mathbb{F}\,B$, i.e., the left branch is satisfiable. In the latter, M, f satisfies $\sigma\,\mathbb{F}\,C$, i.e., the right branch is satisfiable.

2. The branch is expanded by applying $\vee\mathbb{T}$ to $\mathbb{T}\,B \vee C \in \Gamma$: Exercise.

3. The branch is expanded by applying $\rightarrow\mathbb{T}$ to $\mathbb{T}\,B \rightarrow C \in \Gamma$: Exercise. □

Corollary 6.7. *If $\Gamma \vdash A$ then $\Gamma \models A$.*

Proof. If $\Gamma \vdash A$ then for some $B_1, \ldots, B_n \in \Gamma$, $\Delta = \{1\,\mathbb{F}\,A, 1\,\mathbb{T}\,B_1, \ldots, 1\,\mathbb{T}\,B_n\}$ has a closed tableau. We want to show that $\Gamma \models A$. Suppose not, so for some $\mathfrak{M}$ and w, $\mathfrak{M}, w \Vdash B_i$ for $i = 1, \ldots, n$, but $\mathfrak{M}, w \nVdash A$. Let $f(1) = w$; then f is an interpretation of $P(\Delta)$ into $\mathfrak{M}$, and $\mathfrak{M}$ satisfies Δ with respect to f. But by Theorem 6.6, Δ is unsatisfiable since it has a closed tableau, a contradiction. So we must have $\Gamma \vdash A$ after all. □

Corollary 6.8. *If $\vdash A$ then A is true in all models.*

6.5 Rules for Other Accessibility Relations

In order to deal with logics determined by special accessibility relations, we consider the additional rules in table 6.3.

Adding these rules results in systems that are sound and complete for the logics given in table 6.4.

Example 6.9. We give a closed tableau that shows **S5** $\vdash 5$, i.e., $\Box A \rightarrow \Box\Diamond A$.

1.	$1\,\mathbb{F}\ \ \Box A \rightarrow \Box\Diamond A$	Assumption
2.	$1\,\mathbb{T}\ \ \Box A$	$\rightarrow\mathbb{F}\,1$
3.	$1\,\mathbb{F}\ \ \Box\Diamond A$	$\rightarrow\mathbb{F}\,1$
4.	$1.1\,\mathbb{F}\ \ \Diamond A$	$\Box\mathbb{F}\,3$
5.	$1\,\mathbb{F}\ \ \Diamond A$	$4\mathrm{r}\Diamond\ 4$
6.	$1.1\,\mathbb{F}\ \ A$	$\Diamond\mathbb{F}\,5$
7.	$1.1\,\mathbb{T}\ \ A$	$\Box\mathbb{T}\,2$
	⊗	

$\dfrac{\sigma\,\mathbb{T}\,\Box A}{\sigma\,\mathbb{T}\,A}\ \mathrm{T}\Box$	$\dfrac{\sigma\,\mathbb{F}\,\Diamond A}{\sigma\,\mathbb{F}\,A}\ \mathrm{T}\Diamond$
$\dfrac{\sigma\,\mathbb{T}\,\Box A}{\sigma\,\mathbb{T}\,\Diamond A}\ \mathrm{D}\Box$	$\dfrac{\sigma\,\mathbb{F}\,\Diamond A}{\sigma\,\mathbb{F}\,\Box A}\ \mathrm{D}\Diamond$
$\dfrac{\sigma.n\,\mathbb{T}\,\Box A}{\sigma\,\mathbb{T}\,A}\ \mathrm{B}\Box$	$\dfrac{\sigma.n\,\mathbb{F}\,\Diamond A}{\sigma\,\mathbb{F}\,A}\ \mathrm{B}\Diamond$
$\dfrac{\sigma\,\mathbb{T}\,\Box A}{\sigma.n\,\mathbb{T}\,\Box A}\ 4\Box$ $\sigma.n$ is used	$\dfrac{\sigma\,\mathbb{F}\,\Diamond A}{\sigma.n\,\mathbb{F}\,\Diamond A}\ 4\Diamond$ $\sigma.n$ is used
$\dfrac{\sigma.n\,\mathbb{T}\,\Box A}{\sigma\,\mathbb{T}\,\Box A}\ 4\mathrm{r}\Box$	$\dfrac{\sigma.n\,\mathbb{F}\,\Diamond A}{\sigma\,\mathbb{F}\,\Diamond A}\ 4\mathrm{r}\Diamond$

Table 6.3: More modal rules.

6.6 Soundness for Additional Rules

We say a rule is sound for a class of models if, whenever a branch in a tableau is satisfiable in a model from that class, the branch resulting from applying the rule is also satisfiable in a model from that class.

Proposition 6.10. T$\Box$ *and* T$\Diamond$ *are sound for reflexive models.*

Proof. 1. The branch is expanded by applying T$\Box$ to $\sigma\,\mathbb{T}\,\Box B \in \Gamma$: This results in a new signed formula $\sigma\,\mathbb{T}\,B$ on the branch. Suppose $M, f \Vdash \Gamma$, in particular, $M, f(\sigma) \Vdash$

Logic	R is ...	Rules
T = **KT**	reflexive	T$\Box$, T$\Diamond$
D = **KD**	serial	D$\Box$, D$\Diamond$
K4	transitive	4$\Box$, 4$\Diamond$
B = **KTB**	reflexive, symmetric	T$\Box$, T$\Diamond$ B$\Box$, B$\Diamond$
S4 = **KT4**	reflexive, transitive	T$\Box$, T$\Diamond$, 4$\Box$, 4$\Diamond$
S5 = **KT4B**	reflexive, transitive, euclidean	T$\Box$, T$\Diamond$, 4$\Box$, 4$\Diamond$, 4r$\Box$, 4r$\Diamond$

Table 6.4: Tableau rules for various modal logics.

$\Box B$. Since R is reflexive, we know that $Rf(\sigma)f(\sigma)$. Hence, $M, f(\sigma) \Vdash B$, i.e., M, f satisfies $\sigma \mathbb{T} B$.

2. The branch is expanded by applying T$\Diamond$ to $\sigma \mathbb{F} \Diamond B \in \Gamma$: Exercise. □

Proposition 6.11. D$\Box$ *and* D$\Diamond$ *are sound for serial models.*

Proof. 1. The branch is expanded by applying D$\Box$ to $\sigma \mathbb{T} \Box B \in \Gamma$: This results in a new signed formula $\sigma \mathbb{T} \Diamond B$ on the branch. Suppose $M, f \Vdash \Gamma$, in particular, $M, f(\sigma) \Vdash \Box B$. Since R is serial, there is a $w \in W$ such that $Rf(\sigma)w$. Then $M, w \Vdash B$, and hence $M, f(\sigma) \Vdash \Diamond B$. So, M, f satisfies $\sigma \mathbb{T} \Diamond B$.

2. The branch is expanded by applying D$\Diamond$ to $\sigma \mathbb{F} \Diamond B \in \Gamma$: Exercise. □

Proposition 6.12. B$\Box$ *and* B$\Diamond$ *are sound for symmetric models.*

Proof. 1. The branch is expanded by applying B$\Box$ to $\sigma.n\,\mathbb{T}\,\Box B \in \Gamma$: This results in a new signed formula $\sigma\,\mathbb{T}\,B$ on the branch. Suppose $\mathcal{M}, f \Vdash \Gamma$, in particular, $\mathcal{M}, f(\sigma.n) \Vdash \Box B$. Since f is an interpretation of prefixes on the branch into $\mathcal{M}$, we know that $Rf(\sigma)f(\sigma.n)$. Since R is symmetric, $Rf(\sigma.n)f(\sigma)$. Since $\mathcal{M}, f(\sigma.n) \Vdash \Box B$, $\mathcal{M}, f(\sigma) \Vdash B$. Hence, $\mathcal{M}, f$ satisfies $\sigma\,\mathbb{T}\,B$.

2. The branch is expanded by applying B$\Diamond$ to $\sigma.n\,\mathbb{F}\,\Diamond B \in \Gamma$: Exercise. □

Proposition 6.13. 4$\Box$ *and* 4$\Diamond$ *are sound for transitive models.*

Proof. 1. The branch is expanded by applying 4$\Box$ to $\sigma\,\mathbb{T}\,\Box B \in \Gamma$: This results in a new signed formula $\sigma.n\,\mathbb{T}\,\Box B$ on the branch. Suppose $\mathcal{M}, f \Vdash \Gamma$, in particular, $\mathcal{M}, f(\sigma) \Vdash \Box B$. Since f is an interpretation of prefixes on the branch into $\mathcal{M}$ and $\sigma.n$ must be used, we know that $Rf(\sigma)f(\sigma.n)$. Now let w be any world such that $Rf(\sigma.n)w$. Since R is transitive, $Rf(\sigma)w$. Since $\mathcal{M}, f(\sigma) \Vdash \Box B$, $\mathcal{M}, w \Vdash B$. Hence, $\mathcal{M}, f(\sigma.n) \Vdash \Box B$, and $\mathcal{M}, f$ satisfies $\sigma.n\,\mathbb{T}\,\Box B$.

2. The branch is expanded by applying 4$\Diamond$ to $\sigma\,\mathbb{F}\,\Diamond B \in \Gamma$: Exercise. □

Proposition 6.14. 4r$\Box$ *and* 4r$\Diamond$ *are sound for euclidean models.*

Proof. 1. The branch is expanded by applying 4r$\Box$ to $\sigma.n\,\mathbb{T}\,\Box B \in \Gamma$: This results in a new signed formula $\sigma\,\mathbb{T}\,\Box B$ on the branch. Suppose $\mathcal{M}, f \Vdash \Gamma$, in particular, $\mathcal{M}, f(\sigma.n) \Vdash \Box B$. Since f is an interpretation of prefixes on the branch into $\mathcal{M}$, we know that $Rf(\sigma)f(\sigma.n)$. Now let w be any world such that $Rf(\sigma)w$. Since R is euclidean, $Rf(\sigma.n)w$. Since $\mathcal{M}, f(\sigma).n \Vdash \Box B$, $\mathcal{M}, w \Vdash B$. Hence, $\mathcal{M}, f(\sigma) \Vdash \Box B$, and $\mathcal{M}, f$ satisfies $\sigma\,\mathbb{T}\,\Box B$.

2. The branch is expanded by applying $_4\mathrm{r}\Diamond$ to $\sigma.n\,\mathbb{F}\,\Diamond B \in \Gamma$: Exercise. □

Corollary 6.15. *The tableau systems given in table 6.4 are sound for the respective classes of models.*

6.7 Simple Tableaux for S5

S5 is sound and complete with respect to the class of universal models, i.e., models where every world is accessible from every world. In universal models the accessibility relation doesn't matter: "there is a world w where $M, w \Vdash A$" is true if and only if there is such a w that's accessible from u. So in **S5**, we can define models as simply a set of worlds and a valuation V. This suggests that we should be able to simplify the tableau rules as well. In the general case, we take as prefixes sequences of positive integers, so that we can keep track of which such prefixes name worlds which are accessible from others: $\sigma.n$ names a world accessible from σ. But in **S5** any world is accessible from any world, so there is no need to so keep track. Instead, we can use positive integers as prefixes. The simplified rules are given in table 6.5.

Example 6.16. We give a simplified closed tableau that shows **S5** $\vdash 5$, i.e., $\Diamond A \to \Box\Diamond A$.

1.	$1\,\mathbb{F}\ \ \Diamond A \to \Box\Diamond A$	Assumption
2.	$1\,\mathbb{T}\ \ \Diamond A$	$\to\mathbb{F}\,1$
3.	$1\,\mathbb{F}\ \ \Box\Diamond A$	$\to\mathbb{F}\,1$
4.	$2\,\mathbb{F}\ \ \Diamond A$	$\Box\mathbb{F}\,3$
5.	$3\,\mathbb{T}\ \ A$	$\Diamond\mathbb{T}\,2$
6.	$3\,\mathbb{F}\ \ A$	$\Diamond\mathbb{F}\,4$
	⊗	

$\dfrac{n\,\mathbb{T}\,\Box A}{m\,\mathbb{T}\,A}\,\Box\mathbb{T}$ m is used	$\dfrac{n\,\mathbb{F}\,\Box A}{m\,\mathbb{F}\,A}\,\Box\mathbb{F}$ m is new
$\dfrac{n\,\mathbb{T}\,\Diamond A}{m\,\mathbb{T}\,A}\,\Diamond\mathbb{T}$ m is new	$\dfrac{n\,\mathbb{F}\,\Diamond A}{m\,\mathbb{F}\,A}\,\Diamond\mathbb{F}$ m is used

Table 6.5: Simplified rules for **S5**.

6.8 Completeness for K

To show that the method of tableaux is complete, we have to show that whenever there is no closed tableau to show $\Gamma \vdash A$, then $\Gamma \nvDash A$, i.e., there is a countermodel. But "there is no closed tableau" means that every way we could try to construct one has to fail to close. The trick is to see that if every such way fails to close, then a specific, *systematic and exhaustive* way also fails to close. And this systematic and exhaustive way would close if a closed tableau exists. The single tableau will contain, among its open branches, all the information required to define a countermodel. The countermodel given by an open branch in this tableau will contain the all the prefixes used on that branch as the worlds, and a propositional variable p is true at σ iff $\sigma\,\mathbb{T}\,p$ occurs on the branch.

Definition 6.17. A branch in a tableau is called complete if, whenever it contains a prefixed formula $\sigma\, S\, A$ to which a rule can be applied, it also contains

1. the prefixed formulas that are the corresponding conclusions of the rule, in the case of propositional stacking rules;
2. one of the corresponding conclusion formulas in the case of propositional branching rules;
3. at least one possible conclusion in the case of modal rules that require a new prefix;
4. the corresponding conclusion for every prefix occurring on the branch in the case of modal rules that require a used prefix.

For instance, a complete branch contains $\sigma\,\mathbb{T}\,B$ and $\sigma\,\mathbb{T}\,C$ whenever it contains $\mathbb{T}\,B \wedge C$. If it contains $\sigma\,\mathbb{T}\,B \vee C$ it contains at least one of $\sigma\,\mathbb{F}\,B$ and $\sigma\,\mathbb{T}\,C$. If it contains $\sigma\,\mathbb{F}\,\Box$ it also contains $\sigma.n\,\mathbb{F}\,\Box$ for at least one n. And whenever it contains $\sigma\,\mathbb{T}\,\Box$ it also contains $\sigma.n\,\mathbb{T}\,\Box$ for every n such that $\sigma.n$ is used on the branch.

Proposition 6.18. *Every finite Γ has a tableau in which every branch is complete.*

Proof. Consider an open branch in a tableau for Γ. There are finitely many prefixed formulas in the branch to which a rule could be applied. In some fixed order (say, top to bottom), for each of these prefixed formulas for which the conditions (1)–(4) do not already hold, apply the rules that can be applied to it to extend the branch. In some cases this will result in branching; apply the rule at the tip of each resulting branch for all remaining prefixed formulas. Since the number of prefixed formulas is finite, and the number of used prefixes on the branch is finite, this procedure eventually results in (possibly many) branches ex-

tending the original branch. Apply the procedure to each, and repeat. But by construction, every branch is closed. □

Theorem 6.19 (Completeness). *If Γ has no closed tableau, Γ is satisfiable.*

Proof. By the proposition, Γ has a tableau in which every branch is complete. Since it has no closed tableau, it thas has a tableau in which at least one branch is open and complete. Let Δ be the set of prefixed formulas on the branch, and $P(\Delta)$ the set of prefixes occurring in it.

We define a model $\boldsymbol{M}(\Delta) = \langle P(\Delta), R, V \rangle$ where the worlds are the prefixes occurring in Δ, the accessibility relation is given by:

$$R\sigma\sigma' \quad \text{iff} \quad \sigma' = \sigma.n \quad \text{for some } n$$

and

$$V(p) = \{\sigma : \sigma\, \mathbb{T}\, p \in \Delta\}.$$

We show by induction on A that if $\sigma\, \mathbb{T}\, A \in \Delta$ then $\boldsymbol{M}(\Delta), \sigma \Vdash A$, and if $\sigma\, \mathbb{F}\, A \in \Delta$ then $\boldsymbol{M}(\Delta), \sigma \nVdash A$.

1. $A \equiv p$: If $\sigma\, \mathbb{T}\, A \in \Delta$ then $\sigma \in V(p)$ (by definition of V) and so $\boldsymbol{M}(\Delta), \sigma \Vdash A$.

   If $\sigma\, \mathbb{F}\, A \in \Delta$ then $\sigma\, \mathbb{T}\, A \notin \Delta$, since the branch would otherwise be closed. So $\sigma \notin V(p)$ and thus $\boldsymbol{M}(\Delta), \sigma \nVdash A$.

2. $A \equiv \neg B$: If $\sigma\, \mathbb{T}\, A \in \Delta$, then $\sigma\, \mathbb{F}\, B \in \Delta$ since the branch is complete. By induction hypothesis, $\boldsymbol{M}(\Delta), \sigma \nVdash B$ and thus $\boldsymbol{M}(\Delta), \sigma \Vdash A$.

   If $\sigma\, \mathbb{F}\, A \in \Delta$, then $\sigma\, \mathbb{T}\, B \in \Delta$ since the branch is complete. By induction hypothesis, $\boldsymbol{M}(\Delta), \sigma \Vdash B$ and thus $\boldsymbol{M}(\Delta), \sigma \nVdash A$.

3. $A \equiv B \wedge A$: Exercise.

4. $A \equiv B \vee A$: If $\sigma\,\mathbb{T}\,A \in \Delta$, then either $\sigma\,\mathbb{T}\,B \in \Delta$ or $\sigma\,\mathbb{T}\,C \in \Delta$ since the branch is complete. By induction hypothesis, either $\boldsymbol{M}(\Delta), \sigma \Vdash B$ or $\boldsymbol{M}(\Delta), \sigma \Vdash C$. Thus $\boldsymbol{M}(\Delta), \sigma \Vdash A$.

   If $\sigma\,\mathbb{F}\,A \in \Delta$, then both $\sigma\,\mathbb{F}\,B \in \Delta$ and $\sigma\,\mathbb{F}\,C \in \Delta$ since the branch is complete. By induction hypothesis, both $\boldsymbol{M}(\Delta), \sigma \nVdash B$ and $\boldsymbol{M}(\Delta), \sigma \nVdash B$. Thus $\boldsymbol{M}(\Delta), \sigma \nVdash A$.

5. $A \equiv B \to A$: Exercise.

6. $A \equiv \Box B$: If $\sigma\,\mathbb{T}\,A \in \Delta$, then, since the branch is complete, $\sigma.n\,\mathbb{T}\,B \in \Delta$ for every $\sigma.n$ used on the branch, i.e., for every $\sigma' \in P(\Delta)$ such that $R\sigma\sigma'$. By induction hypothesis, $\boldsymbol{M}(\Delta), \sigma' \Vdash B$ for every σ' such that $R\sigma\sigma'$. Therefore, $\boldsymbol{M}(\Delta), \sigma \Vdash A$.

   If $\sigma\,\mathbb{F}\,A \in \Delta$, then for some $\sigma.n$, $\sigma.n\,\mathbb{F}\,B \in \Delta$ since the branch is complete. By induction hypothesis, $\boldsymbol{M}(\Delta), \sigma.n \nVdash B$. Since $R\sigma(\sigma.n)$, there is a σ' such that $\boldsymbol{M}(\Delta), \sigma' \nVdash B$. Thus $\boldsymbol{M}(\Delta), \sigma \nVdash A$.

7. $A \equiv \Diamond B$: Exercise.

Since $\Gamma \subseteq \Delta$, $\boldsymbol{M}(\Delta) \Vdash \Gamma$. □

Corollary 6.20. *If $\Gamma \vDash A$ then $\Gamma \vdash A$.*

Corollary 6.21. *If A is true in all models, then $\vdash A$.*

6.9 Countermodels from Tableaux

The proof of the completeness theorem doesn't just show that if $\vDash A$ then $\vdash A$, it also gives us a method for constructing countermodels to A if $\nvDash A$. In the case of **K**, this method constitutes a *decision procedure*. For suppose $\nvDash A$. Then the proof of Proposition 6.18 gives a method for constructing a complete tableau.

The method in fact always terminates. The propositional rules for **K** only add prefixed formulas of lower complexity, i.e., each propositional rule need only be applied once on a branch for any signed formula $\sigma\, S\, A$. New prefixes are only generated by the $\Box\mathbb{F}$ and $\Diamond\mathbb{T}$ rules, and also only have to be applied once (and produce a single new prefix). $\Box\mathbb{T}$ and $\Diamond\mathbb{F}$ have to be applied potentially multiple times, but only once per prefix, and only finitely many new prefixes are generated. So the construction either results in a closed branch or a complete branch after finitely many stages.

Once a tableau with an open complete branch is constructed, the proof of Theorem 6.19 gives us an explict model that satisfies the original set of prefixed formulas. So not only is it the case that if $\Gamma \vDash A$, then a closed tableau exists and $\Gamma \vdash A$, if we look for the closed tableau in the right way and end up with a "complete" tableau, we'll not only know that $\Gamma \nvDash A$ but actually be able to construct a countermodel.

Example 6.22. We know that $\nvdash \Box(p \vee q) \to (\Box p \vee \Box q)$. The construction of a tableau begins with:

1.	$1\,\mathbb{F}\ \ \Box(p \vee q) \to (\Box p \vee \Box q)$ ✓	Assumption
2.	$1\,\mathbb{T}\ \ \Box(p \vee q)$	$\to\mathbb{F}\,1$
3.	$1\,\mathbb{F}\ \ \Box p \vee \Box q$ ✓	$\to\mathbb{F}\,1$
4.	$1\,\mathbb{F}\ \ \Box p$ ✓	$\vee\mathbb{F}\,3$
5.	$1\,\mathbb{F}\ \ \Box q$ ✓	$\vee\mathbb{F}\,3$
6.	$1.1\,\mathbb{F}\ \ p$ ✓	$\Box\mathbb{F}\,4$
7.	$1.2\,\mathbb{F}\ \ q$ ✓	$\Box\mathbb{F}\,5$

The tableau is of course not finished yet. In the next step, we consider the only line without a checkmark: the prefixed formula $1\,\mathbb{T}\,\Box(p \vee q)$ on line 2. The construction of the closed tableau says to apply the $\Box\mathbb{T}$ rule for every prefix used on the branch, i.e., for both 1.1 and 1.2:

1.	$1\mathbb{F}\ \ \Box(p \vee q) \to (\Box p \vee \Box q)\ \checkmark$	Assumption
2.	$1\mathbb{T}\ \ \Box(p \vee q)$	$\to\mathbb{F}1$
3.	$1\mathbb{F}\ \ \Box p \vee \Box q\ \checkmark$	$\to\mathbb{F}1$
4.	$1\mathbb{F}\ \ \Box p\ \checkmark$	$\vee\mathbb{F}3$
5.	$1\mathbb{F}\ \ \Box q\ \checkmark$	$\vee\mathbb{F}3$
6.	$1.1\mathbb{F}\ \ p\ \checkmark$	$\Box\mathbb{F}4$
7.	$1.2\mathbb{F}\ \ q\ \checkmark$	$\Box\mathbb{F}5$
8.	$1.1\mathbb{T}\ \ p \vee q$	$\Box\mathbb{T}2$
9.	$1.2\mathbb{T}\ \ p \vee q$	$\Box\mathbb{T}2$

Now lines 2, 8, and 9, don't have checkmarks. But no new prefix has been added, so we apply $\vee\mathbb{T}$ to lines 8 and 9, on all resulting branches (as long as they don't close):

1.	$1\mathbb{F}\ \ \Box(p \vee q) \to (\Box p \vee \Box q)\ \checkmark$	Assumption
2.	$1\mathbb{T}\ \ \Box(p \vee q)$	$\to\mathbb{F}1$
3.	$1\mathbb{F}\ \ \Box p \vee \Box q\ \checkmark$	$\to\mathbb{F}1$
4.	$1\mathbb{F}\ \ \Box p\ \checkmark$	$\vee\mathbb{F}3$
5.	$1\mathbb{F}\ \ \Box q\ \checkmark$	$\vee\mathbb{F}3$
6.	$1.1\mathbb{F}\ \ p\ \checkmark$	$\Box\mathbb{F}4$
7.	$1.2\mathbb{F}\ \ q\ \checkmark$	$\Box\mathbb{F}5$
8.	$1.1\mathbb{T}\ \ p \vee q\ \checkmark$	$\Box\mathbb{T}2$
9.	$1.2\mathbb{T}\ \ p \vee q\ \checkmark$	$\Box\mathbb{T}2$
10.	$1.1\mathbb{T}\ \ p\ \checkmark$ ⊗ $1.1\mathbb{T}\ \ q\ \checkmark$	$\vee\mathbb{T}8$
11.	$1.2\mathbb{T}\ \ p\ \checkmark$ $1.2\mathbb{T}\ \ q\ \checkmark$ ⊗	$\vee\mathbb{T}9$

There is one remaining open branch, and it is complete. From it we define the model with worlds $W = \{1, 1.1, 1.2\}$ (the only prefixes appearing on the open branch), the accessibility relation $R = \{\langle 1, 1.1\rangle, \langle 1, 1.2\rangle\}$, and the assignment $V(p) = \{1.2\}$ (because line 11 contains $1.2\mathbb{T}\,p$) and $V(q) = \{1.1\}$ (because line 10 con-

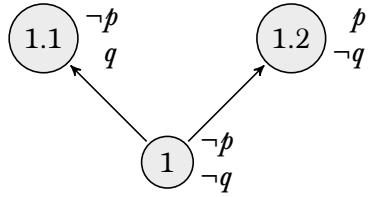


Figure 6.1: A countermodel to $\Box(p \vee q) \rightarrow (\Box p \vee \Box q)$.

tains $1.1\,\mathbb{T}\,q$). The model is pictured in Figure 6.1, and you can verify that it is a countermodel to $\Box(p \vee q) \rightarrow (\Box p \vee \Box q)$.

Problems

Problem 6.1. Find closed tableaux in **K** for the following formulas:

1. $\Box\neg p \rightarrow \Box(p \rightarrow q)$
2. $(\Box p \vee \Box q) \rightarrow \Box(p \vee q)$
3. $\Diamond p \rightarrow \Diamond(p \vee q)$

Problem 6.2. Complete the proof of Theorem 6.6.

Problem 6.3. Give closed tableaux that show the following:

1. **KT5** $\vdash$ B;
2. **KT5** $\vdash$ 4;
3. **KDB4** $\vdash$ T;
4. **KB4** $\vdash$ 5;
5. **KB5** $\vdash$ 4;
6. **KT** $\vdash$ D.

Problem 6.4. Complete the proof of Proposition 6.10

Problem 6.5. Complete the proof of Proposition 6.11

Problem 6.6. Complete the proof of Proposition 6.12

Problem 6.7. Complete the proof of Proposition 6.13

Problem 6.8. Complete the proof of Proposition 6.14

Problem 6.9. Complete the proof of Theorem 6.19.

PART II

Intuitionistic Logic

CHAPTER 7

Introduction

7.1 Constructive Reasoning

In constrast to extensions of classical logic by modal operators or second-order quantifiers, intuitionistic logic is "non-classical" in that it restricts classical logic. Classical logic is *non-constructive* in various ways. Intuitionistic logic is intended to capture a more "constructive" kind of reasoning characteristic of a kind of constructive mathematics. The following examples may serve to illustrate some of the underlying motivations.

Suppose someone claimed that they had determined a natural number n with the property that if n is even, the Riemann hypothesis is true, and if n is odd, the Riemann hypothesis is false. Great news! Whether the Riemann hypothesis is true or not is one of the big open questions of mathematics, and they seem to have reduced the problem to one of calculation, that is, to the determination of whether a specific number is even or not.

What is the magic value of n? They describe it as follows: n is the natural number that is equal to 2 if the Riemann hypothesis is true, and 3 otherwise.

Angrily, you demand your money back. From a classical point of view, the description above does in fact determine a unique value of n; but what you really want is a value of n that is given *explicitly*.

To take another, perhaps less contrived example, consider

the following question. We know that it is possible to raise an irrational number to a rational power, and get a rational result. For example, $\sqrt{2}^2 = 2$. What is less clear is whether or not it is possible to raise an irrational number to an *irrational* power, and get a rational result. The following theorem answers this in the affirmative:

Theorem 7.1. *There are irrational numbers a and b such that a^b is rational.*

Proof. Consider $\sqrt{2}^{\sqrt{2}}$. If this is rational, we are done: we can let $a = b = \sqrt{2}$. Otherwise, it is irrational. Then we have

$$(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^2 = 2,$$

which is rational. So, in this case, let a be $\sqrt{2}^{\sqrt{2}}$, and let b be $\sqrt{2}$.□

Does this constitute a valid proof? Most mathematicians feel that it does. But again, there is something a little bit unsatisfying here: we have proved the existence of a pair of real numbers with a certain property, without being able to say *which* pair of numbers it is. It is possible to prove the same result, but in such a way that the pair a, b *is* given in the proof: take $a = \sqrt{3}$ and $b = \log_3 4$. Then

$$a^b = \sqrt{3}^{\log_3 4} = 3^{1/2\cdot\log_3 4} = (3^{\log_3 4})^{1/2} = 4^{1/2} = 2,$$

since $3^{\log_3 x} = x$.

Intuitionistic logic is designed to capture a kind of reasoning where moves like the one in the first proof are disallowed. Proving the existence of an x satisfying $A(x)$ means that you have to give a specific x, and a proof that it satisfies A, like in the second proof. Proving that A or B holds requires that you can prove one or the other.

Formally speaking, intuitionistic logic is what you get if you restrict a proof system for classical logic in a certain way. From

the mathematical point of view, these are just formal deductive systems, but, as already noted, they are intended to capture a kind of mathematical reasoning. One can take this to be the kind of reasoning that is justified on a certain philosophical view of mathematics (such as Brouwer's intuitionism); one can take it to be a kind of mathematical reasoning which is more "concrete" and satisfying (along the lines of Bishop's constructivism); and one can argue about whether or not the formal description captures the informal motivation. But whatever philosophical positions we may hold, we can study intuitionistic logic as a formally presented logic; and for whatever reasons, many mathematical logicians find it interesting to do so.

7.2 Syntax of Intuitionistic Logic

The syntax of intuitionistic logic is the same as that for propositional logic. In classical propositional logic it is possible to define connectives by others, e.g., one can define $A \to B$ by $\neg A \vee B$, or $A \vee B$ by $\neg(\neg A \wedge \neg B)$. Thus, presentations of classical logic often introduce some connectives as abbreviations for these definitions. This is not so in intuitionistic logic, with two exceptions: $\neg A$ can be—and often is—defined as an abbreviation for $A \to \bot$. Then, of course, $\bot$ must not itself be defined! Also, $A \leftrightarrow B$ can be defined, as in classical logic, as $(A \to B) \wedge (B \to A)$.

Formulas of propositional intuitionistic logic are built up from *propositional variables* and the propositional constant $\bot$ using *logical connectives*. We have:

1. A countably infinite set At_0 of propositional variables p_0, $p_1, \ldots$
2. The propositional constant for falsity $\bot$.
3. The logical connectives: $\wedge$ (conjunction), $\vee$ (disjunction), $\to$ (conditional)
4. Punctuation marks: (,), and the comma.

Definition 7.2 (Formula). The set $\mathrm{Frm}(\mathscr{L}_0)$ of *formulas* of propositional intuitionistic logic is defined inductively as follows:

1. $\bot$ is an atomic formula.
2. Every propositional variable p_i is an atomic formula.
3. If A and B are formulas, then $(A \wedge B)$ is a formula.
4. If A and B are formulas, then $(A \vee B)$ is a formula.
5. If A and B are formulas, then $(A \to B)$ is a formula.
6. Nothing else is a formula.

In addition to the primitive connectives introduced above, we also use the following *defined* symbols: $\neg$ (negation) and $\leftrightarrow$ (biconditional). Formulas constructed using the defined operators are to be understood as follows:

1. $\neg A$ abbreviates $A \to \bot$.
2. $A \leftrightarrow B$ abbreviates $(A \to B) \wedge (B \to A)$.

Although $\neg$ is officially treated as an abbreviation, we will sometimes give explicit rules and clauses in definitions for $\neg$ as if it were primitive. This is mostly so we can state practice problems.

7.3 The Brouwer-Heyting-Kolmogorov Interpretation

There is an informal constructive interpretation of the intuitionist connectives, usually known as the Brouwer-Heyting-Kolmogorov interpretation. It uses the notion of a "construction," which you may think of as a constructive proof. (We don't use "proof" in the BHK interpretation so as not to get confused with the notion of a derivation in a formal proof system.) Based on this intuitive

notion, the BHK interpretation explains the meanings of the intuitionistic connectives.

1. We assume that we know what constitutes a construction of an atomic statement.
2. A construction of $A_1 \wedge A_2$ is a pair $\langle M_1, M_2 \rangle$ where M_1 is a construction of A_1 and M_2 is a construction of A_2.
3. A construction of $A_1 \vee A_2$ is a pair $\langle s, M \rangle$ where s is 1 and M is a construction of A_1, or s is 2 and M is a construction of A_2.
4. A construction of $A \to B$ is a function that converts a construction of A into a construction of B.
5. There is no construction for $\bot$ (absurdity).
6. $\neg A$ is defined as synonym for $A \to \bot$. That is, a construction of $\neg A$ is a function converting a construction of A into a construction of $\bot$.

Example 7.3. Take $\neg\bot$ for example. A construction of it is a function which, given any construction of $\bot$ as input, provides a construction of $\bot$ as output. Obviously, the identity function Id is such a construction: given a construction M of $\bot$, $\mathrm{Id}(M) = M$ yields a construction of $\bot$.

Generally speaking, $\neg A$ means "A construction of A is impossible".

Example 7.4. Let us prove $A \to \neg\neg A$ for any proposition A, which is $A \to ((A \to \bot) \to \bot)$. The construction should be a function f that, given a construction M of A, returns a construction $f(M)$ of $(A \to \bot) \to \bot$. Here is how f constructs the construction of $(A \to \bot) \to \bot$: We have to define a function g which, when given a construction h of $A \to \bot$ as input, outputs a construction of $\bot$. We can define g as follows: apply the input h

to the construction M of A (that we received earlier). Since the output $h(M)$ of h is a construction of $\bot$, $f(M)(h) = h(M)$ is a construction of $\bot$ if M is a construction of A.

Example 7.5. Let us give a construction for $\neg(A \wedge \neg A)$, i.e., $(A \wedge (A \to \bot)) \to \bot$. This is a function f which, given as input a construction M of $A \wedge (A \to \bot)$, yields a construction of $\bot$. A construction of a conjunction $B_1 \wedge B_2$ is a pair $\langle N_1, N_2 \rangle$ where N_1 is a construction of B_1 and N_2 is a construction of B_2. We can define functions p_1 and p_2 which recover from a construction of $B_1 \wedge B_2$ the constructions of B_1 and B_2, respectively:

$$p_1(\langle N_1, N_2 \rangle) = N_1$$
$$p_2(\langle N_1, N_2 \rangle) = N_2$$

Here is what f does: First it applies p_1 to its input M. That yields a construction of A. Then it applies p_2 to M, yielding a construction of $A \to \bot$. Such a construction, in turn, is a function $p_2(M)$ which, if given as input a construction of A, yields a construction of $\bot$. In other words, if we apply $p_2(M)$ to $p_1(M)$, we get a construction of $\bot$. Thus, we can define $f(M) = p_2(M)(p_1(M))$.

Example 7.6. Let us give a construction of $((A \wedge B) \to C) \to (A \to (B \to C))$, i.e., a function f which turns a construction g of $(A \wedge B) \to C$ into a construction of $(A \to (B \to C))$. The construction g is itself a function (from constructions of $A \wedge B$ to constructions of C). And the output $f(g)$ is a function h_g from constructions of A to functions from constructions of B to constructions of C.

Ok, this is confusing. We have to construct a certain function h_g, which will be the output of f for input g. The input of h_g is a construction M of A. The output of $h_g(M)$ should be a function k_M from constructions N of B to constructions of C. Let $k_{g,M}(N) = g(\langle M, N \rangle)$. Remember that $\langle M, N \rangle$ is a construction of $A \wedge B$. So $k_{g,M}$ is a construction of $B \to C$: it maps constructions N of B to constructions of C. Now let $h_g(M) = k_{g,M}$. That's a function that maps constructions M of A to constructions $k_{g,M}$

of $B \to C$. Now let $f(g) = h_g$. That's a function that maps constructions g of $(A \wedge B) \to C$ to constructions of $A \to (B \to C)$. Whew!

The statement $A \vee \neg A$ is called the Law of Excluded Middle. We can prove it for some specific A (e.g., $\bot \vee \neg\bot$), but not in general. This is because the intuitionistic disjunction requires a construction of one of the disjuncts, but there are statements which currently can neither be proved nor refuted (say, Goldbach's conjecture). However, you can't refute the law of excluded middle either: that is, $\neg\neg(A \vee \neg A)$ holds.

Example 7.7. To prove $\neg\neg(A \vee \neg A)$, we need a function f that transforms a construction of $\neg(A \vee \neg A)$, i.e., of $(A \vee (A \to \bot)) \to \bot$, into a construction of $\bot$. In other words, we need a function f such that $f(g)$ is a construction of $\bot$ if g is a construction of $\neg(A \vee \neg A)$.

Suppose g is a construction of $\neg(A \vee \neg A)$, i.e., a function that transforms a construction of $A \vee \neg A$ into a construction of $\bot$. A construction of $A \vee \neg A$ is a pair $\langle s, M\rangle$ where either $s = 1$ and M is a construction of A, or $s = 2$ and M is a construction of $\neg A$. Let h_1 be the function mapping a construction M_1 of A to a construction of $A \vee \neg A$: it maps M_1 to $\langle 1, M_2\rangle$. And let h_2 be the function mapping a construction M_2 of $\neg A$ to a construction of $A \vee \neg A$: it maps M_2 to $\langle 2, M_2\rangle$.

Let k be $g \circ h_1$: it is a function which, if given a construction of A, returns a construction of $\bot$, i.e., it is a construction of $A \to \bot$ or $\neg A$. Now let l be $g \circ h_2$. It is a function which, given a construction of $\neg A$, provides a construction of $\bot$. Since k is a construction of $\neg A$, $l(k)$ is a construction of $\bot$.

Together, what we've done is describe how we can turn a construction g of $\neg(A \vee \neg A)$ into a construction of $\bot$, i.e., the function f mapping a construction g of $\neg(A \vee \neg A)$ to the construction $l(k)$ of $\bot$ is a construction of $\neg\neg(A \vee \neg A)$.

As you can see, using the BHK interpretation to show the intuitionistic validity of formulas quickly becomes cumbersome

and confusing. Luckily, there are better derivation systems for intuitionistic logic, and more precise semantic interpretations.

7.4 Natural Deduction

Natural deduction without the $\bot_C$ rules is a standard derivation system for intuitionistic logic. We repeat the rules here and indicate the motivation using the BHK interpretation. In each case, we can think of a rule which allows us to conclude that if the premises have constructions, so does the conclusion.

Since natural deduction derivations have undischarged assumptions, we should consider such a derivation, say, of A from undischarged assumptions Γ, as a function that turns constructions of all $B \in \Gamma$ into a construction of A. If there is a derivation of A from no undischarged assumptions, then there is a construction of A in the sense of the BHK interpretation. For the purpose of the discussion, however, we'll suppress the Γ when not needed.

An assumption A by itself is a derivation of A from the undischarged assumption A. This agrees with the BHK-interpretation: the identity function on constructions turns any construction of A into a construction of A.

Conjunction

$$\frac{A \qquad B}{A \wedge B}\,\wedge\text{Intro} \qquad \frac{A \wedge B}{A}\,\wedge\text{Elim} \qquad \frac{A \wedge B}{B}\,\wedge\text{Elim}$$

Suppose we have constructions N_1, N_2 of A_1 and A_2, respectively. Then we also have a construction $A_1 \wedge A_2$, namely the pair $\langle N_1, N_2 \rangle$.

A construction of $A_1 \wedge A_1$ on the BHK interpretation is a pair $\langle N_1, N_2 \rangle$. So assume we have such a pair. Then we also have a

construction of each conjunct: N_1 is a construction of A_1 and N_2 is a construction of A_2.

Conditional

$$\begin{array}{c}[A]^u \\ \vdots \\ B \\ \hline A \to B\ {}^u\end{array} \to\text{Intro} \qquad \frac{A \to B \quad A}{B} \to\text{Elim}$$

If we have a derivation of B from undischarged assumption A, then there is a function f that turns constructions of A into constructions of B. That same function is a construction of $A \to B$. So, if the premise of $\to$Intro has a construction conditional on a construction of A, the conclusion $A \to B$ has a construction.

On the other hand, suppose there are constructions N of A and f of $A \to B$. A construction of $A \to B$ is a function that turns constructions of A into constructions of B. So, $f(N)$ is a construction of B, i.e., the conclusion of $\to$Elim has a construction.

Disjunction

$$\frac{A}{A \vee B} \vee\text{Intro} \qquad \frac{B}{A \vee B} \vee\text{Intro}$$

$$\begin{array}{ccc} & [A]^n & [B]^n \\ & \vdots & \vdots \\ A \vee B & C & C \\ \hline & C & \end{array}\ {}^n \vee\text{Elim}$$

If we have a construction N_i of A_i we can turn it into a construction $\langle i, N_i \rangle$ of $A_1 \vee A_2$. On the other hand, suppose we have a construction of $A_1 \vee A_2$, i.e., a pair $\langle i, N_i \rangle$ where N_i is a construction of A_i, and also functions f_1, f_2, which turn constructions of A_1, A_2, respectively, into constructions of C. Then $f_i(N_i)$ is a construction of C, the conclusion of $\vee$Elim.

Absurdity

$$\frac{\bot}{A}\,\bot_I$$

If we have a derivation of $\bot$ from undischarged assumptions $B_1, \ldots, B_n$, then there is a function $f(M_1, \ldots, M_n)$ that turns constructions of $B_1, \ldots, B_n$ into a construction of $\bot$. Since $\bot$ has no construction, there cannot be any constructions of all of $B_1, \ldots, B_n$ either. Hence, f also has the property that *if* $M_1, \ldots, M_n$ are constructions of $B_1, \ldots, B_n$, respectively, *then* $f(M_1, \ldots, M_n)$ is a construction of A.

Rules for ¬

Since $\neg A$ is defined as $A \to \bot$, we strictly speaking do not need rules for ¬. But if we did, this is what they'd look like:

$$\begin{array}{c} [A]^n \\ \vdots \\ \bot \end{array} \Big/ \neg A \;\neg\text{Intro}_n \qquad\qquad \frac{\neg A \quad A}{\bot}\,\neg\text{Elim}$$

Examples of Derivations

1. $\vdash A \to (\neg A \to \bot)$, i.e., $\vdash A \to ((A \to \bot) \to \bot)$

$$\dfrac{\dfrac{\dfrac{[A]^2 \qquad [A \to \bot]^1}{\bot}\,\to\text{Elim}}{(A \to \bot) \to \bot}\,\to\text{Intro}_1}{A \to (A \to \bot) \to \bot}\,\to\text{Intro}_2$$

2. $\vdash ((A \wedge B) \to C) \to (A \to (B \to C))$

$$\dfrac{\dfrac{\dfrac{\dfrac{[(A \wedge B) \to C]^3 \qquad \dfrac{[A]^2 \qquad [B]^1}{A \wedge B}\,\wedge\text{Intro}}{C}\,\to\text{Elim}}{B \to C}\,{}^1\to\text{Intro}}{A \to (B \to C)}\,{}^2\to\text{Intro}}{((A \wedge B) \to C) \to (A \to (B \to C))}\,{}^3\to\text{Intro}$$

3. $\vdash \neg(A \wedge \neg A)$, i.e., $\vdash (A \wedge (A \to \bot)) \to \bot$

$$\dfrac{\dfrac{\dfrac{[A \wedge (A \to \bot)]^1}{A \to \bot}\,\wedge\text{Elim} \qquad \dfrac{[A \wedge (A \to \bot)]^1}{A}\,\wedge\text{Elim}}{\bot}\,\to\text{Elim}}{(A \wedge (A \to \bot)) \to \bot}\,{}^1\to\text{Intro}$$

4. $\vdash \neg\neg(A \vee \neg A)$, i.e., $\vdash ((A \vee (A \to \bot)) \to \bot) \to \bot$

$$\dfrac{\dfrac{\dfrac{[(A \vee (A \to \bot)) \to \bot]^2 \qquad \dfrac{\dfrac{[A]^1}{A \vee (A \to \bot)}\,\vee\text{Intro}}{\bot}\,\to\text{Elim}}{A \to \bot}\,{}^1\to\text{Intro}}{A \vee (A \to \bot)}\,\vee\text{Intro} \qquad [(A \vee (A \to \bot)) \to \bot]^2}{\bot}\,\to\text{Elim}}{((A \vee (A \to \bot)) \to \bot) \to \bot}\,{}^2\to\text{Intro}$$

Proposition 7.8. *If $\Gamma \vdash A$ in intuitionistic logic, $\Gamma \vdash A$ in classical logic. In particular, if A is an intuitionistic theorem, it is also a classical theorem.*

Proof. Every natural deduction rule is also a rule in classical natural deduction, so every derivation in intuitionistic logic is also a derivation in classical logic. □

7.5 Axiomatic Derivations

Axiomatic derivations for intuitionistic propositional logic are the conceptually simplest, and historically first, derivation systems. They work just as in classical propositional logic.

Definition 7.9 (Derivability). If Γ is a set of formulas of $\mathcal{L}$ then a *derivation* from Γ is a finite sequence $A_1, \ldots, A_n$ of formulas where for each $i \leq n$ one of the following holds:

1. $A_i \in \Gamma$; or
2. A_i is an axiom; or
3. A_i follows from some A_j and A_k with $j < i$ and $k < i$ by modus ponens, i.e., $A_k \equiv A_j \to A_i$.

Definition 7.10 (Axioms). The set of Ax_0 of *axioms* for the intuitionistic propositional logic are all formulas of the following forms:

$$(A \wedge B) \to A \tag{7.1}$$
$$(A \wedge B) \to B \tag{7.2}$$
$$A \to (B \to (A \wedge B)) \tag{7.3}$$
$$A \to (A \vee B) \tag{7.4}$$
$$A \to (B \vee A) \tag{7.5}$$
$$(A \to C) \to ((B \to C) \to ((A \vee B) \to C)) \tag{7.6}$$
$$A \to (B \to A) \tag{7.7}$$
$$(A \to (B \to C)) \to ((A \to B) \to (A \to C)) \tag{7.8}$$
$$\bot \to A \tag{7.9}$$

Definition 7.11 (Derivability). A formula A is *derivable* from Γ, written $\Gamma \vdash A$, if there is a derivation from Γ ending in A.

Definition 7.12 (Theorems). A formula A is a *theorem* if there is a derivation of A from the empty set. We write $\vdash A$ if A is a theorem and $\nvdash A$ if it is not.

Proposition 7.13. *If $\Gamma \vdash A$ in intuitionistic logic, $\Gamma \vdash A$ in classical logic. In particular, if A is an intuitionistic theorem, it is also a classical theorem.*

Proof. Every intuitionistic axiom is also a classical axiom, so every derivation in intuitionistic logic is also a derivation in classical logic. □

Problems

Problem 7.1. Give derivations in intutionistic logic of the following.

1. $(\neg A \vee B) \to (A \to B)$
2. $\neg\neg\neg A \to \neg A$
3. $\neg\neg(A \wedge B) \leftrightarrow (\neg\neg A \wedge \neg\neg B)$

CHAPTER 8

Semantics

8.1 Introduction

No logic is satisfactorily described without a semantics, and intuitionistic logic is no exception. Whereas for classical logic, the semantics based on valuations is canonical, there are several competing semantics for intuitionistic logic. None of them are completely satisfactory in the sense that they give an intuitionistically acceptable account of the meanings of the connectives.

The semantics based on relational models, similar to the semantics for modal logics, is perhaps the most popular one. In this semantics, propositional variables are assigned to worlds, and these worlds are related by an accessibility relation. That relation is always a partial order, i.e., it is reflexive, antisymmetric, and transitive.

Intuitively, you might think of these worlds as states of knowledge or "evidentiary situations." A state w' is accessible from w iff, for all we know, w' is a possible (future) state of knowledge, i.e., one that is compatible with what's known at w. Once a proposition is known, it can't become un-known, i.e., whenever A is known at w and Rww', A is known at w' as well. So "knowledge" is monotonic with respect to the accessibility relation.

If we define "A is known" as in epistemic logic as "true in all epistemic alternatives," then $A \wedge B$ is known at w if in all epistemic alternatives, both A and B are known. But since knowledge is

monotonic and R is reflexive, that means that $A \wedge B$ is known at w iff A and B are known at w. For the same reason, $A \vee B$ is known at w iff at least one of them is known. So for $\wedge$ and $\vee$, the truth conditions of the connectives coincide with those in classical logic.

The truth conditions for the conditional, however, differ from classical logic. $A \rightarrow B$ is known at w iff at no w' with Rww', A is known without B also being known. This is not the same as the condition that A is unknown or B is known at w. For if we know neither A nor B at w, there might be a future epistemic state w' with Rww' such that at w', A is known without also coming to know B.

We know $\neg A$ only if there is no possible future epistemic state in which we know A. Here the idea is that if A were knowable, then in some possible future epistemic state A becomes known. Since we can't know $\bot$, in that future epistemic state, we would know A but not know $\bot$.

On this interpretation the principle of excluded middle fails. For there are some A which we don't yet know, but which we might come to know. For such an A, both A and $\neg A$ are unknown, so $A \vee \neg A$ is not known. But we do know, e.g., that $\neg(A \wedge \neg A)$. For no future state in which we know both A and $\neg A$ is possible, and we know this independently of whether or not we know A or $\neg A$.

Relational models are not the only available semantics for intuitionistic logic. The topological semantics is another: here propositions are interpreted as open sets in a topological space, and the connectives are interpreted as operations on these sets (e.g., $\wedge$ corresponds to intersection).

8.2 Relational models

In order to give a precise semantics for intuitionistic propositional logic, we have to give a definition of what counts as a model relative to which we can evaluate formulas. On the basis of such a definition it is then also possible to define semantics notions

such as validity and entailment. One such semantics is given by relational models.

Definition 8.1. A relational model for intuitionistic propositional logic is a triple $\boldsymbol{M} = \langle W, R, V \rangle$, where

1. W is a non-empty set,
2. R is a partial order (i.e., a reflexive, antisymmetric, and transitive binary relation) on W, and
3. V is a function assigning to each propositional variable p a subset of W, such that
4. V is monotone with respect to R, i.e., if $w \in V(p)$ and Rww', then $w' \in V(p)$.

Definition 8.2. We define the notion of A *being true at* w *in* $\boldsymbol{M}$, $\boldsymbol{M}, w \Vdash A$, inductively as follows:

1. $A \equiv p$: $\boldsymbol{M}, w \Vdash A$ iff $w \in V(p)$.
2. $A \equiv \bot$: not $\boldsymbol{M}, w \Vdash A$.
3. $A \equiv \neg B$: $\boldsymbol{M}, w \Vdash A$ iff for no w' such that Rww', $\boldsymbol{M}, w' \Vdash B$.
4. $A \equiv B \wedge C$: $\boldsymbol{M}, w \Vdash A$ iff $\boldsymbol{M}, w \Vdash B$ and $\boldsymbol{M}, w \Vdash C$.
5. $A \equiv B \vee C$: $\boldsymbol{M}, w \Vdash A$ iff $\boldsymbol{M}, w \Vdash B$ or $\boldsymbol{M}, w \Vdash C$ (or both).
6. $A \equiv B \to C$: $\boldsymbol{M}, w \Vdash A$ iff for every w' such that Rww', not $\boldsymbol{M}, w' \Vdash B$ or $\boldsymbol{M}, w' \Vdash C$ (or both).

We write $\boldsymbol{M}, w \nVdash A$ if not $\boldsymbol{M}, w \Vdash A$. If Γ is a set of formulas, $\boldsymbol{M}, w \Vdash \Gamma$ means $\boldsymbol{M}, w \Vdash B$ for all $B \in \Gamma$.

Proposition 8.3. *Truth at worlds is monotonic with respect to R, i.e., if $M, w \Vdash A$ and Rww', then $M, w' \Vdash A$.*

Proof. Exercise. □

8.3 Semantic Notions

Definition 8.4. We say A is *true in the model* $M = \langle W, R, V \rangle$, $M \Vdash A$, iff $M, w \Vdash A$ for all $w \in W$. A is *valid*, $\vDash A$, iff it is true in all models. We say a set of formulas Γ *entails* A, $\Gamma \vDash A$, iff for every model M and every w such that $M, w \Vdash \Gamma$, $M, w \Vdash A$.

Proposition 8.5. *1. If $M, w \Vdash \Gamma$ and $\Gamma \vDash A$, then $M, w \Vdash A$.*

2. If $M \Vdash \Gamma$ and $\Gamma \vDash A$, then $M \Vdash A$.

Proof. 1. Suppose $M \Vdash \Gamma$. Since $\Gamma \vDash A$, we know that if $M, w \Vdash \Gamma$, then $M, w \Vdash A$. Since $M, u \Vdash \Gamma$ for all every $u \in W$, $M, w \Vdash \Gamma$. Hence $M, w \Vdash A$.

2. Follows immediately from (1). □

8.4 Topological Semantics

Another way to provide a semantics for intuitionistic logic is using the mathematical concept of a topology.

Definition 8.6. Let X be a set. A *topology on* X is a set $\mathcal{O} \subseteq \wp(X)$ that satisfies the properties below. The elements of $\mathcal{O}$ are called the *open sets* of the topology. The set X together with $\mathcal{O}$ is called a *topological space*.

1. The empty set and the entire space open: $\emptyset, X \in \mathcal{O}$.

2. Open sets are closed under finite intersections: if $U, V \in \mathcal{O}$ then $U \cap V \in \mathcal{O}$

3. Open sets are closed under arbitrary unions: if $U_i \in \mathcal{O}$ for all $i \in I$, then $\bigcup\{U_i : i \in I\} \in \mathcal{O}$.

We may write X for a topology if the collection of open sets can be inferred from the context; note that, still, only after X is endowed with open sets can it be called a topology.

Definition 8.7. A *topological model* of intuitionistic propositional logic is a triple $\mathbf{X} = \langle X, \mathcal{O}, V\rangle$ where $\mathcal{O}$ is a topology on X and V is a function assigning an open set in $\mathcal{O}$ to each propositional variable.

Given a topological model $\mathbf{X}$, we can define $[A]_X$ inductively as follows:

1. $V(\bot) = \emptyset$
2. $[p]_X = V(p)$
3. $[A \wedge B]_X = [A]_X \cap [B]_X$
4. $[A \vee B]_X = [A]_X \cup [B]_X$
5. $[A \to B]_X = \mathrm{Int}((X \setminus [A]_X) \cup [B]_X)$

Here, $\mathrm{Int}(V)$ is the function that maps a set $V \subseteq X$ to its *interior*, that is, the union of all open sets it contains. In other words,

$$\mathrm{Int}(V) = \bigcup\{U : U \subseteq V \text{ and } U \in \mathcal{O}\}.$$

Note that the interior of any set is always open, since it is a union of open sets. Thus, $[A]_X$ is always an open set.

Although topological semantics is highly abstract, there are ways to think about it that might motivate it. Suppose that the elements, or "points," of X are points at which statements can be evaluated. The set of all points where A is true is the proposition

expressed by A. Not every set of points is a potential proposition; only the elements of $\mathfrak{O}$ are. $A \vDash B$ iff B is true at every point at which A is true, i.e., $[A]_X \subseteq [B]_X$, for all X. The absurd statement $\bot$ is never true, so $[\bot]_X = \emptyset$. How must the propositions expressed by $B \wedge C$, $B \vee C$, and $B \to C$ be related to those expressed by B and C for the intuitionistically valid laws to hold, i.e., so that $A \vdash B$ iff $[A]_X \subset [B]_X$. $\bot \vdash A$ for any A, and only $\emptyset \subseteq U$ for all U. Since $B \wedge C \vdash B$, $[B \wedge C]_X \subseteq [B]_X$, and similarly $[B \wedge C]_X \subseteq [C]_X$. The largest set satisfying $W \subseteq U$ and $W \subseteq V$ is $U \cap V$. Conversely, $B \vdash B \vee C$ and $C \vdash B \vee C$, and so $[B]_X \subseteq [B \vee C]_X$ and $[C]_X \subseteq [B \vee C]_X$. The smallest set W such that $U \subseteq W$ and $V \subseteq W$ is $U \cup V$. The definition for $\to$ is tricky: $A \to B$ expresses the weakest proposition that, combined with A, entails B. That $A \to B$ combined with A entails B is clear from $(A \to B) \wedge A \vdash B$. So $[A \to B]_X$ should be the greatest open set such that $[A \to B]_X \cap [A]_X \subset [B]_X$, leading to our definition.

Problems

Problem 8.1. Show that according to Definition 8.2, $M, w \Vdash \neg A$ iff $M, w \Vdash A \to \bot$.

Problem 8.2. Prove Proposition 8.3.

CHAPTER 9

Soundness and Completeness

9.1 Soundness of Axiomatic Derivations

Theorem 9.1 (Soundness). *If $\Gamma \vdash A$, then $\Gamma \vDash A$.*

Proof. We prove that if $\Gamma \vdash A$, then $\Gamma \vDash A$. The proof is by induction on the number n of formulas in the derivation of A from Γ. We show that if $A_1, \ldots, A_n = A$ is a derivation from Γ, then $\Gamma \vDash A_n$. Note that if $A_1, \ldots, A_n$ is a derivation, so is $A_1, \ldots, A_k$ for any $k < n$.

There are no derivations of length 0, so for $n = 0$ the claim holds vacuously. So the claim holds for all derivations of length $< n$. We distinguish cases according to the justification of A_n.

1. A_n is an axiom. All axioms are valid, so $\Gamma \vDash A_n$ for any Γ.

2. $A_n \in \Gamma$. Then for any $\boldsymbol{M}$ and w, if $\boldsymbol{M}, w \Vdash \Gamma$, obviously $\boldsymbol{M} \Vdash \Gamma A_n[w]$, i.e., $\Gamma \vDash A$.

3. A_n follows by MP from A_i and $A_j \equiv A_i \to A_n$. $A_1, \ldots, A_i$ and $A_1, \ldots, A_j$ are derivations from Γ, so by inductive hypothesis, $\Gamma \vDash A_i$ and $\Gamma \vDash A_i \to A_n$.

Suppose $M, w \Vdash \Gamma$. Since $M, w \Vdash \Gamma$ and $\Gamma \vDash A_i \to A_n$, $M, w \Vdash A_i \to A_n$. By definition, this means that for all w' such that Rww', if $M, w' \Vdash A_i$ then $M, w' \Vdash A_n$. Since R is reflexive, w is among the w' such that Rww', i.e., we have that if $M, w \Vdash A_i$ then $M, w \Vdash A_n$. Since $\Gamma \vDash A_i$, $M, w \Vdash A_i$. So, $M, w \Vdash A_n$, as we wanted to show. □

9.2 Soundness of Natural Deduction

Theorem 9.2 (Soundness). *If $\Gamma \vdash A$, then $\Gamma \vDash A$.*

Proof. We prove that if $\Gamma \vdash A$, then $\Gamma \vDash A$. The proof is by induction on the derivation of A from Γ.

1. If the derivation consists of just the assumption A, we have $A \vdash A$, and want to show that $A \vDash A$. Consider any model M such that $M \Vdash A$. Then trivially $M \Vdash A$.

2. The derivation ends in ∧Intro: Exercise.

3. The derivation ends in ∧Elim: Exercise.

4. The derivation ends in ∨Intro: Suppose the premise is B, and the undischarged assumptions of the derivation ending in B are Γ. Then we have $\Gamma \vdash B$ and by inductive hypothesis, $\Gamma \vDash B$. We have to show that $\Gamma \vDash B \vee C$. Suppose $M \Vdash \Gamma$. Since $\Gamma \vDash B$, $M \Vdash B$. But then also $M \Vdash B \vee C$. Similarly, if the premise is C, we have that $\Gamma \vDash C$.

5. The derivation ends in ∨Elim: The derivations ending in the premises are of $B \vee C$ from undischarged assumptions Γ, of D from undischarged assumptions $\Delta_1 \cup \{B\}$, and of D from undischarged assumptions $\Delta_2 \cup \{C\}$. So we have $\Gamma \vdash B \vee C$, $\Delta_1 \cup \{B\} \vdash D$, and $\Delta_2 \cup \{C\} \vdash D$. By induction hypothesis, $\Gamma \vDash B \vee C$, $\Delta_1 \cup \{B\} \vDash D$, and $\Delta_2 \cup \{C\} \vDash D$. We have to prove that $\Gamma \cup \Delta_1 \cup \Delta_2 \vDash D$.

Suppose $M \Vdash \Gamma \cup \Delta_1 \cup \Delta_2$. Then $M \Vdash \Gamma$ and since $\Gamma \vDash B \vee C$, $M \Vdash B \vee C$. By definition of $M \Vdash$, either $M \Vdash B$ or $M \Vdash C$. So we distinguish cases: (a) $M \Vdash B$. Then $M \Vdash \Delta_1 \cup \{B\}$. Since $\Delta_1 \cup B \vDash D$, we have $M \Vdash D$. (b) $M \Vdash C$. Then $M \Vdash \Delta_2 \cup \{C\}$. Since $\Delta_2 \cup C \vDash D$, we have $M \Vdash D$. So in either case, $M \Vdash D$, as we wanted to show.

6. The derivation ends with $\to$Intro concluding $B \to C$. Then the premise is C, and the derivation ending in the premise has undischarged assumptions $\Gamma \cup \{B\}$. So we have that $\Gamma \cup \{B\} \vdash C$, and by induction hypothesis that $\Gamma \cup \{B\} \vDash C$. We have to show that $\Gamma \vDash B \to C$.

   Suppose $M, w \Vdash \Gamma$. We want to show that for all w' such that Rww', if $M, w' \Vdash B$, then $M, w' \Vdash C$. So assume that Rww' and $M, w' \Vdash B$. By Proposition 8.3, $M, w' \Vdash \Gamma$. Since $\Gamma \cup \{B\} \vDash C$, $M, w' \Vdash C$, which is what we wanted to show.

7. The derivation ends in $\to$Elim and conclusion C. The premises are $B \to C$ and B, with derivations from undischarged assumptions Γ, Δ. So we have $\Gamma \vdash B \to C$ and $\Delta \vdash B$. By inductive hypothesis, $\Gamma \vDash B \to C$ and $\Delta \vDash B$. We have to show that $\Gamma \cup \Delta \vDash C$.

   Suppose $M, w \Vdash \Gamma \cup \Delta$. Since $M, w \Vdash \Gamma$ and $\Gamma \vDash B \to C$, $M, w \Vdash B \to C$. By definition, this means that for all w' such that Rww', if $M, w' \Vdash B$ then $M, w' \Vdash C$. Since R is reflexive, w is among the w' such that Rww', i.e., we have that if $M, w \Vdash B$ then $M, w \Vdash C$. Since $M, w \Vdash \Delta$ and $\Delta \vDash B$, $M, w \Vdash B$. So, $M, w \Vdash C$, as we wanted to show.

8. The derivation ends in $\bot_I$, concluding A. The premise is $\bot$ and the undischarged assumptions of the derivation of the premise are Γ. Then $\Gamma \vdash \bot$. By inductive hypothesis, $\Gamma \vDash \bot$. We have to show $\Gamma \vDash A$.

   We proceed indirectly. If $\Gamma \nvDash A$ there is a model M and world w such that $M, w \Vdash \Gamma$ and $M, w \nVdash A$. Since $\Gamma \vDash$

$\bot$, $M, w \Vdash \bot$. But that's impossible, since by definition, $M, w \nVdash \bot$. So $\Gamma \vDash A$.

9. The derivation ends in ¬Intro: Exercise.

10. The derivation ends in ¬Elim: Exercise. □

9.3 Lindenbaum's Lemma

Definition 9.3. A set of formulas Γ is *prime* iff

1. Γ is consistent.
2. If $\Gamma \vdash A$ then $A \in \Gamma$, and
3. If $A \vee B \in \Gamma$ then $A \in \Gamma$ or $B \in \Gamma$.

Lemma 9.4 (Lindenbaum's Lemma). *If $\Gamma \nvdash A$, there is a $\Gamma^* \supseteq \Gamma$ such that Γ^* is prime and $\Gamma^* \nvdash A$.*

Proof. Let $B_1 \vee C_1$, $B_2 \vee C_2$, ..., be an enumeration of all formulas of the form $B \vee C$. We'll define an increasing sequence of sets of formulas Γ_n, where each Γ_{n+1} is defined as Γ_n together with one new formula. Γ^* will be the union of all Γ_n. The new formulas are selected so as to ensure that Γ^* is prime and still $\Gamma^* \nvdash A$. This means that at each step we should find the first disjunction $B_i \vee C_i$ such that:

1. $\Gamma_n \vdash B_i \vee C_i$
2. $B_i \notin \Gamma_n$ and $C_i \notin \Gamma_n$

We add to Γ_n either B_i if $\Gamma_n \cup \{B_i\} \nvdash A$, or C_i otherwise. We'll have to show that this works. For now, let's define $i(n)$ as the least i such that (1) and (2) hold.

Define $\Gamma_0 = \Gamma$ and

$$\Gamma_{n+1} = \begin{cases} \Gamma_n \cup \{B_{i(n)}\} & \text{if } \Gamma_n \cup \{B_{i(n)}\} \nvdash A \\ \Gamma_n \cup \{C_{i(n)}\} & \text{otherwise} \end{cases}$$

If $i(n)$ is undefined, i.e., whenever $\Gamma_n \vdash B \vee C$, either $B \in \Gamma_n$ or $C \in \Gamma_n$, we let $\Gamma_{n+1} = \Gamma_n$. Now let $\Gamma^* = \bigcup_{n=0}^{\infty} \Gamma_n$

First we show that for all n, $\Gamma_n \nvdash A$. We proceed by induction on n. For $n = 0$ the claim holds by the hypothesis of the theorem, i.e., $\Gamma \nvdash A$. If $n > 0$, we have to show that if $\Gamma_n \nvdash A$ then $\Gamma_{n+1} \nvdash A$. If $i(n)$ is undefined, $\Gamma_{n+1} = \Gamma_n$ and there is nothing to prove. So suppose $i(n)$ is defined. For simplicity, let $i = i(n)$.

We'll prove the contrapositive of the claim. Suppose $\Gamma_{n+1} \vdash A$. By construction, $\Gamma_{n+1} = \Gamma_n \cup \{B_i\}$ if $\Gamma_n \cup \{B_i\} \nvdash A$, or else $\Gamma_{n+1} = \Gamma_n \cup \{C_i\}$. It clearly can't be the first, since then $\Gamma_{n+1} \nvdash A$. Hence, $\Gamma_n \cup \{B_i\} \vdash A$ and $\Gamma_{n+1} = \Gamma_n \cup \{C_i\}$. By definition of $i(n)$, we have that $\Gamma_n \vdash B_i \vee C_i$. We have $\Gamma_n \cup \{B_i\} \vdash A$. We also have $\Gamma_{n+1} = \Gamma_n \cup \{C_i\} \vdash A$. Hence, $\Gamma_n \vdash A$, which is what we wanted to show.

If $\Gamma^* \vdash A$, there would be some finite subset $\Gamma' \subseteq \Gamma^*$ such that $\Gamma' \vdash A$. Each $D \in \Gamma'$ must be in Γ_i for some i. Let n be the largest of these. Since $\Gamma_i \subseteq \Gamma_n$ if $i \leq n$, $\Gamma' \subseteq \Gamma_n$. But then $\Gamma_n \vdash A$, contrary to our proof above that $\Gamma_n \nvdash A$.

Lastly, we show that Γ^* is prime, i.e., satisfies conditions (1), (2), and (3) of Definition 9.3.

First, $\Gamma^* \nvdash A$, so Γ^* is consistent, so (1) holds.

We now show that if $\Gamma^* \vdash B \vee C$, then either $B \in \Gamma^*$ or $C \in \Gamma^*$. This proves (3), since if $B \in \Gamma^*$ then also $\Gamma^* \vdash B$, and similarly for C. So assume $\Gamma^* \vdash B \vee C$ but $B \notin \Gamma^*$ and $C \notin \Gamma^*$. Since $\Gamma^* \vdash B \vee C$, $\Gamma_n \vdash B \vee C$ for some n. $B \vee C$ appears on the enumeration of all disjunctions, say as $B_j \vee C_j$. $B_j \vee C_j$ satisfies the properties in the definition of $i(n)$, namely we have $\Gamma_n \vdash B_j \vee C_j$, while $B_j \notin \Gamma_n$ and $C_j \notin \Gamma_n$. At each stage, at least one fewer disjunction $B_i \vee C_i$ satisfies the conditions (since at each stage we add either B_i or C_i), so at some stage m we will have $j = i(\Gamma_m)$. But then either $B \in \Gamma_{m+1}$ or $C \in \Gamma_{m+1}$, contrary to the assumption that $B \notin \Gamma^*$ and $C \notin \Gamma^*$.

Now suppose $\Gamma^* \vdash A$. Then $\Gamma^* \vdash A \vee A$. But we've just proved that if $\Gamma^* \vdash A \vee A$ then $A \in \Gamma^*$. Hence, Γ^* satisfies (2) of Definition 9.3. □

9.4 The Canonical Model

The worlds in our model will be finite sequences σ of natural numbers, i.e., $\sigma \in \mathbb{N}^*$. Note that $\mathbb{N}^*$ is inductively defined by:

1. $\Lambda \in \mathbb{N}^*$.
2. If $\sigma \in \mathbb{N}^*$ and $n \in \mathbb{N}$, then $\sigma.n \in \mathbb{N}^*$ (where $\sigma.n$ is $\sigma \frown \langle n \rangle$ and $\sigma \frown \sigma'$ is the concatenation if σ and σ').
3. Nothing else is in $\mathbb{N}^*$.

So we can use $\mathbb{N}^*$ to give inductive definitions.

Let $\langle B_1, C_1 \rangle$, $\langle B_2, C_s \rangle$, ..., be an enumeration of all pairs of formulas. Given a set of formulas Δ, define $\Delta(\sigma)$ by induction as follows:

1. $\Delta(\Lambda) = \Delta$
2. $\Delta(\sigma.n) =$

$$\begin{cases} (\Delta(\sigma) \cup \{B_n\})^* & \text{if } \Delta(\sigma) \cup \{B_n\} \nvdash C_n \\ \Delta(\sigma) & \text{otherwise} \end{cases}$$

Here by $(\Delta(\sigma) \cup \{B_n\})^*$ we mean the prime set of formulas which exists by Lemma 9.4 applied to the set $\Delta(\sigma) \cup \{B_n\}$ and the formula C_n. Note that by this definition, if $\Delta(\sigma) \cup \{B_n\} \nvdash C_n$, then $\Delta(\sigma.n) \vdash B_n$ and $\Delta(\sigma.n) \nvdash C_n$. Note also that $\Delta(\sigma) \subseteq \Delta(\sigma.n)$ for any n. If Δ is prime, then $\Delta(\sigma)$ is prime for all σ.

Definition 9.5. Suppose Δ is prime. Then the *canonical model* $M(\Delta)$ for Δ is defined by:

1. $W = \mathbb{N}^*$, the set of finite sequences of natural numbers.
2. R is the partial order according to which $R\sigma\sigma'$ iff σ is an initial segment of σ' (i.e., $\sigma' = \sigma \frown \sigma''$ for some sequence σ'').
3. $V(p) = \{\sigma : p \in \Delta(\sigma)\}$.

It is easy to verify that R is indeed a partial order. Also, the monotonicity condition on V is satisfied. Since $\Delta(\sigma) \subseteq \Delta(\sigma.n)$ we get $\Delta(\sigma) \subseteq \Delta(\sigma')$ whenever $R\sigma\sigma'$ by induction on σ.

9.5 The Truth Lemma

Lemma 9.6. *If Δ is prime, then $M(\Delta), \sigma \Vdash A$ iff $\Delta(\sigma) \vdash A$.*

Proof. By induction on A.

1. $A \equiv \bot$: Since $\Delta(\sigma)$ is prime, it is consistent, so $\Delta(\sigma) \nvdash A$. By definition, $M(\Delta), \sigma \nVdash A$.
2. $A \equiv p$: By definition of $\Vdash$, $M(\Delta), \sigma \Vdash A$ iff $\sigma \in V(p)$, i.e., $\Delta(\sigma) \vdash A$.
3. $A \equiv \neg B$: exercise.
4. $A \equiv B \wedge C$: $M(\Delta), \sigma \Vdash A$ iff $M(\Delta), \sigma \Vdash B$ and $M(\Delta), \sigma \Vdash C$. By induction hypothesis, $M(\Delta), \sigma \Vdash B$ iff $\Delta(\sigma) \vdash B$, and similarly for C. But $\Delta(\sigma) \vdash B$ and $\Delta(\sigma) \vdash C$ iff $\Delta(\sigma) \vdash A$.
5. $A \equiv B \vee C$: $M(\Delta), \sigma \Vdash A$ iff $M(\Delta), \sigma \Vdash B$ or $M(\Delta), \sigma \Vdash C$. By induction hypothesis, this holds iff $\Delta(\sigma) \vdash B$ of $\Delta(\sigma) \vdash C$. We have to show that this in turn holds iff $\Delta(\sigma) \vdash A$. The left-to-right direction is clear. The right-to-left direction follows since $\Delta(\sigma)$ is prime.

6. $A \equiv B \to C$: First the contrapositive of the left-to-right direction: Assume $\Delta(\sigma) \nvdash B \to C$. Then also $\Delta(\sigma) \cup \{B\} \nvdash C$. Since $\langle B, C\rangle$ is $\langle B_n, C_n\rangle$ for some n, we have $\Delta(\sigma.n) = (\Delta(\sigma) \cup \{B\})^*$, and $\Delta(\sigma.n) \vdash B$ but $\Delta(\sigma.n) \nvdash C$. By inductive hypothesis, $\mathfrak{M}(\Delta), \sigma.n \Vdash B$ and $\mathfrak{M}(\Delta), \sigma.n \nVdash C$. Since $R\sigma(\sigma.n)$, this means that $\mathfrak{M}(\Delta), \sigma \nVdash A$.

   Now assume $\Delta(\sigma) \vdash B \to C$, and let $R\sigma\sigma'$. Since $\Delta(\sigma) \subseteq \Delta(\sigma')$, we have: if $\Delta(\sigma') \vdash B$, then $\Delta(\sigma') \vdash C$. In other words, for every σ' such that $R\sigma\sigma'$, either $\Delta(\sigma') \nvdash B$ or $\Delta(\sigma') \vdash C$. By induction hypothesis, this means that whenever $R\sigma\sigma'$, either $\mathfrak{M}(\Delta), \sigma' \nVdash B$ or $\mathfrak{M}(\Delta), \sigma' \Vdash C$, i.e., $\mathfrak{M}(\Delta), \sigma \Vdash A$. □

9.6 The Completeness Theorem

Theorem 9.7. *If $\Gamma \vDash A$ then $\Gamma \vdash A$.*

Proof. We prove the contrapositive: Suppose $\Gamma \nvdash A$. Then by Lemma 9.4, there is a prime set $\Gamma^* \supseteq \Gamma$ such that $\Gamma^* \nvdash A$. Consider the canonical model $\mathfrak{M}(\Gamma^*)$ for Γ^* as defined in Definition 9.5. For any $B \in \Gamma$, $\Gamma^* \vdash B$. Note that $\Gamma^*(\Lambda) = \Gamma^*$. By the Truth Lemma (Lemma 9.6), we have $\mathfrak{M}(\Gamma^*), \Lambda \Vdash B$ for all $B \in \Gamma$ and $\mathfrak{M}(\Gamma^*), \Lambda \nVdash A$. This shows that $\Gamma \nvDash A$. □

Problems

Problem 9.1. Complete the proof of Theorem 9.2. For the cases for ¬Intro and ¬Elim, use the definition of $\mathfrak{M}, w \Vdash \neg A$ in Definition 8.2, i.e., don't treat $\neg A$ as defined by $A \to \bot$.

PART III

Counter-factuals

CHAPTER 10

Introduction

10.1 The Material Conditional

In its simplest form in English, a conditional is a sentence of the form "If ... then ...," where the ... are themselves sentences, such as "If the butler did it, then the gardener is innocent." In introductory logic courses, we earn to symbolize conditionals using the $\rightarrow$ connective: symbolize the parts indicated by ..., e.g., by formulas A and B, and the entire conditional is symbolized by $A \rightarrow B$.

The connective $\rightarrow$ is *truth-functional*, i.e., the truth value—$\mathbb{T}$ or $\mathbb{F}$—of $A \rightarrow B$ is determined by the truth values of A and B: $A \rightarrow B$ is true iff A is false or B is true, and false otherwise. Relative to a truth value assignment v, we define $v \vDash A \rightarrow B$ iff $v \nvDash A$ or $v \vDash B$. The connective $\rightarrow$ with this semantics is called the *material conditional*.

This definition results in a number of elementary logical facts. First of all, the deduction theorem holds for the material conditional:

$$\text{If } \Gamma, A \vDash B \text{ then } \Gamma \vDash A \rightarrow B \tag{10.1}$$

It is truth-functional: $A \rightarrow B$ and $\neg A \vee B$ are equivalent:

$$A \rightarrow B \vDash \neg A \vee B \tag{10.2}$$

$$\neg A \vee B \vDash A \rightarrow B \tag{10.3}$$

A material conditional is entailed by its consequent and by the negation of its antecedent:

$$B \models A \to B \tag{10.4}$$
$$\neg A \models A \to B \tag{10.5}$$

A false material conditional is equivalent to the conjunction of its antecedent and the negation of its consequent: if $A \to B$ is false, $A \wedge \neg B$ is true, and vice versa:

$$\neg(A \to B) \models A \wedge \neg B \tag{10.6}$$
$$A \wedge \neg B \models \neg(A \to B) \tag{10.7}$$

The material conditional supports modus ponens:

$$A, A \to B \models B \tag{10.8}$$

The material conditional agglomerates:

$$A \to B, A \to C \models A \to (B \wedge C) \tag{10.9}$$

We can always strengthen the antecedent, i.e., the conditional is *monotonic*:

$$A \to B \models (A \wedge C) \to B \tag{10.10}$$

The material conditional is transitive, i.e., the chain rule is valid:

$$A \to B, B \to C \models A \to C \tag{10.11}$$

The material conditional is equivalent to its contrapositive:

$$A \to B \models \neg B \to \neg A \tag{10.12}$$
$$\neg B \to \neg A \models A \to B \tag{10.13}$$

These are all useful and unproblematic inferences in mathematical reasoning. However, the philosophical and linguistic literature is replete with purported counterexamples to the equivalent inferences in non-mathematical contexts. These suggest that the material conditional $\rightarrow$ is not—or at least not always—the appropriate connective to use when symbolizing English "if ... then ..." statements.

10.2 Paradoxes of the Material Conditional

One of the first to criticize the use of $A \rightarrow B$ as a way to symbolize "if ... then ..." statements of English was C. I. Lewis. Lewis was criticizing the use of the material condition in Whitehead and Russell's *Principia Mathematica*, who pronounced $\rightarrow$ as "implies." Lewis rightly complained that if $\rightarrow$ meant "implies," then any false proposition p implies that p implies q, since $p \rightarrow (p \rightarrow q)$ is true if p is false, and that any true proposition q implies that p implies q, since $q \rightarrow (p \rightarrow q)$ is true if q is true.

Logicians of course know that *implication*, i.e., logical entailment, is not a connective but a relation between formulas or statements. So we should just not read $\rightarrow$ as "implies" to avoid confusion.[1] As long as we don't, the particular worry that Lewis had simply does not arise: p does not "imply" q even if we think of p as standing for a false English sentence. To determine if $p \vDash q$ we must consider *all* valuations, and $p \nvDash q$ even when we use p to symbolize a sentence which happens to be false.

But there is still something odd about "if ... then..." statements such as Lewis's

> If the moon is made of green cheese, then $2 + 2 = 4$.

and about the inferences

[1] Reading "$\rightarrow$" as "implies" is still widely practised by mathematicians and computer scientists, although philosophers try to avoid the confusions Lewis highlighted by pronouncing it as "only if."

> The moon is not made of green cheese. Therefore, if the moon is made of green cheese, then $2 + 2 = 4$.
>
> $2 + 2 = 4$. Therefore, if the moon is made of green cheese, then $2 + 2 = 4$.

Yet, if "if ... then ..." were just $\rightarrow$, the sentence would be unproblematically true, and the inferences unproblematically valid.

Another example of concerns the tautology $(A \rightarrow B) \vee (B \rightarrow A)$. This would suggest that if you take two indicative sentences S and T from the newspaper at random, the sentence "If S then T, or if T then S" should be true.

10.3 The Strict Conditional

Lewis introduced the *strict conditional* $\dashv$ and argued that it, not the material conditional, corresponds to implication. In alethic modal logic, $A \dashv B$ can be defined as $\Box(A \rightarrow B)$. A strict conditional is thus true (at a world) iff the corresponding material conditional is necessary.

How does the strict conditional fare vis-a-vis the paradoxes of the material conditional? A strict conditional with a false antecedent and one with a true consequent, may be true, or it may be false. Moreover, $(A \dashv B) \vee (B \dashv A)$ is not valid. The strict conditional $A \dashv B$ is also not equivalent to $\neg A \vee B$, so it is not truth functional.

We have:

$$A \dashv B \vDash \neg A \vee B \text{ but:} \tag{10.14}$$
$$\neg A \vee B \nvDash A \dashv B \tag{10.15}$$
$$B \nvDash A \dashv B \tag{10.16}$$
$$\neg A \nvDash A \dashv B \tag{10.17}$$
$$\neg(A \rightarrow B) \nvDash A \wedge \neg B \text{ but:} \tag{10.18}$$
$$A \wedge \neg B \vDash \neg(A \dashv B) \tag{10.19}$$

However, the strict conditional still supports modus ponens:

$$A, A \dashrightarrow B \vDash B \tag{10.20}$$

The strict conditional agglomerates:

$$A \dashrightarrow B, A \dashrightarrow C \vDash A \dashrightarrow (B \wedge C) \tag{10.21}$$

Antecedent strengthening holds for the strict conditional:

$$A \dashrightarrow B \vDash (A \wedge C) \dashrightarrow B \tag{10.22}$$

The strict conditional is also transitive:

$$A \dashrightarrow B, B \dashrightarrow C \vDash A \dashrightarrow C \tag{10.23}$$

Finally, the strict conditional is equivalent to its contrapositive:

$$A \dashrightarrow B \vDash \neg B \dashrightarrow \neg A \tag{10.24}$$

$$\neg B \dashrightarrow \neg A \vDash A \dashrightarrow B \tag{10.25}$$

However, the strict conditional still has its own "paradoxes." Just as a material conditional with a false antecedent or a true consequent is true, a strict conditional with a *necessarily* false antecedent or a necessarily true consequent is true. Moreover, any true strict conditional is necessarily true, and any false strict conditional is necessarily false. In other words, we have

$$\Box A \vDash A \dashrightarrow B \tag{10.26}$$

$$\Box \neg B \vDash A \dashrightarrow B \tag{10.27}$$

$$A \dashrightarrow B \vDash \Box (A \dashrightarrow B) \tag{10.28}$$

$$\neg (A \dashrightarrow B) \vDash \Box \neg (A \dashrightarrow B) \tag{10.29}$$

These are not problems if you think of $\dashrightarrow$ as "implies." Logical entailment relationships are, after all, mathematical facts and so can't be contingent. But they do raise issues if you want to use $\dashrightarrow$ as a logical connective that is supposed to capture "if ... then ...," especially the last two. For surely there are "if ... then ..." statements that are contingently true or contingently false—in fact, they generally are neither necessary nor impossible.

10.4 Counterfactuals

A very common and important form of "if . . . then . . . " constructions in English are built using the past subjunctive form of *to be*: "if it were the case that . . . then it would be the case that . . . " Because usually the antecedent of such a conditional is false, i.e., counter to fact, they are called *counterfactual conditionals* (and because they use the subjunctive form of *to be*, also *subjunctive conditionals*. They are distinguished from *indicative* conditionals which take the form of "if it is the case that . . . then it is the case that . . . " Counterfactual and indicative conditionals differ in truth conditions. Consider Adams's famous example:

> If Oswald didn't kill Kennedy, someone else did.
>
> If Oswald hadn't killed Kennedy, someone else would have.

The first is indicative, the second counterfactual. The first is clearly true: we know JFK was killed by someone, and if that someone wasn't (contrary to the Warren Report) Lee Harvey Oswald, then someone else killed JFK. The second one says something different. It claims that if Oswald hadn't killed Kennedy, i.e., if the Dallas shooting had been avoided or had been unsuccessful, history would have subsequently unfolded in such a way that another assassination would have been successful. In order for it to be true, it would have to be the case that powerful forces had conspired to ensure JFK's death (as many JFK conspiracy theorists believe).

It is a live debate whether the *indicative* conditional is correctly captured by the material conditional, in particular, whether the paradoxes of the material conditional can be "explained" in a way that is compatible with it giving the truth conditions for English indicative conditionals. By contrast, it is uncontroversial that counterfactual conditionals cannot be symbolized correctly by the material conditionals. That is clear because, even though generally the antecedents of counterfactuals are false, not

all counterfactuals with false antecedents are true—for instance, if you believe the Warren Report, and there was no conspiracy to assassinate JFK, then Adams's counterfactual conditional is an example.

Counterfactual conditionals play an important role in causal reasoning: a prime example of the use of counterfactuals is to express causal relationships. E.g., striking a match causes it to light, and you can express this by saying "if this match were struck, it would light." Material, and generally indicative conditionals, cannot be used to express this: "the match is struck $\rightarrow$ the match lights" is true if the match is never struck, regardless of what would happen if it were. Even worse, "the match is struck $\rightarrow$ the match turns into a bouquet of flowers" is also true if it is never struck, but the match would certainly not turn into a bouquet of flowers if it were struck.

It is still debated What exactly the correct logic of counterfactuals is. An influential analysis of counterfactuals was given by Stalnaker and Lewis. According to them, a counterfactual "if it were the case that S then it would be the case that T" is true iff T is true in the counterfactual situation ("possible world") that is closest to the way the actual world is and where S is true. This is called an "ontic" analysis, since it makes reference to an ontology of possible worlds. Other analyses make use of conditional probabilities or theories of belief revision. There is a proliferation of different proposed logics of counterfactuals. There isn't even a single Lewis-Stalnaker logic of counterfactuals: even though Stalnaker and Lewis proposed accounts along similar lines with reference to closest possible worlds, the assumptions they made result in different valid inferences.

Problems

Problem 10.1. Give **S5**-counterexamples to the entailment relations which do not hold for the strict conditional, i.e., for:

1. $\neg p \nvDash \Box(p \rightarrow q)$

2. $q \not\models \Box(p \to q)$
3. $\neg\Box(p \to q) \not\models p \wedge \neg q$
4. $\not\models \Box(p \to q) \vee \Box(q \to p)$

Problem 10.2. Show that the valid entailment relations hold for the strict conditional by giving **S5**-proofs of:

1. $\Box(A \to B) \models \neg A \vee B$
2. $A \wedge \neg B \models \neg\Box(A \to B)$
3. $A, \Box(A \to B) \models B$
4. $\Box(A \to B), \Box(A \to C) \models \Box(A \to (B \wedge C))$
5. $\Box(A \to B) \models \Box((A \wedge C) \to B)$
6. $\Box(A \to B), \Box(B \to C) \models \Box(A \to C)$
7. $\Box(A \to B) \models \Box(\neg B \to \neg A)$
8. $\Box(\neg B \to \neg A) \models \Box(A \to B)$

Problem 10.3. Give proofs in **S5** of:

1. $\Box\neg B \models A \dashv B$
2. $A \dashv B \models \Box(A \dashv B)$
3. $\neg(A \dashv B) \models \Box\neg(A \dashv B)$

Use the definition of $\dashv$ to do so.

CHAPTER 11

Minimal Change Semantics

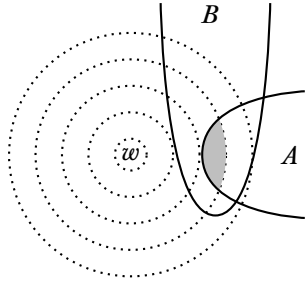
11.1 Introduction

Stalnaker and Lewis proposed accounts of counterfactual conditionals such as "If the match were struck, it would light." Their accounts were proposals for how to properly understand the truth conditions for such sentences. The idea behind both proposals is this: to evaluate whether a counterfactual conditional is true, we have to consider those possible worlds which are minimally different from the way the world actually is to make the antecedent true. If the consequent is true in these possible worlds, then the counterfactual is true. For instance, suppose I hold a match and a matchbook in my hand. In the actual world I only look at them and ponder what would happen if I were to strike the match. The minimal change from the actual world where I strike the match is that where I decide to act and strike the match. It is minimal in that nothing else changes: I don't also jump in the air, striking the match doesn't also light my hair on fire, I don't suddenly lose

all strength in my fingers, I am not simultaneously doused with water in a SuperSoaker ambush, etc. In that alternative possibility, the match lights. Hence, it's true that if I were to strike the match, it would light.

This intuitive account can be paired with formal semantics for logics of counterfactuals. Lewis introduced the symbol "$\Box\!\!\rightarrow$" for the counterfactual while Stalnaker used the symbol ">". We'll use $\Box\!\!\rightarrow$, and add it as a binary connective to propositional logic. So, we have, in addition to formulas of the form $A \rightarrow B$ also formulas of the form $A \mathrel{\Box\!\!\rightarrow} B$. The formal semantics, like the relational semantics for modal logic, is based on models in which formulas are evaluated at worlds, and the satisfaction condition defining $\boldsymbol{M}, w \Vdash A \mathrel{\Box\!\!\rightarrow} B$ is given in terms of $\boldsymbol{M}, w' \Vdash A$ and $\boldsymbol{M}, w' \Vdash B$ for some (other) worlds w'. Which w'? Intuitively, the one(s) closest to w for which it holds that $\boldsymbol{M}, w' \Vdash A$. This requires that a relation of "closeness" has to be included in the model as well.

Lewis introduced an instructive way of representing counterfactual situations graphically. Each possible world is at the center of a set of nested spheres containing other worlds—we draw these spheres as concentric circles. The worlds between two spheres are equally close to the world at the center as each other, those contained in a nested sphere are closer, and those in a surrounding sphere further away.



The closest A-worlds are those worlds w' where A is satisfied which lie in the smallest sphere around the center world w (the

gray area). Intuitively, $A \boxrightarrow B$ is satisfied at w if B is true at all closest A-worlds.

11.2 Sphere Models

One way of providing a formal semantics for counterfactuals is to turn Lewis's informal account into a mathematical structure. The spheres around a world w then are sets of worlds. Since the spheres are nested, the sets of worlds around w have to be linearly ordered by the subset relation.

Definition 11.1. A *sphere model* is a triple $M = \langle W, O, V \rangle$ where W is a non-empty set of worlds, $V\colon \mathrm{At}_0 \to \wp(W)$ is a valuation, and $O\colon W \to \wp(\wp(W))$ assigns to each world w a *system of spheres* O_w. For each w, O_w is a set of sets of worlds, and must satisfy:

1. O_w is *centered* on w: $\{w\} \in O_w$.
2. O_w is *nested*: whenever $S_1, S_2 \in O_w$, $S_1 \subseteq S_2$ or $S_2 \subseteq S_1$, i.e., O_w is linearly ordered by $\subseteq$.
3. O_w is closed under non-empty unions.
4. O_w is closed under non-empty intersections.

The intuition behind O_w is that the worlds "around" w are stratified according to how far away they are from w. The innermost sphere is just w by itself, i.e., the set $\{w\}$: w is closer to w than the worlds in any other sphere. If $S \subsetneq S'$, then the worlds in $S' \setminus S$ are further way from w than the worlds in S: $S' \setminus S$ is the "layer" between the S and the worlds outside of S'. In particular, we have to think of the spheres as containing all the worlds within their outer surface; they are not just the individual layers.

The diagram in Figure 11.1 corresponds to the sphere model with $W = \{w, w_1, \ldots, w_7\}$, $V(p) = \{w_5, w_6, w_7\}$. The innermost sphere $S_1 = \{w\}$. The closest worlds to w are w_1, w_2, w_3, so the

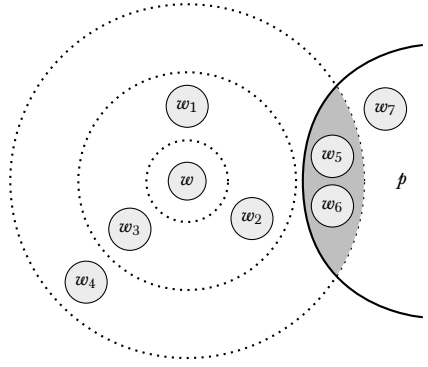


Figure 11.1: Diagram of a sphere model

next larger sphere is $S_2 = \{w, w_1, w_2, w_3\}$. The worlds further out are w_4, w_5, w_6, so the outermost sphere is $S_3 = \{w, w_1, \ldots, w_6\}$. The system of spheres around w is $O_w = \{S_1, S_2, S_3\}$. The world w_7 is not in any sphere around w. The closest worlds in which p is true are w_5 and w_6, and so the smallest p-admitting sphere is S_3.

To define satisfaction of a formula A at world w in a sphere model $M, M, w \Vdash A$, we expand the definition for modal formulas to include a clause for $B \boxdot\!\!\rightarrow C$:

Definition 11.2. $M, w \Vdash B \boxdot\!\!\rightarrow C$ iff either

1. For all $u \in \bigcup O_w$, $M, u \nVdash C$, or
2. For some $S \in O_w$,
   a) $M, u \Vdash B$ for some $u \in S$, and
   b) for all $v \in S$, either $M, v \nVdash B$ or $M, v \Vdash C$.

According to this definition, $M, w \Vdash B \boxdot\!\!\rightarrow C$ iff either the antecedent B is false everywhere in the spheres around w, or there is a sphere S where B is true, and the material conditional $B \rightarrow C$ is true at all worlds in that "B-admitting" sphere. Note

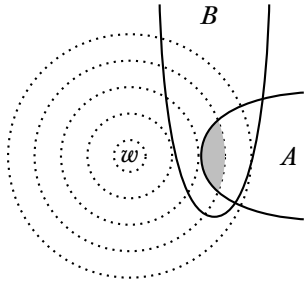


Figure 11.2: Non-vacuously true counterfactual

that we didn't require in the definition that S is the *innermost* B-admitting sphere, contrary to what one might expect from the intuitive explanation. But if the condition in (2) is satisfied for some sphere S, then it is also satisfied for all spheres S contains, and hence in particular for the innermost sphere.

Note also that the definition of sphere models does not require that there *is* an innermost B-admitting sphere: we may have an infinite sequence $S_1 \supsetneq S_2 \supsetneq \cdots \supsetneq \{w\}$ of B-admitting spheres, and hence no innermost B-admitting spheres. In that case, $M, w \Vdash B \boxright C$ iff $B \rightarrow C$ holds throughout the spheres S_i, S_{i+1}, ..., for some i.

11.3 Truth and Falsity of Counterfactuals

A counterfactual $A \boxright B$ is (non-vacuously) true if the closest A-worlds are all B-worlds, as depicted in Figure 11.2. A counterfactual is also true at w if the system of spheres around w has no A-admitting spheres at all. In that case it is *vacuously* true (see Figure 11.3).

It can be false in two ways. One way is if the closest A-worlds are not all B-worlds, but some of them are. In this case, $A \boxright \neg B$ is also false (see Figure 11.4). If the closest A-worlds do not overlap with the B-worlds at all, then $A \boxright B$. But, in this case

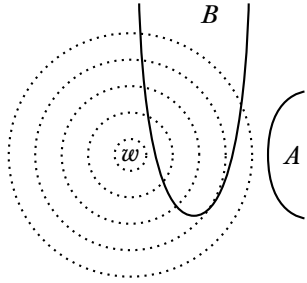


Figure 11.3: Vacuously true counterfactual

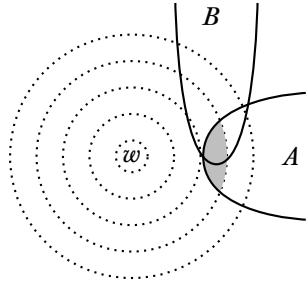


Figure 11.4: False counterfactual, false opposite

all the closest A-worlds are $\neg B$-worlds, and so $A \boxrightarrow \neg B$ is true (see Figure 11.5).

In contrast to the strict conditional, counterfactuals may be contingent. Consider the sphere model in Figure 11.6. The A-worlds closest to u are all B-worlds, so $M, u \Vdash A \boxrightarrow B$. But there are A-worlds closest to v which are not B-worlds, so $M, v \nVdash A \boxrightarrow B$.

11.4 Antecedent Strengthenng

"Strengthening the antecedent" refers to the inference $A \rightarrow C \vDash (A \wedge B) \rightarrow C$. It is valid for the material conditional, but invalid

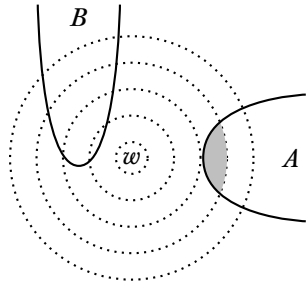


Figure 11.5: False counterfactual, true opposite

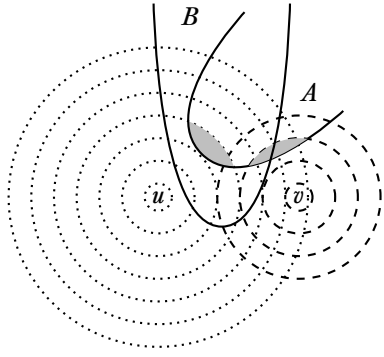


Figure 11.6: Contingent counterfactual

for counterfactuals. Suppose it is true that if I were to strike this match, it would light. (That means, there is nothing wrong with the match or the matchbook surface, I will not break the match, etc.) But it is not true that if I were to light this match in outer space, it would light. So the following inference is invalid:

> I the match were struck, it would light.
>
> Therefore, if the match were struck in outer space, it would light.

The Lewis-Stalnaker account of conditionals explains this:

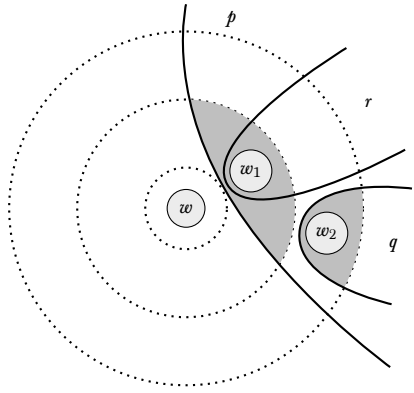


Figure 11.7: Counterexample to antecedent strengthening

the closest world where I light the match and I do so in outer space is much further removed from the actual world than the closest world where I light the match is. So although it's true that the match lights in the latter, it is not in the former. And that is as it schould be.

Example 11.3. The sphere semantics invalidates the inference, i.e., we have $p \boxright r \not\models (p \wedge q) \boxright r$. Consider the model $\mathfrak{M} = \langle W, O, V \rangle$ where $W = \{w, w_1, w_2\}$, $O_w = \{\{w\}, \{w, w_1\}, \{w, w_1, w_2\}\}$, $V(p) = \{w_1, w_2\}$, $V(q) = \{w_2\}$, and $V(r) = \{w_1\}$. There is a p-admitting sphere $S = \{w, w_1\}$ and $p \rightarrow r$ is true at all worlds in it, so $\mathfrak{M}, w \Vdash p \boxright r$. There is also a $(p \wedge q)$-admitting sphere $S' = \{w, w_1, w_2\}$ but $\mathfrak{M}, w_2 \nVdash (p \wedge q) \rightarrow r$, so $\mathfrak{M}, w \nVdash (p \wedge q) \boxright r$ (see Figure 11.7).

11.5 Transitivity

For the material conditional, the chain rule holds: $A \rightarrow B, B \rightarrow C \models A \rightarrow C$. In other words, the material conditional is transitive. Is the same true for counterfactuals? Consider the following example due to Stalnaker.

> If J. Edgar Hoover had been born a Russian, he would have been a Communist.
>
> If J. Edgar Hoover were a Communist, he would have been be a traitor.
>
> Therefore, If J. Edgar Hoover had been born a Russian, he would have been be a traitor.

If Hoover had been born (at the same time he actually did), not in the United States, but in Russia, he would have grown up in the Soviet Union and become a Communist (let's assume). So the first premise is true. Likewise, the second premise, considered in isolation is true. The conclusion, however, is false: in all likelihood, Hoover would have been a fervent Communist if he had been born in the USSR, and not been a traitor (to his country). The intuitive assignment of truth values is borne out by the Stalnaker-Lewis account. The closest possible world to ours with the only change being Hoover's place of birth is the one where Hoover grows up to be a good citizen of the USSR. This is the closest possible world where the antecedent of the first premise and of the conclusion is true, and in that world Hoover is a loyal member of the Communist party, and so not a traitor. To evaluate the second premise, we have to look at a different world, however: the closest world where Hoover is a Communist, which is one where he was born in the United States, turned, and thus became a traitor.[1]

Example 11.4. The sphere semantics invalidates the inference, i.e., we have $p \boxrightarrow q, q \boxrightarrow r \not\models p \boxrightarrow r$. Consider the model $M = \langle W, O, V\rangle$ where $W = \{w, w_1, w_2\}$, $O_w = \{\{w\}, \{w, w_1\}, \{w, w_1, w_2\}\}$, $V(p) = \{w_2\}$, $V(q) = \{w_1, w_2\}$, and $V(r) = \{w_1\}$. There is a p-admitting sphere $S = \{w, w_1, w_2\}$ and $q \rightarrow q$ is true at all worlds in it, so $M, w \Vdash p \boxrightarrow q$. There is also

[1] Of course, to appreciate the force of the example we have to take on board some metaphysical and political assumptions, e.g., that it is possible that Hoover could have been born to Russian parents, or that Communists in the US of the 1950s were traitors to their country.

a q-admitting sphere $S' = \{w, w_1\}$ and $M \nVdash q \to r$ is true at all worlds in it, so $M, w \Vdash q \boxrightarrow r$. However, the p-admitting sphere $\{w, w_1, w_2\}$ contains a world, namely w_2, where $M, w_2 \nVdash p \to r$.

11.6 Contraposition

Material and strict conditionals are equivalent to their contrapositives. Counterfactuals are not. Here is an example due to Kratzer:

> If Goethe hadn't died in 1832, he would (still) be dead now.
>
> If Goethe weren't dead now, he would have died in 1832.

The first sentence is true: humans don't live hundreds of years. The second is clearly false: if Goethe weren't dead now, he would be still alive, and so couldn't have died in 1832.

Example 11.5. The sphere semantics invalidates contraposition, i.e., we have $p \boxrightarrow q \not\models \neg q \boxrightarrow \neg p$. Think of p as "Goethe didn't die in 1832" and q as "Goethe is dead now." We can capture this in a model $M_1 = \langle W, O, V \rangle$ with $W = \{w, w_1, w_2\}$, $O = \{\{w\}, \{w, w_1\}, \{w, w_1, w_2\}\}$, $V(p) = \{w_1, w_2\}$ and $V(q) = \{w, w_1\}$. So w is the actual world where Goethe died in 1832 and is still dead; w_1 is the (close) world where Goethe died in, say, 1833, and is still dead; and w_2 is a (remote) world where Goethe is still alive. There is a p-admitting sphere $S = \{w, w_1\}$ and $p \to q$ is true at all worlds in it, so $M, w \Vdash p \boxrightarrow q$. However, the $\neg q$-admitting sphere $\{w, w_1, w_2\}$ contains a world, namely w_2, where q is false and p is true, so $M, w_2 \nVdash \neg q \to \neg p$.

Problems

Problem 11.1. Find a convincing, intuitive example for the failure of transitivity of counterfactuals.

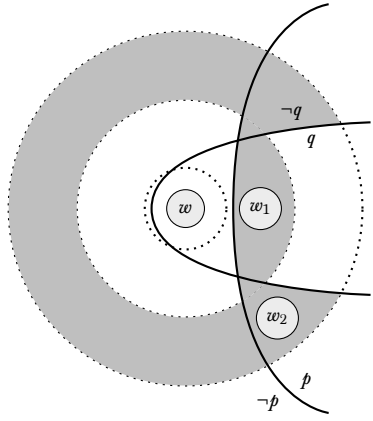


Figure 11.8: Counterexample to contraposition

Problem 11.2. Draw the sphere diagram corresponding to the counterexample in Example 11.4.

Problem 11.3. In Example 11.4, world w_2 is where Hoover is born in Russia, is a communist, and not a traitor, and w_1 is the world where Hoover is born in the US, is a communist, and a traitor. In this model, w_1 is closer to w than w_2 is. Is this necessary? Can you give a counterexample that does not assume that Hoover's being born in Russia is a more remote possibility than him being a Communist?

PART IV

Appendices

APPENDIX A

Sets

A.1 Extensionality

A *set* is a collection of objects, considered as a single object. The objects making up the set are called *elements* or *members* of the set. If x is an element of a set a, we write $x \in a$; if not, we write $x \notin a$. The set which has no elements is called the *empty* set and denoted "$\emptyset$".

It does not matter how we *specify* the set, or how we *order* its elements, or indeed how *many times* we count its elements. All that matters are what its elements are. We codify this in the following principle.

Definition A.1 (Extensionality). If A and B are sets, then $A = B$ iff every element of A is also an element of B, and vice versa.

Extensionality licenses some notation. In general, when we have some objects $a_1, \ldots, a_n$, then $\{a_1, \ldots, a_n\}$ is *the* set whose elements are $a_1, \ldots, a_n$. We emphasise the word "*the*", since extensionality tells us that there can be only *one* such set. Indeed, extensionality also licenses the following:

$$\{a, a, b\} = \{a, b\} = \{b, a\}.$$

This delivers on the point that, when we consider sets, we don't care about the order of their elements, or how many times they are specified.

Example A.2. Whenever you have a bunch of objects, you can collect them together in a set. The set of Richard's siblings, for instance, is a set that contains one person, and we could write it as $S = \{\text{Ruth}\}$. The set of positive integers less than 4 is $\{1, 2, 3\}$, but it can also be written as $\{3, 2, 1\}$ or even as $\{1, 2, 1, 2, 3\}$. These are all the same set, by extensionality. For every element of $\{1, 2, 3\}$ is also an element of $\{3, 2, 1\}$ (and of $\{1, 2, 1, 2, 3\}$), and vice versa.

Frequently we'll specify a set by some property that its elements share. We'll use the following shorthand notation for that: $\{x : \varphi(x)\}$, where the $\varphi(x)$ stands for the property that x has to have in order to be counted among the elements of the set.

Example A.3. In our example, we could have specified S also as

$$S = \{x : x \text{ is a sibling of Richard}\}.$$

Example A.4. A number is called *perfect* iff it is equal to the sum of its proper divisors (i.e., numbers that evenly divide it but aren't identical to the number). For instance, 6 is perfect because its proper divisors are 1, 2, and 3, and $6 = 1 + 2 + 3$. In fact, 6

is the only positive integer less than 10 that is perfect. So, using extensionality, we can say:

$$\{6\} = \{x : x \text{ is perfect and } 0 \leq x \leq 10\}$$

We read the notation on the right as "the set of x's such that x is perfect and $0 \leq x \leq 10$". The identity here confirms that, when we consider sets, we don't care about how they are specified. And, more generally, extensionality guarantees that there is always only one set of x's such that $\varphi(x)$. So, extensionality justifies calling $\{x : \varphi(x)\}$ *the* set of x's such that $\varphi(x)$.

Extensionality gives us a way for showing that sets are identical: to show that $A = B$, show that whenever $x \in A$ then also $x \in B$, and whenever $y \in B$ then also $y \in A$.

A.2 Subsets and Power Sets

We will often want to compare sets. And one obvious kind of comparison one might make is as follows: *everything in one set is in the other too.* This situation is sufficiently important for us to introduce some new notation.

Definition A.5 (Subset). If every element of a set A is also an element of B, then we say that A is a *subset* of B, and write $A \subseteq B$. If A is not a subset of B we write $A \nsubseteq B$. If $A \subseteq B$ but $A \neq B$, we write $A \subsetneq B$ and say that A is a *proper subset* of B.

Example A.6. Every set is a subset of itself, and $\emptyset$ is a subset of every set. The set of even numbers is a subset of the set of natural numbers. Also, $\{a, b\} \subseteq \{a, b, c\}$. But $\{a, b, e\}$ is not a subset of $\{a, b, c\}$.

Example A.7. The number 2 is an element of the set of integers, whereas the set of even numbers is a subset of the set of integers. However, a set may happen to *both* be an element and a subset of some other set, e.g., $\{0\} \in \{0, \{0\}\}$ and also $\{0\} \subseteq \{0, \{0\}\}$.

Extensionality gives a criterion of identity for sets: $A = B$ iff every element of A is also an element of B and vice versa. The definition of "subset" defines $A \subseteq B$ precisely as the first half of this criterion: every element of A is also an element of B. Of course the definition also applies if we switch A and B: that is, $B \subseteq A$ iff every element of B is also an element of A. And that, in turn, is exactly the "vice versa" part of extensionality. In other words, extensionality entails that sets are equal iff they are subsets of one another.

Proposition A.8. *$A = B$ iff both $A \subseteq B$ and $B \subseteq A$.*

Now is also a good opportunity to introduce some further bits of helpful notation. In defining when A is a subset of B we said that "every element of A is ...," and filled the "..." with "an element of B". But this is such a common *shape* of expression that it will be helpful to introduce some formal notation for it.

Definition A.9. $(\forall x \in A)\varphi$ abbreviates $\forall x(x \in A \rightarrow \varphi)$. Similarly, $(\exists x \in A)\varphi$ abbreviates $\exists x(x \in A \wedge \varphi)$.

Using this notation, we can say that $A \subseteq B$ iff $(\forall x \in A)x \in B$.

Now we move on to considering a certain kind of set: the set of all subsets of a given set.

Definition A.10 (Power Set). The set consisting of all subsets of a set A is called the *power set of* A, written $\wp(A)$.

$$\wp(A) = \{B : B \subseteq A\}$$

Example A.11. What are all the possible subsets of $\{a, b, c\}$? They are: $\emptyset$, $\{a\}$, $\{b\}$, $\{c\}$, $\{a, b\}$, $\{a, c\}$, $\{b, c\}$, $\{a, b, c\}$. The set of all these subsets is $\wp(\{a, b, c\})$:

$$\wp(\{a, b, c\}) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$$

A.3 Some Important Sets

Example A.12. We will mostly be dealing with sets whose elements are mathematical objects. Four such sets are important enough to have specific names:

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\}$$
the set of natural numbers

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$
the set of integers

$$\mathbb{Q} = \{m/n : m, n \in \mathbb{Z} \text{ and } n \neq 0\}$$
the set of rationals

$$\mathbb{R} = (-\infty, \infty)$$
the set of real numbers (the continuum)

These are all *infinite* sets, that is, they each have infinitely many elements.

As we move through these sets, we are adding *more* numbers to our stock. Indeed, it should be clear that $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$: after all, every natural number is an integer; every integer is a rational; and every rational is a real. Equally, it should be clear that $\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q}$, since -1 is an integer but not a natural number, and $1/2$ is rational but not integer. It is less obvious that $\mathbb{Q} \subsetneq \mathbb{R}$, i.e., that there are some real numbers which are not rational.

We'll sometimes also use the set of positive integers $\mathbb{Z}^+ = \{1, 2, 3, \ldots\}$ and the set containing just the first two natural numbers $\mathbb{B} = \{0, 1\}$.

Example A.13 (Strings). Another interesting example is the set A^* of *finite strings* over an alphabet A: any finite sequence of elements of A is a string over A. We include the *empty string* Λ among the strings over A, for every alphabet A. For instance,

$$\mathbb{B}^* = \{\Lambda, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, 0000, \ldots\}.$$

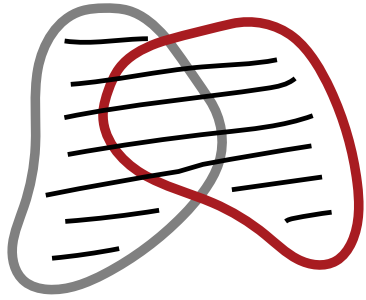


Figure A.1: The union $A \cup B$ of two sets is set of elements of A together with those of B.

If $x = x_1 \dots x_n \in A^*$is a string consisting of n "letters" from A, then we say *length* of the string is n and write $\text{len}(x) = n$.

Example A.14 (Infinite sequences). For any set A we may also consider the set A^ω of infinite sequences of elements of A. An infinite sequence $a_1 a_2 a_3 a_4 \dots$ consists of a one-way infinite list of objects, each one of which is an element of A.

A.4 Unions and Intersections

In appendix A.1, we introduced definitions of sets by abstraction, i.e., definitions of the form $\{x : \varphi(x)\}$. Here, we invoke some property φ, and this property can mention sets we've already defined. So for instance, if A and B are sets, the set $\{x : x \in A \vee x \in B\}$ consists of all those objects which are elements of either A or B, i.e., it's the set that combines the elements of A and B. We can visualize this as in Figure A.1, where the highlighted area indicates the elements of the two sets A and B together.

This operation on sets—combining them—is very useful and common, and so we give it a formal name and a symbol.

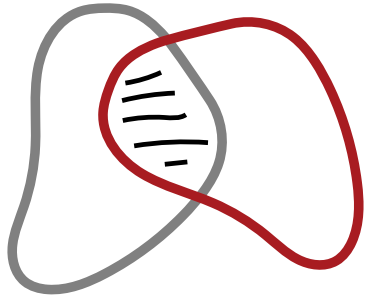


Figure A.2: The intersection $A \cap B$ of two sets is the set of elements they have in common.

Definition A.15 (Union). The *union* of two sets A and B, written $A \cup B$, is the set of all things which are elements of A, B, or both.

$$A \cup B = \{x : x \in A \vee x \in B\}$$

Example A.16. Since the multiplicity of elements doesn't matter, the union of two sets which have an element in common contains that element only once, e.g., $\{a, b, c\} \cup \{a, 0, 1\} = \{a, b, c, 0, 1\}$.

The union of a set and one of its subsets is just the bigger set: $\{a, b, c\} \cup \{a\} = \{a, b, c\}$.

The union of a set with the empty set is identical to the set: $\{a, b, c\} \cup \emptyset = \{a, b, c\}$.

We can also consider a "dual" operation to union. This is the operation that forms the set of all elements that are elements of A and are also elements of B. This operation is called *intersection*, and can be depicted as in Figure A.2.

Definition A.17 (Intersection). The *intersection* of two sets A and B, written $A \cap B$, is the set of all things which are elements of both A and B.

$$A \cap B = \{x : x \in A \wedge x \in B\}$$

Two sets are called *disjoint* if their intersection is empty. This means they have no elements in common.

Example A.18. If two sets have no elements in common, their intersection is empty: $\{a, b, c\} \cap \{0, 1\} = \emptyset$.

If two sets do have elements in common, their intersection is the set of all those: $\{a, b, c\} \cap \{a, b, d\} = \{a, b\}$.

The intersection of a set with one of its subsets is just the smaller set: $\{a, b, c\} \cap \{a, b\} = \{a, b\}$.

The intersection of any set with the empty set is empty: $\{a, b, c\} \cap \emptyset = \emptyset$.

We can also form the union or intersection of more than two sets. An elegant way of dealing with this in general is the following: suppose you collect all the sets you want to form the union (or intersection) of into a single set. Then we can define the union of all our original sets as the set of all objects which belong to at least one element of the set, and the intersection as the set of all objects which belong to every element of the set.

Definition A.19. If A is a set of sets, then $\bigcup A$ is the set of elements of elements of A:

$$\begin{aligned}\bigcup A &= \{x : x \text{ belongs to an element of } A\}, \text{ i.e.,} \\ &= \{x : \text{there is a } B \in A \text{ so that } x \in B\}\end{aligned}$$

Definition A.20. If A is a set of sets, then $\bigcap A$ is the set of objects which all elements of A have in common:

$$\begin{aligned}\bigcap A &= \{x : x \text{ belongs to every element of } A\}, \text{ i.e.,} \\ &= \{x : \text{for all } B \in A, x \in B\}\end{aligned}$$

Example A.21. Suppose $A = \{\{a, b\}, \{a, d, e\}, \{a, d\}\}$. Then $\bigcup A = \{a, b, d, e\}$ and $\bigcap A = \{a\}$.

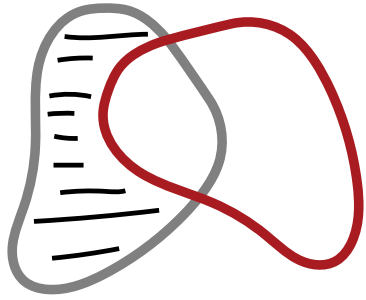


Figure A.3: The difference $A \setminus B$ of two sets is the set of those elements of A which are not also elements of B.

We could also do the same for a sequence of sets $A_1, A_2, \ldots$

$$\bigcup_i A_i = \{x : x \text{ belongs to one of the } A_i\}$$
$$\bigcap_i A_i = \{x : x \text{ belongs to every } A_i\}.$$

When we have an *index* of sets, i.e., some set I such that we are considering A_i for each $i \in I$, we may also use these abbreviations:

$$\bigcup_{i \in I} A_i = \bigcup \{A_i : i \in I\}$$
$$\bigcap_{i \in I} A_i = \bigcap \{A_i : i \in I\}$$

Finally, we may want to think about the set of all elements in A which are not in B. We can depict this as in Figure A.3.

Definition A.22 (Difference). The *set difference* $A \setminus B$ is the set of all elements of A which are not also elements of B, i.e.,

$$A \setminus B = \{x : x \in A \text{ and } x \notin B\}.$$

A.5 Pairs, Tuples, Cartesian Products

It follows from extensionality that sets have no order to their elements. So if we want to represent order, we use *ordered pairs* $\langle x, y\rangle$. In an unordered pair $\{x, y\}$, the order does not matter: $\{x, y\} = \{y, x\}$. In an ordered pair, it does: if $x \neq y$, then $\langle x, y\rangle \neq \langle y, x\rangle$.

How should we think about ordered pairs in set theory? Crucially, we want to preserve the idea that ordered pairs are identical iff they share the same first element and share the same second element, i.e.:

$$\langle a, b\rangle = \langle c, d\rangle \text{ iff both } a = c \text{ and } b = d.$$

We can define ordered pairs in set theory using the Wiener-Kuratowski definition.

Definition A.23 (Ordered pair). $\langle a, b\rangle = \{\{a\}, \{a, b\}\}$.

Having fixed a definition of an ordered pair, we can use it to define further sets. For example, sometimes we also want ordered sequences of more than two objects, e.g., *triples* $\langle x, y, z\rangle$, *quadruples* $\langle x, y, z, u\rangle$, and so on. We can think of triples as special ordered pairs, where the first element is itself an ordered pair: $\langle x, y, z\rangle$ is $\langle\langle x, y\rangle, z\rangle$. The same is true for quadruples: $\langle x, y, z, u\rangle$ is $\langle\langle\langle x, y\rangle, z\rangle, u\rangle$, and so on. In general, we talk of *ordered n-tuples* $\langle x_1, \ldots, x_n\rangle$.

Certain sets of ordered pairs, or other ordered n-tuples, will be useful.

Definition A.24 (Cartesian product). Given sets A and B, their *Cartesian product* $A \times B$ is defined by

$$A \times B = \{\langle x, y\rangle : x \in A \text{ and } y \in B\}.$$

Example A.25. If $A = \{0, 1\}$, and $B = \{1, a, b\}$, then their product is

$$A \times B = \{\langle 0,1\rangle, \langle 0,a\rangle, \langle 0,b\rangle, \langle 1,1\rangle, \langle 1,a\rangle, \langle 1,b\rangle\}.$$

Example A.26. If A is a set, the product of A with itself, $A \times A$, is also written A^2. It is the set of *all* pairs $\langle x, y\rangle$ with $x, y \in A$. The set of all triples $\langle x, y, z\rangle$ is A^3, and so on. We can give a recursive definition:

$$A^1 = A$$
$$A^{k+1} = A^k \times A$$

Proposition A.27. *If A has n elements and B has m elements, then $A \times B$ has $n \cdot m$ elements.*

Proof. For every element x in A, there are m elements of the form $\langle x, y\rangle \in A \times B$. Let $B_x = \{\langle x, y\rangle : y \in B\}$. Since whenever $x_1 \neq x_2$, $\langle x_1, y\rangle \neq \langle x_2, y\rangle$, $B_{x_1} \cap B_{x_2} = \emptyset$. But if $A = \{x_1, \ldots, x_n\}$, then $A \times B = B_{x_1} \cup \cdots \cup B_{x_n}$, and so has $n \cdot m$ elements.

To visualize this, arrange the elements of $A \times B$ in a grid:

$$\begin{array}{rccccc} B_{x_1} = & \{\langle x_1, y_1\rangle & \langle x_1, y_2\rangle & \ldots & \langle x_1, y_m\rangle\} \\ B_{x_2} = & \{\langle x_2, y_1\rangle & \langle x_2, y_2\rangle & \ldots & \langle x_2, y_m\rangle\} \\ \vdots & & \vdots & & \\ B_{x_n} = & \{\langle x_n, y_1\rangle & \langle x_n, y_2\rangle & \ldots & \langle x_n, y_m\rangle\} \end{array}$$

Since the x_i are all different, and the y_j are all different, no two of the pairs in this grid are the same, and there are $n \cdot m$ of them.□

Example A.28. If A is a set, a *word* over A is any sequence of elements of A. A sequence can be thought of as an n-tuple of elements of A. For instance, if $A = \{a, b, c\}$, then the sequence "bac" can be thought of as the triple $\langle b, a, c\rangle$. Words, i.e., sequences of symbols, are of crucial importance in computer science. By convention, we count elements of A as sequences of length 1, and $\emptyset$ as the sequence of length 0. The set of *all* words over A then is

$$A^* = \{\emptyset\} \cup A \cup A^2 \cup A^3 \cup \dots$$

A.6 Russell's Paradox

Extensionality licenses the notation $\{x : \varphi(x)\}$, for *the* set of x's such that $\varphi(x)$. However, all that extensionality *really* licenses is the following thought. *If* there is a set whose members are all and only the φ's, *then* there is only one such set. Otherwise put: having fixed some φ, the set $\{x : \varphi(x)\}$ is unique, *if it exists.*

But this conditional is important! Crucially, not every property lends itself to *comprehension.* That is, some properties do *not* define sets. If they all did, then we would run into outright contradictions. The most famous example of this is Russell's Paradox.

Sets may be elements of other sets—for instance, the power set of a set A is made up of sets. And so it makes sense to ask or investigate whether a set is an element of another set. Can a set be a member of itself? Nothing about the idea of a set seems to rule this out. For instance, if *all* sets form a collection of objects, one might think that they can be collected into a single set—the set of all sets. And it, being a set, would be an element of the set of all sets.

Russell's Paradox arises when we consider the property of not having itself as an element, of being *non-self-membered.* What if we suppose that there is a set of all sets that do not have themselves as an element? Does

$$R = \{x : x \notin x\}$$

exist? It turns out that we can prove that it does not.

Theorem A.29 (Russell's Paradox). *There is no set* $R = \{x : x \notin x\}$.

Proof. For reductio, suppose that $R = \{x : x \notin x\}$ exists. Then $R \in R$ iff $R \notin R$, since sets are extensional. But this is a contradicion. □

Let's run through the proof that no set R of non-self-membered sets can exist more slowly. If R exists, it makes sense to ask if $R \in R$ or not—it must be either $\in R$ or $\notin R$. Suppose the former is true, i.e., $R \in R$. R was defined as the set of all sets that are not elements of themselves, and so if $R \in R$, then R does not have this defining property of R. But only sets that have this property are in R, hence, R cannot be an element of R, i.e., $R \notin R$. But R can't both be and not be an element of R, so we have a contradiction.

Since the assumption that $R \in R$ leads to a contradiction, we have $R \notin R$. But this also leads to a contradiction! For if $R \notin R$, it does have the defining property of R, and so would be an element of R just like all the other non-self-membered sets. And again, it can't both not be and be an element of R.

How do we set up a set theory which avoids falling into Russell's Paradox, i.e., which avoids making the *inconsistent* claim that $R = \{x : x \notin x\}$ exists? Well, we would need to lay down axioms which give us very precise conditions for stating when sets exist (and when they don't).

The set theory sketched in this chapter doesn't do this. It's *genuinely naïve*. It tells you only that sets obey extensionality and that, if you have some sets, you can form their union, intersection, etc. It is possible to develop set theory more rigorously than this.

Problems

Problem A.1. Prove that there is at most one empty set, i.e., show that if A and B are sets without elements, then $A = B$.

Problem A.2. List all subsets of $\{a, b, c, d\}$.

Problem A.3. Show that if A has n elements, then $\wp(A)$ has 2^n elements.

Problem A.4. Prove that if $A \subseteq B$, then $A \cup B = B$.

Problem A.5. Prove rigorously that if $A \subseteq B$, then $A \cap B = A$.

Problem A.6. Show that if A is a set and $A \in B$, then $A \subseteq \bigcup B$.

Problem A.7. Prove that if $A \subsetneq B$, then $B \setminus A \neq \emptyset$.

Problem A.8. Using Definition A.23, prove that $\langle a, b\rangle = \langle c, d\rangle$ iff both $a = c$ and $b = d$.

Problem A.9. List all elements of $\{1, 2, 3\}^3$.

Problem A.10. Show, by induction on k, that for all $k \geq 1$, if A has n elements, then A^k has n^k elements.

APPENDIX B

Relations

B.1 Relations as Sets

In appendix A.3, we mentioned some important sets: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$. You will no doubt remember some interesting relations between the elements of some of these sets. For instance, each of these sets has a completely standard *order relation* on it. There is also the relation *is identical with* that every object bears to itself and to no other thing. There are many more interesting relations that we'll encounter, and even more possible relations. Before we review them, though, we will start by pointing out that we can look at relations as a special sort of set.

For this, recall two things from appendix A.5. First, recall the notion of a *ordered pair*: given a and b, we can form $\langle a, b\rangle$. Importantly, the order of elements *does* matter here. So if $a \neq b$ then $\langle a, b\rangle \neq \langle b, a\rangle$. (Contrast this with unordered pairs, i.e., 2-element sets, where $\{a, b\} = \{b, a\}$.) Second, recall the notion of a *Cartesian product*: if A and B are sets, then we can form $A \times B$, the set of all pairs $\langle x, y\rangle$ with $x \in A$ and $y \in B$. In particular, $A^2 = A \times A$ is the set of all ordered pairs from A.

Now we will consider a particular relation on a set: the $<$-relation on the set $\mathbb{N}$ of natural numbers. Consider the set of all pairs of numbers $\langle n, m\rangle$ where $n < m$, i.e.,

$$R = \{\langle n, m\rangle : n, m \in \mathbb{N} \text{ and } n < m\}.$$

There is a close connection between n being less than m, and the pair $\langle n, m\rangle$ being a member of R, namely:

$$n < m \text{ iff } \langle n, m\rangle \in R.$$

Indeed, without any loss of information, we can consider the set R to *be* the $<$-relation on $\mathbb{N}$.

In the same way we can construct a subset of $\mathbb{N}^2$ for any relation between numbers. Conversely, given any set of pairs of numbers $S \subseteq \mathbb{N}^2$, there is a corresponding relation between numbers, namely, the relationship n bears to m if and only if $\langle n, m\rangle \in S$. This justifies the following definition:

Definition B.1 (Binary relation). A *binary relation* on a set A is a subset of A^2. If $R \subseteq A^2$ is a binary relation on A and $x, y \in A$, we sometimes write Rxy (or xRy) for $\langle x, y\rangle \in R$.

Example B.2. The set $\mathbb{N}^2$ of pairs of natural numbers can be listed in a 2-dimensional matrix like this:

$$\begin{array}{ccccc}
\langle \mathbf{0,0}\rangle & \langle 0,1\rangle & \langle 0,2\rangle & \langle 0,3\rangle & \dots \\
\langle 1,0\rangle & \langle \mathbf{1,1}\rangle & \langle 1,2\rangle & \langle 1,3\rangle & \dots \\
\langle 2,0\rangle & \langle 2,1\rangle & \langle \mathbf{2,2}\rangle & \langle 2,3\rangle & \dots \\
\langle 3,0\rangle & \langle 3,1\rangle & \langle 3,2\rangle & \langle \mathbf{3,3}\rangle & \dots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

We have put the diagonal, here, in bold, since the subset of $\mathbb{N}^2$ consisting of the pairs lying on the diagonal, i.e.,

$$\{\langle 0,0\rangle, \langle 1,1\rangle, \langle 2,2\rangle, \dots\},$$

is the *identity relation on* $\mathbb{N}$. (Since the identity relation is popular, let's define $\mathrm{Id}_A = \{\langle x, x\rangle : x \in X\}$ for any set A.) The subset of all pairs lying above the diagonal, i.e.,

$$L = \{\langle 0,1\rangle, \langle 0,2\rangle, \dots, \langle 1,2\rangle, \langle 1,3\rangle, \dots, \langle 2,3\rangle, \langle 2,4\rangle, \dots\},$$

is the *less than* relation, i.e., Lnm iff $n < m$. The subset of pairs below the diagonal, i.e.,

$$G = \{\langle 1,0\rangle, \langle 2,0\rangle, \langle 2,1\rangle, \langle 3,0\rangle, \langle 3,1\rangle, \langle 3,2\rangle, \dots\},$$

is the *greater than* relation, i.e., Gnm iff $n > m$. The union of L with I, which we might call $K = L \cup I$, is the *less than or equal to* relation: Knm iff $n \leq m$. Similarly, $H = G \cup I$ is the *greater than or equal to relation.* These relations L, G, K, and H are special kinds of relations called *orders.* L and G have the property that no number bears L or G to itself (i.e., for all n, neither Lnn nor Gnn). Relations with this property are called *irreflexive*, and, if they also happen to be orders, they are called *strict orders.*

Although orders and identity are important and natural relations, it should be emphasized that according to our definition *any* subset of A^2 is a relation on A, regardless of how unnatural or contrived it seems. In particular, $\emptyset$ is a relation on any set (the *empty relation*, which no pair of elements bears), and A^2 itself is a relation on A as well (one which every pair bears), called the *universal relation.* But also something like $E = \{\langle n, m\rangle : n > 5 \text{ or } m \times n \geq 34\}$ counts as a relation.

B.2 Special Properties of Relations

Some kinds of relations turn out to be so common that they have been given special names. For instance, $\leq$ and $\subseteq$ both relate their respective domains (say, $\mathbb{N}$ in the case of $\leq$ and $\wp(A)$ in the case of $\subseteq$) in similar ways. To get at exactly how these relations are similar, and how they differ, we categorize them according to some special properties that relations can have. It turns out that (combinations of) some of these special properties are especially important: orders and equivalence relations.

Definition B.3 (Reflexivity). A relation $R \subseteq A^2$ is *reflexive* iff, for every $x \in A$, Rxx.

Definition B.4 (Transitivity). A relation $R \subseteq A^2$ is *transitive* iff, whenever Rxy and Ryz, then also Rxz.

Definition B.5 (Symmetry). A relation $R \subseteq A^2$ is *symmetric* iff, whenever Rxy, then also Ryx.

Definition B.6 (Anti-symmetry). A relation $R \subseteq A^2$ is *anti-symmetric* iff, whenever both Rxy and Ryx, then $x = y$ (or, in other words: if $x \neq y$ then either $\neg Rxy$ or $\neg Ryx$).

In a symmetric relation, Rxy and Ryx always hold together, or neither holds. In an anti-symmetric relation, the only way for Rxy and Ryx to hold together is if $x = y$. Note that this does not *require* that Rxy and Ryx holds when $x = y$, only that it isn't ruled out. So an anti-symmetric relation can be reflexive, but it is not the case that every anti-symmetric relation is reflexive. Also note that being anti-symmetric and merely not being symmetric are different conditions. In fact, a relation can be both symmetric and anti-symmetric at the same time (e.g., the identity relation is).

Definition B.7 (Connectivity). A relation $R \subseteq A^2$ is *connected* if for all $x, y \in X$, if $x \neq y$, then either Rxy or Ryx.

Definition B.8 (Irreflexivity). A relation R on A is called *irreflexive* if, for all $x \in A$, $\neg Rxx$.

Definition B.9 (Asymmetry). A relation R on A is called *asymmetric* if for no pair $x, y \in A$ we have Rxy and Ryx.

B.3 Equivalence Relations

The identity relation on a set is reflexive, symmetric, and transitive. Relations R that have all three of these properties are very common.

Definition B.10 (Equivalence relation). A relation $R \subseteq A^2$ that is reflexive, symmetric, and transitive is called an *equivalence relation.* Elements x and y of A are said to be *R-equivalent* if Rxy.

Equivalence relations give rise to the notion of an *equivalence class.* An equivalence relation "chunks up" the domain into different partitions. Within each partition, all the objects are related to one another; and no objects from different partitions relate to one another. Sometimes, it's helpful just to talk about these partitions *directly.* To that end, we introduce a definition:

Definition B.11. Let $R \subseteq A^2$ be an equivalence relation. For each $x \in A$, the *equivalence class* of x in A is the set $[x]_R = \{y \in A : Rxy\}$. The *quotient* of A under R is $A/_R = \{[x]_R : x \in A\}$, i.e., the set of these equivalence classes.

The next result vindicates the definition of an equivalence class, in proving that the equivalence classes are indeed the partitions of A:

Proposition B.12. *If $R \subseteq A^2$ is an equivalence relation, then Rxy iff $[x]_R = [y]_R$.*

Proof. For the left-to-right direction, suppose Rxy, and let $z \in [x]_R$. By definition, then, Rxz. Since R is an equivalence relation, Ryz. (Spelling this out: as Rxy and R is symmetric we have Ryx, and as Rxz and R is transitive we have Ryz.) So $z \in [y]_R$. Generalising, $[x]_R \subseteq [y]_R$. But exactly similarly, $[y]_R \subseteq [x]_R$. So $[x]_R = [y]_R$, by extensionality.

For the right-to-left direction, suppose $[x]_R = [y]_R$. Since R is reflexive, Ryy, so $y \in [y]_R$. Thus also $y \in [x]_R$ by the assumption that $[x]_R = [y]_R$. So Rxy. □

Example B.13. A nice example of equivalence relations comes from modular arithmetic. For any a, b, and $n \in \mathbb{N}$, say that $a \equiv_n b$ iff dividing a by n gives remainder b. (Somewhat more symbolically: $a \equiv_n b$ iff $(\exists k \in \mathbb{N})a - b = kn$.) Now, $\equiv_n$ is an equivalence relation, for any n. And there are exactly n distinct equivalence classes generated by $\equiv_n$; that is, $\mathbb{N}/_{\equiv_n}$ has n elements. These are: the set of numbers divisible by n without remainder, i.e., $[0]_{\equiv_n}$; the set of numbers divisible by n with remainder 1, i.e., $[1]_{\equiv_n}$; …; and the set of numbers divisible by n with remainder $n-1$, i.e., $[n-1]_{\equiv_n}$.

B.4 Orders

Many of our comparisons involve describing some objects as being "less than", "equal to", or "greater than" other objects, in a certain respect. These involve *order* relations. But there are different kinds of order relations. For instance, some require that any two objects be comparable, others don't. Some include identity (like $\leq$) and some exclude it (like $<$). It will help us to have a taxonomy here.

Definition B.14 (Preorder). A relation which is both reflexive and transitive is called a *preorder*.

Definition B.15 (Partial order). A preorder which is also anti-symmetric is called a *partial order*.

Definition B.16 (Linear order). A partial order which is also connected is called a *total order* or *linear order*.

Every linear order is also a partial order, and every partial order is also a preorder, but the converses don't hold.

Example B.17. Every linear order is also a partial order, and every partial order is also a preorder, but the converses don't hold. The universal relation on A is a preorder, since it is reflexive and transitive. But, if A has more than one element, the universal relation is not anti-symmetric, and so not a partial order.

Example B.18. Consider the *no longer than* relation $\preccurlyeq$ on $\mathbb{B}^*$: $x \preccurlyeq y$ iff $\text{len}(x) \leq \text{len}(y)$. This is a preorder (reflexive and transitive), and even connected, but not a partial order, since it is not anti-symmetric. For instance, $01 \preccurlyeq 10$ and $10 \preccurlyeq 01$, but $01 \neq 10$.

Example B.19. An important partial order is the relation $\subseteq$ on a set of sets. This is not in general a linear order, since if $a \neq b$ and we consider $\wp(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$, we see that $\{a\} \not\subseteq \{b\}$ and $\{a\} \neq \{b\}$ and $\{b\} \not\subseteq \{a\}$.

Example B.20. The relation of *divisibility without remainder* gives us a partial order which isn't a linear order. For integers n, m, we write $n \mid m$ to mean n (evenly) divides m, i.e., iff there is some integer k so that $m = kn$. On $\mathbb{N}$, this is a partial order, but not a linear order: for instance, $2 \nmid 3$ and also $3 \nmid 2$. Considered as a relation on $\mathbb{Z}$, divisibility is only a preorder since it is not anti-symmetric: $1 \mid -1$ and $-1 \mid 1$ but $1 \neq -1$.

Definition B.21 (Strict order). A *strict order* is a relation which is irreflexive, asymmetric, and transitive.

Definition B.22 (Strict linear order). A strict order which is also connected is called a *strict linear order.*

Example B.23. $\leq$ is the linear order corresponding to the strict linear order $<$. $\subseteq$ is the partial order corresponding to the strict order $\subsetneq$.

Definition B.24 (Total order). A strict order which is also connected is called a *total order.* This is also sometimes called a *strict linear order.*

Any strict order R on A can be turned into a partial order by adding the diagonal Id_A, i.e., adding all the pairs $\langle x, x\rangle$. (This is called the *reflexive closure* of R.) Conversely, starting from a partial order, one can get a strict order by removing Id_A. These next two results make this precise.

Proposition B.25. *If R is a strict order on A, then $R^+ = R \cup \mathrm{Id}_A$ is a partial order. Moreover, if R is total, then R^+ is a linear order.*

Proof. Suppose R is a strict order, i.e., $R \subseteq A^2$ and R is irreflexive, asymmetric, and transitive. Let $R^+ = R \cup \mathrm{Id}_A$. We have to show that R^+ is reflexive, antisymmetric, and transitive.

R^+ is clearly reflexive, since $\langle x, x\rangle \in \mathrm{Id}_A \subseteq R^+$ for all $x \in A$.

To show R^+ is antisymmetric, suppose for reductio that R^+xy and R^+yx but $x \neq y$. Since $\langle x, y\rangle \in R \cup \mathrm{Id}_X$, but $\langle x, y\rangle \notin \mathrm{Id}_X$, we must have $\langle x, y\rangle \in R$, i.e., Rxy. Similarly, Ryx. But this contradicts the assumption that R is asymmetric.

To establish transitivity, suppose that R^+xy and R^+yz. If both $\langle x, y\rangle \in R$ and $\langle y, z\rangle \in R$, then $\langle x, z\rangle \in R$ since R is transitive. Otherwise, either $\langle x, y\rangle \in \mathrm{Id}_X$, i.e., $x = y$, or $\langle y, z\rangle \in \mathrm{Id}_X$, i.e.,

$y = z$. In the first case, we have that R^+yz by assumption, $x = y$, hence R^+xz. Similarly in the second case. In either case, R^+xz, thus, R^+ is also transitive.

Concerning the "moreover" clause, suppose R is a total order, i.e., that R is connected. So for all $x \neq y$, either Rxy or Ryx, i.e., either $\langle x, y\rangle \in R$ or $\langle y, x\rangle \in R$. Since $R \subseteq R^+$, this remains true of R^+, so R^+ is connected as well. □

Proposition B.26. *If R is a partial order on X, then $R^- = R \setminus \mathrm{Id}_X$ is a strict order. Moreover, if R is linear, then R^- is total.*

Proof. This is left as an exercise. □

Example B.27. $\leq$ is the linear order corresponding to the total order $<$. $\subseteq$ is the partial order corresponding to the strict order $\subsetneq$.

The following simple result which establishes that total orders satisfy an extensionality-like property:

Proposition B.28. *If $<$ totally orders A, then:*

$$(\forall a, b \in A)((\forall x \in A)(x < a \leftrightarrow x < b) \rightarrow a = b)$$

Proof. Suppose $(\forall x \in A)(x < a \leftrightarrow x < b)$. If $a < b$, then $a < a$, contradicting the fact that $<$ is irreflexive; so $a \not< b$. Exactly similarly, $b \not< a$. So $a = b$, as $<$ is connected. □

B.5 Graphs

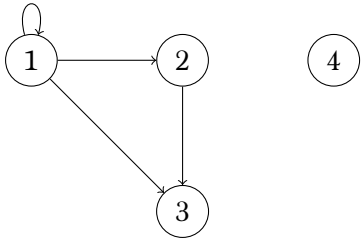
A *graph* is a diagram in which points—called "nodes" or "vertices" (plural of "vertex")—are connected by edges. Graphs are a ubiquitous tool in discrete mathematics and in computer science. They are incredibly useful for representing, and visualizing, relationships and structures, from concrete things like networks of various kinds to abstract structures such as the possible outcomes of decisions. There are many different kinds of graphs in

the literature which differ, e.g., according to whether the edges are directed or not, have labels or not, whether there can be edges from a node to the same node, multiple edges between the same nodes, etc. *Directed graphs* have a special connection to relations.

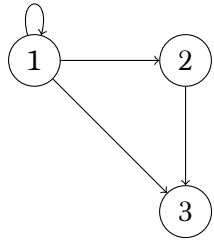
Definition B.29 (Directed graph). A *directed graph* $G = \langle V, E \rangle$ is a set of *vertices* V and a set of *edges* $E \subseteq V^2$.

According to our definition, a graph just is a set together with a relation on that set. Of course, when talking about graphs, it's only natural to expect that they are graphically represented: we can draw a graph by connecting two vertices v_1 and v_2 by an arrow iff $\langle v_1, v_2 \rangle \in E$. The only difference between a relation by itself and a graph is that a graph specifies the set of vertices, i.e., a graph may have isolated vertices. The important point, however, is that every relation R on a set X can be seen as a directed graph $\langle X, R \rangle$, and conversely, a directed graph $\langle V, E \rangle$ can be seen as a relation $E \subseteq V^2$ with the set V explicitly specified.

Example B.30. The graph $\langle V, E \rangle$ with $V = \{1, 2, 3, 4\}$ and $E = \{\langle 1, 1 \rangle, \langle 1, 2 \rangle, \langle 1, 3 \rangle, \langle 2, 3 \rangle\}$ looks like this:



This is a different graph than $\langle V', E\rangle$ with $V' = \{1, 2, 3\}$, which looks like this:



B.6 Operations on Relations

It is often useful to modify or combine relations. In Proposition B.25, we considered the *union* of relations, which is just the union of two relations considered as sets of pairs. Similarly, in Proposition B.26, we considered the relative difference of relations. Here are some other operations we can perform on relations.

Definition B.31. Let R, S be relations, and A be any set.

The *inverse* of R is $R^{-1} = \{\langle y, x\rangle : \langle x, y\rangle \in R\}$.

The *relative product* of R and S is $(R \mid S) = \{\langle x, z\rangle : \exists y(Rxy \wedge Syz)\}$.

The *restriction* of R to A is $R\restriction_A = R \cap A^2$.

The *application* of R to A is $R[A] = \{y : (\exists x \in A)Rxy\}$

Example B.32. Let $S \subseteq \mathbb{Z}^2$ be the successor relation on $\mathbb{Z}$, i.e., $S = \{\langle x, y\rangle \in \mathbb{Z}^2 : x + 1 = y\}$, so that Sxy iff $x + 1 = y$.

S^{-1} is the predecessor relation on $\mathbb{Z}$, i.e., $\{\langle x, y\rangle \in \mathbb{Z}^2 : x - 1 = y\}$.

$S \mid S$ is $\{\langle x, y\rangle \in \mathbb{Z}^2 : x + 2 = y\}$

$S\restriction_{\mathbb{N}}$ is the successor relation on $\mathbb{N}$.

$S[\{1, 2, 3\}]$ is $\{2, 3, 4\}$.

Definition B.33 (Transitive closure). Let $R \subseteq A^2$ be a binary relation.

The *transitive closure* of R is $R^+ = \bigcup_{0<n\in\mathbb{N}} R^n$, where we recursively define $R^1 = R$ and $R^{n+1} = R^n \mid R$.

The *reflexive transitive closure* of R is $R^* = R^+ \cup \mathrm{Id}_X$.

Example B.34. Take the successor relation $S \subseteq \mathbb{Z}^2$. S^2xy iff $x + 2 = y$, S^3xy iff $x + 3 = y$, etc. So S^+xy iff $x + n = y$ for some $n > 1$. In other words, S^+xy iff $x < y$, and S^*xy iff $x \leq y$.

Problems

Problem B.1. List the elements of the relation $\subseteq$ on the set $\wp(\{a, b, c\})$.

Problem B.2. Give examples of relations that are (a) reflexive and symmetric but not transitive, (b) reflexive and anti-symmetric, (c) anti-symmetric, transitive, but not reflexive, and (d) reflexive, symmetric, and transitive. Do not use relations on numbers or sets.

Problem B.3. Show that $\equiv_n$ is an equivalence relation, for any $n \in \mathbb{N}$, and that $\mathbb{N}/_{\equiv_n}$ has exactly n members.

Problem B.4. Give a proof of Proposition B.26.

Problem B.5. Consider the less-than-or-equal-to relation $\leq$ on the set $\{1, 2, 3, 4\}$ as a graph and draw the corresponding diagram.

Problem B.6. Show that the transitive closure of R is in fact transitive.

APPENDIX C

Syntax and Semantics

C.1 Introduction

Propositional logic deals with formulas that are built from propositional variables using the propositional connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$. Intuitively, a propositional variable p stands for a sentence or proposition that is true or false. Whenever the "truth value" of the propositional variable in a formula is determined, so is the truth value of any formulas formed from them using propositional connectives. We say that propositional logic is *truth functional*, because its semantics is given by functions of truth values. In particular, in propositional logic we leave out of consideration any further determination of truth and falsity, e.g., whether something is necessarily true rather than just contingently true, or whether something is known to be true, or whether something is true now rather than was true or will be true. We only consider two truth values true ($\mathbb{T}$) and false ($\mathbb{F}$), and so exclude from discussion the possibility that a statement may be neither true nor false, or only half true. We also concentrate only on connectives where the truth value of a formula built from them is completely determined by the truth values of its parts (and not, say, on its meaning). In particular, whether the truth value of conditionals

in English is truth functional in this sense is contentious. The material conditional $\rightarrow$ is; other logics deal with conditionals that are not truth functional.

In order to develop the theory and metatheory of truth-functional propositional logic, we must first define the syntax and semantics of its expressions. We will describe one way of constructing formulas from propositional variables using the connectives. Alternative definitions are possible. Other systems will chose different symbols, will select different sets of connectives as primitive, will use parentheses differently (or even not at all, as in the case of so-called Polish notation). What all approaches have in common, though, is that the formation rules define the set of formulas *inductively*. If done properly, every expression can result essentially in only one way according to the formation rules. The inductive definition resulting in expressions that are *uniquely readable* means we can give meanings to these expressions using the same method—inductive definition.

Giving the meaning of expressions is the domain of semantics. The central concept in semantics for propositonal logic is that of satisfaction in a valuation. A valuation v assigns truth values $\mathbb{T}$, $\mathbb{F}$ to the propositional variables. Any valuation determines a truth value $\overline{v}(A)$ for any formula A. A formula is satisfied in a valuation v iff $\overline{v}(A) = \mathbb{T}$—we write this as $v \vDash A$. This relation can also be defined by induction on the structure of A, using the truth functions for the logical connectives to define, say, satisfaction of $A \wedge B$ in terms of satisfaction (or not) of A and B.

On the basis of the satisfaction relation $v \vDash A$ for sentences we can then define the basic semantic notions of tautology, entailment, and satisfiability. A formula is a tautology, $\vDash A$, if every valuation satisfies it, i.e., $\overline{v}(A) = \mathbb{T}$ for any v. It is entailed by a set of formulas, $\Gamma \vDash A$, if every valuation that satisfies all the formulas in Γ also satisfies A. And a set of formulas is satisfiable if some valuation satisfies all formulas in it at the same time. Because formulas are inductively defined, and satisfaction is in turn defined by induction on the structure of formulas, we can use induction to prove properties of our semantics and to relate

the semantic notions defined.

C.2 Propositional Formulas

Formulas of propositional logic are built up from *propositional variables* and the propositional constant $\bot$ using *logical connectives.*

1. A countably infinite set At_0 of propositional variables p_0, $p_1, \ldots$
2. The propositional constant for falsity $\bot$.
3. The logical connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional)
4. Punctuation marks: (,), and the comma.

We denote this language of propositional logic by $\mathscr{L}_0$.

In addition to the primitive connectives introduced above, we also use the following *defined* symbols: $\leftrightarrow$ (biconditional), $\top$ (truth)

A defined symbol is not officially part of the language, but is introduced as an informal abbreviation: it allows us to abbreviate formulas which would, if we only used primitive symbols, get quite long. This is obviously an advantage. The bigger advantage, however, is that proofs become shorter. If a symbol is primitive, it has to be treated separately in proofs. The more primitive symbols, therefore, the longer our proofs.

You may be familiar with different terminology and symbols than the ones we use above. Logic texts (and teachers) commonly use either $\sim$, $\neg$, and ! for "negation", $\wedge$, $\cdot$, and & for "conjunction". Commonly used symbols for the "conditional" or "implication" are $\rightarrow$, $\Rightarrow$, and $\supset$. Symbols for "biconditional," "bi-implication," or "(material) equivalence" are $\leftrightarrow$, $\Leftrightarrow$, and $\equiv$. The $\bot$ symbol is variously called "falsity," "falsum," "absurdity," or "bottom." The $\top$ symbol is variously called "truth," "verum," or "top."

Definition C.1 (Formula). The set $\mathrm{Frm}(\mathscr{L}_0)$ of *formulas* of propositional logic is defined inductively as follows:

1. $\bot$ is an atomic formula.
2. Every propositional variable p_i is an atomic formula.
3. If A is a formula, then $\neg A$ is formula.
4. If A and B are formulas, then $(A \wedge B)$ is a formula.
5. If A and B are formulas, then $(A \vee B)$ is a formula.
6. If A and B are formulas, then $(A \to B)$ is a formula.
7. Nothing else is a formula.

The definition of formulas is an *inductive definition*. Essentially, we construct the set of formulas in infinitely many stages. In the initial stage, we pronounce all atomic formulas to be formulas; this corresponds to the first few cases of the definition, i.e., the cases for $\bot$, p_i. "Atomic formula" thus means any formula of this form.

The other cases of the definition give rules for constructing new formulas out of formulas already constructed. At the second stage, we can use them to construct formulas out of atomic formulas. At the third stage, we construct new formulas from the atomic formulas and those obtained in the second stage, and so on. A formula is anything that is eventually constructed at such a stage, and nothing else.

Definition C.2. Formulas constructed using the defined operators are to be understood as follows:

1. $\top$ abbreviates $\neg\bot$.
2. $A \leftrightarrow B$ abbreviates $(A \to B) \wedge (B \to A)$.

Definition C.3 (Syntactic identity). The symbol $\equiv$ expresses syntactic identity between strings of symbols, i.e., $A \equiv B$ iff A and B are strings of symbols of the same length and which contain the same symbol in each place.

The $\equiv$ symbol may be flanked by strings obtained by concatenation, e.g., $A \equiv (B \vee C)$ means: the string of symbols A is the same string as the one obtained by concatenating an opening parenthesis, the string B, the $\vee$ symbol, the string C, and a closing parenthesis, in this order. If this is the case, then we know that the first symbol of A is an opening parenthesis, A contains B as a substring (starting at the second symbol), that substring is followed by $\vee$, etc.

C.3 Preliminaries

Theorem C.4 (*Principle of induction on formulas*). *If some property P holds for all the atomic formulas and is such that*

1. *it holds for $\neg A$ whenever it holds for A;*
2. *it holds for $(A \wedge B)$ whenever it holds for A and B;*
3. *it holds for $(A \vee B)$ whenever it holds for A and B;*
4. *it holds for $(A \to B)$ whenever it holds for A and B;*

then P holds for all formulas.

Proof. Let S be the collection of all formulas with property P. Clearly $S \subseteq \mathrm{Frm}(\mathcal{L}_0)$. S satisfies all the conditions of Definition C.1: it contains all atomic formulas and is closed under the logical operators. $\mathrm{Frm}(\mathcal{L}_0)$ is the smallest such class, so $\mathrm{Frm}(\mathcal{L}_0) \subseteq S$. So $\mathrm{Frm}(\mathcal{L}_0) = S$, and every formula has property P. □

Proposition C.5. *Any formula in* $\mathrm{Frm}(\mathscr{L}_0)$ *is* balanced, *in that it has as many left parentheses as right ones.*

Proposition C.6. *No proper initial segment of a formula is a formula.*

Proposition C.7 (Unique Readability). *Any formula* A *in* $\mathrm{Frm}(\mathscr{L}_0)$ *has exactly one parsing as one of the following*

1. $\bot$.
2. p_n *for some* $p_n \in \mathrm{At}_0$.
3. $\neg B$ *for some formula* B.
4. $(B \wedge C)$ *for some formulas* B *and* C.
5. $(B \vee C)$ *for some formulas* B *and* C.
6. $(B \to C)$ *for some formulas* B *and* C.

Moreover, this parsing is unique.

Proof. By induction on A. For instance, suppose that A has two distinct readings as $(B \to C)$ and $(B' \to C')$. Then B and B' must be the same (or else one would be a proper initial segment of the other); so if the two readings of A are distinct it must be because C and C' are distinct readings of the same sequence of symbols, which is impossible by the inductive hypothesis. □

Definition C.8 (Uniform Substitution). If A and B are formulas, and p_i is a propositional variable, then $A[B/p_i]$ denotes the result of replacing each occurrence of p_i by an occurrence of B in A; similarly, the simultaneous substitution of $p_1, \ldots, p_n$ by formulas $B_1, \ldots, B_n$ is denoted by $A[B_1/p_1, \ldots, B_n/p_n]$.

C.4 Valuations and Satisfaction

Definition C.9 (Valuations). Let $\{\mathbb{T}, \mathbb{F}\}$ be the set of the two truth values, "true" and "false." A *valuation* for $\mathscr{L}_0$ is a function v assigning either $\mathbb{T}$ or $\mathbb{F}$ to the propositional variables of the language, i.e., $v\colon \mathrm{At}_0 \to \{\mathbb{T}, \mathbb{F}\}$.

Definition C.10. Given a valuation v, define the evaluation function $\overline{v}\colon \mathrm{Frm}(\mathscr{L}_0) \to \{\mathbb{T}, \mathbb{F}\}$ inductively by:

$$\overline{v}(\bot) = \mathbb{F};$$

$$\overline{v}(p_n) = v(p_n);$$

$$\overline{v}(\neg A) = \begin{cases} \mathbb{T} & \text{if } \overline{v}(A) = \mathbb{F}; \\ \mathbb{F} & \text{otherwise.} \end{cases}$$

$$\overline{v}(A \wedge B) = \begin{cases} \mathbb{T} & \text{if } \overline{v}(A) = \mathbb{T} \text{ and } \overline{v}(B) = \mathbb{T}; \\ \mathbb{F} & \text{if } \overline{v}(A) = \mathbb{F} \text{ or } \overline{v}(B) = \mathbb{F}. \end{cases}$$

$$\overline{v}(A \vee B) = \begin{cases} \mathbb{T} & \text{if } \overline{v}(A) = \mathbb{T} \text{ or } \overline{v}(B) = \mathbb{T}; \\ \mathbb{F} & \text{if } \overline{v}(A) = \mathbb{F} \text{ and } \overline{v}(B) = \mathbb{F}. \end{cases}$$

$$\overline{v}(A \to B) = \begin{cases} \mathbb{T} & \text{if } \overline{v}(A) = \mathbb{F} \text{ or } \overline{v}(B) = \mathbb{T}; \\ \mathbb{F} & \text{if } \overline{v}(A) = \mathbb{T} \text{ and } \overline{v}(B) = \mathbb{F}. \end{cases}$$

The clauses correspond to the following truth tables:

A	$\neg A$
$\mathbb{T}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$

A	B	$A \wedge B$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{F}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$

A	B	$A \vee B$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$

A	B	$A \to B$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{F}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{T}$

Theorem C.11 (Local Determination). *Suppose that v_1 and v_2 are valuations that agree on the propositional letters occurring in A, i.e., $v_1(p_n) = v_2(p_n)$ whenever p_n occurs in some formula A. Then $\overline{v_1}$ and $\overline{v_2}$ also agree on A, i.e., $\overline{v_1}(A) = \overline{v_2}(A)$.*

Proof. By induction on A. □

Definition C.12 (Satisfaction). Using the evaluation function, we can define the notion of *satisfaction of a formula A by a valuation v*, $v \vDash A$, inductively as follows. (We write $v \nvDash A$ to mean "not $v \vDash A$.")

1. $A \equiv \bot$: $v \nvDash A$.
2. $A \equiv p_i$: $v \vDash A$ iff $v(p_i) = \mathbb{T}$.
3. $A \equiv \neg B$: $v \vDash A$ iff $v \nvDash B$.
4. $A \equiv (B \wedge C)$: $v \vDash A$ iff $v \vDash B$ and $v \vDash C$.
5. $A \equiv (B \vee C)$: $v \vDash A$ iff $v \vDash A$ or $v \vDash B$ (or both).
6. $A \equiv (B \to C)$: $v \vDash A$ iff $v \nvDash B$ or $v \vDash C$ (or both).

If Γ is a set of formulas, $v \vDash \Gamma$ iff $v \vDash A$ for every $A \in \Gamma$.

Proposition C.13. *$v \vDash A$ iff $\overline{v}(A) = \mathbb{T}$.*

Proof. By induction on A. □

C.5 Semantic Notions

We define the following semantic notions:

Definition C.14. 1. A formula A is *satisfiable* if for some v, $v \vDash A$; it is *unsatisfiable* if for no v, $v \vDash A$;

2. A formula A is a *tautology* if $v \vDash A$ for all valuations v;

3. A formula A is *contingent* if it is satisfiable but not a tautology;

4. If Γ is a set of formulas, $\Gamma \vDash A$ ("Γ entails A") if and only if $v \vDash A$ for every valuation v for which $v \vDash \Gamma$.

5. If Γ is a set of formulas, Γ is *satisfiable* if there is a valuation v for which $v \vDash \Gamma$, and Γ is *unsatisfiable* otherwise.

Proposition C.15. *1. A is a tautology if and only if $\emptyset \vDash A$;*

2. If $\Gamma \vDash A$ and $\Gamma \vDash A \to B$ then $\Gamma \vDash B$;

3. If Γ is satisfiable then every finite subset of Γ is also satisfiable;

4. Monotony: if $\Gamma \subseteq \Delta$ and $\Gamma \vDash A$ then also $\Delta \vDash A$;

5. Transitivity: if $\Gamma \vDash A$ and $\Delta \cup \{A\} \vDash B$ then $\Gamma \cup \Delta \vDash B$;

Proof. Exercise. □

Proposition C.16. *$\Gamma \vDash A$ if and only if $\Gamma \cup \{\neg A\}$ is unsatisfiable;*

Proof. Exercise. □

Theorem C.17 (Semantic Deduction Theorem). *$\Gamma \vDash A \to B$ if and only if $\Gamma \cup \{A\} \vDash B$.*

Proof. Exercise. □

Problems

Problem C.1. Prove Proposition C.5

Problem C.2. Prove Proposition C.6

Problem C.3. Give a mathematically rigorous definition of $A[B/p]$ by induction.

Problem C.4. Prove Proposition C.13

Problem C.5. Prove Proposition C.15

Problem C.6. Prove Proposition C.16

Problem C.7. Prove Theorem C.17

APPENDIX D

Axiomatic Derivations

D.1 Introduction

Logics commonly have both a semantics and a derivation system. The semantics concerns concepts such as truth, satisfiability, validity, and entailment. The purpose of derivation systems is to provide a purely syntactic method of establishing entailment and validity. They are purely syntactic in the sense that a derivation in such a system is a finite syntactic object, usually a sequence (or other finite arrangement) of sentences or formulas. Good derivation systems have the property that any given sequence or arrangement of sentences or formulas can be verified mechanically to be "correct."

The simplest (and historically first) derivation systems for first-order logic were *axiomatic*. A sequence of formulas counts as a derivation in such a system if each individual formula in it is either among a fixed set of "axioms" or follows from formulas coming before it in the sequence by one of a fixed number of "inference rules"—and it can be mechanically verified if a formula is an axiom and whether it follows correctly from other formulas by one of the inference rules. Axiomatic proof systems are easy to describe—and also easy to handle meta-theoretically—

but derivations in them are hard to read and understand, and are also hard to produce.

Other derivation systems have been developed with the aim of making it easier to construct derivations or easier to understand derivations once they are complete. Examples are natural deduction, truth trees, also known as tableaux proofs, and the sequent calculus. Some derivation systems are designed especially with mechanization in mind, e.g., the resolution method is easy to implement in software (but its derivations are essentially impossible to understand). Most of these other proof systems represent derivations as trees of formulas rather than sequences. This makes it easier to see which parts of a derivation depend on which other parts.

So for a given logic, such as first-order logic, the different derivation systems will give different explications of what it is for a sentence to be a *theorem* and what it means for a sentence to be derivable from some others. However that is done (via axiomatic derivations, natural deductions, sequent derivations, truth trees, resolution refutations), we want these relations to match the semantic notions of validity and entailment. Let's write $\vdash A$ for "A is a theorem" and "$\Gamma \vdash A$" for "A is derivable from Γ." However $\vdash$ is defined, we want it to match up with $\models$, that is:

1. $\vdash A$ if and only if $\models A$
2. $\Gamma \vdash A$ if and only if $\Gamma \models A$

The "only if" direction of the above is called *soundness*. A derivation system is sound if derivability guarantees entailment (or validity). Every decent derivation system has to be sound; unsound derivation systems are not useful at all. After all, the entire purpose of a derivation is to provide a syntactic guarantee of validity or entailment. We'll prove soundness for the derivation systems we present.

The converse "if" direction is also important: it is called *completeness*. A complete derivation system is strong enough to show

that A is a theorem whenever A is valid, and that $\Gamma \vdash A$ whenever $\Gamma \models A$. Completeness is harder to establish, and some logics have no complete derivation systems. First-order logic does. Kurt Gödel was the first one to prove completeness for a derivation system of first-order logic in his 1929 dissertation.

Another concept that is connected to derivation systems is that of *consistency*. A set of sentences is called inconsistent if anything whatsoever can be derived from it, and consistent otherwise. Inconsistency is the syntactic counterpart to unsatisfiablity: like unsatisfiable sets, inconsistent sets of sentences do not make good theories, they are defective in a fundamental way. Consistent sets of sentences may not be true or useful, but at least they pass that minimal threshold of logical usefulness. For different derivation systems the specific definition of consistency of sets of sentences might differ, but like $\vdash$, we want consistency to coincide with its semantic counterpart, satisfiability. We want it to always be the case that Γ is consistent if and only if it is satisfiable. Here, the "if" direction amounts to completeness (consistency guarantees satisfiability), and the "only if" direction amounts to soundness (satisfiability guarantees consistency). In fact, for classical first-order logic, the two versions of soundness and completeness are equivalent.

D.2 Axiomatic Derivations

Axiomatic derivations are the oldest and simplest logical derivation systems. Its derivations are simply sequences of sentences. A sequence of sentences conunts as a correct derivation if every sentence A in it satisfies one of the following conditions:

1. A is an axiom, or
2. A is an element of a given set Γ of sentences, or
3. A is justified by a rule of inference.

To be an axiom, A has to have the form of on of a number of fixed sentence schemas. There are many sets of axiom schemas that provide a satisfactory (sound and complete) derivation system for first-order logic. Some are organized according to the connectives they govern, e.g., the schemas

$$A \to (B \to A) \qquad B \to (B \vee C) \qquad (B \wedge C) \to B$$

are common axioms that govern $\to$, $\vee$ and $\wedge$. Some axiom systems aim at a minimal number of axioms. Depending on the connectives that are taken as primitives, it is even possible to find axiom systems that consist of a single axiom.

A rule of inference is a conditional statement that gives a sufficient condition for a sentence in a derivation to be justified. Modus ponens is one very common such rule: it says that if A and $A \to B$ are already justified, then B is justified. This means that a line in a derivation containing the sentence B is justified, provided that both A and $A \to B$ (for some sentence A) appear in the derivation before B.

The $\vdash$ relation based on axiomatic derivations is defined as follows: $\Gamma \vdash A$ iff there is a derivation with the sentence A as its last formula (and Γ is taken as the set of sentences in that derivation which are justified by (2) above). A is a theorem if A has a derivation where Γ is empty, i.e., every sentence in the derivation is justfied either by (1) or (3). For instance, here is a derivation that shows that $\vdash A \to (B \to (B \vee A))$:

1. $B \to (B \vee A)$
2. $(B \to (B \vee A)) \to (A \to (B \to (B \vee A)))$
3. $A \to (B \to (B \vee A))$

The sentence on line 1 is of the form of the axiom $A \to (A \vee B)$ (with the roles of A and B reversed). The sentence on line 2 is of the form of the axiom $A \to (B \to A)$. Thus, both lines are justified. Line 3 is justified by modus ponens: if we abbreviate it as D, then line 2 has the form $C \to D$, where C is $B \to (B \vee A)$, i.e., line 1.

A set Γ is inconsistent if $\Gamma \vdash \bot$. A complete axiom system will also prove that $\bot \to A$ for any A, and so if Γ is inconsistent, then $\Gamma \vdash A$ for any A.

Systems of axiomatic derivations for logic were first given by Gottlob Frege in his 1879 *Begriffsschrift*, which for this reason is often considered the first work of modern logic. They were perfected in Alfred North Whitehead and Bertrand Russell's *Principia Mathematica* and by David Hilbert and his students in the 1920s. They are thus often called "Frege systems" or "Hilbert systems." They are very versatile in that it is often easy to find an axiomatic system for a logic. Because derivations have a very simple structure and only one or two inference rules, it is also relatively easy to prove things *about* them. However, they are very hard to use in practice, i.e., it is difficult to find and write proofs.

D.3 Rules and Derivations

Axiomatic derivations are perhaps the simplest proof system for logic. A derivation is just a sequence of formulas. To count as a derivation, every formula in the sequence must either be an instance of an axiom, or must follow from one or more formulas that precede it in the sequence by a rule of inference. A derivation derives its last formula.

Definition D.1 (Derivability). If Γ is a set of formulas of $\mathcal{L}$ then a *derivation* from Γ is a finite sequence $A_1, \ldots, A_n$ of formulas where for each $i \leq n$ one of the following holds:

1. $A_i \in \Gamma$; or
2. A_i is an axiom; or
3. A_i follows from some A_j (and A_k) with $j < i$ (and $k < i$) by a rule of inference.

What counts as a correct derivation depends on which inference rules we allow (and of course what we take to be axioms).

And an inference rule is an if-then statement that tells us that, under certain conditions, a step A_i in is a correct inference step.

Definition D.2 (Rule of inference). A *rule of inference* gives a sufficient condition for what counts as a correct inference step in a derivation from Γ.

For instance, since any one-element sequence A with $A \in \Gamma$ trivially counts as a derivation, the following might be a very simple rule of inference:

> If $A \in \Gamma$, then A is always a correct inference step in any derivation from Γ.

Similarly, if A is one of the axioms, then A by itself is a derivation, and so this is also a rule of inference:

> If A is an axiom, then A is a correct inference step.

It gets more interesting if the rule of inference appeals to formulas that appear before the step considered. The following rule is called *modus ponens:*

> If $B \to A$ and B occur higher up in the derivation, then A is a correct inference step.

If this is the only rule of inference, then our definition of derivation above amounts to this: $A_1, \ldots, A_n$ is a derivation iff for each $i \leq n$ one of the following holds:

1. $A_i \in \Gamma$; or
2. A_i is an axiom; or
3. for some $j < i$, A_j is $B \to A_i$, and for some $k < i$, A_k is B.

The last clause says that A_i follows from A_j (B) and A_k ($B \to A_i$) by modus ponens. If we can go from 1 to n, and each time we find a formula A_i that is either in Γ, an axiom, or which a rule of inference tells us that it is a correct inference step, then the entire sequence counts as a correct derivation.

Definition D.3 (Derivability). A formula A is *derivable* from Γ, written $\Gamma \vdash A$, if there is a derivation from Γ ending in A.

Definition D.4 (Theorems). A formula A is a *theorem* if there is a derivation of A from the empty set. We write $\vdash A$ if A is a theorem and $\nvdash A$ if it is not.

D.4 Axiom and Rules for the Propositional Connectives

Definition D.5 (Axioms). The set of Ax_0 of *axioms* for the propositional connectives comprises all formulas of the following forms:

$$(A \wedge B) \to A \tag{D.1}$$
$$(A \wedge B) \to B \tag{D.2}$$
$$A \to (B \to (A \wedge B)) \tag{D.3}$$
$$A \to (A \vee B) \tag{D.4}$$
$$A \to (B \vee A) \tag{D.5}$$
$$(A \to C) \to ((B \to C) \to ((A \vee B) \to C)) \tag{D.6}$$
$$A \to (B \to A) \tag{D.7}$$
$$(A \to (B \to C)) \to ((A \to B) \to (A \to C)) \tag{D.8}$$
$$(A \to B) \to ((A \to \neg B) \to \neg A) \tag{D.9}$$
$$\neg A \to (A \to B) \tag{D.10}$$
$$\top \tag{D.11}$$
$$\bot \to A \tag{D.12}$$
$$(A \to \bot) \to \neg A \tag{D.13}$$
$$\neg\neg A \to A \tag{D.14}$$

Definition D.6 (Modus ponens). If B and $B \to A$ already occur in a derivation, then A is a correct inference step.

We'll abbreviate the rule modus ponens as "MP."

D.5 Examples of Derivations

Example D.7. Suppose we want to prove $(\neg D \vee E) \to (D \to E)$. Clearly, this is not an instance of any of our axioms, so we have to use the MP rule to derive it. Our only rule is MP, which given A and $A \to B$ allows us to justify B. One strategy would be to use eq. (D.6) with A being $\neg D$, B being E, and C being $D \to E$, i.e., the instance

$$(\neg D \to (D \to E)) \to ((E \to (D \to E)) \to ((\neg D \vee E) \to (D \to E))).$$

Why? Two applications of MP yield the last part, which is what we want. And we easily see that $\neg D \to (D \to E)$ is an instance of eq. (D.10), and $E \to (D \to E)$ is an instance of eq. (D.7). So our derivation is:

1.	$\neg D \to (D \to E)$	eq. (D.7)
2.	$(\neg D \to (D \to E)) \to$ $((E \to (D \to E)) \to ((\neg D \vee E) \to (D \to E)))$	eq. (D.6)
3.	$((E \to (D \to E)) \to ((\neg D \vee E) \to (D \to E))$	1, 2, MP
4.	$E \to (D \to E)$	eq. (D.7)
5.	$(\neg D \vee E) \to (D \to E)$	3, 4, MP

Example D.8. Let's try to find a derivation of $D \to D$. It is not an instance of an axiom, so we have to use MP to derive it. eq. (D.7) is an axiom of the form $A \to B$ to which we could apply MP. To be useful, of course, the B which MP would justify as a correct step in this case would have to be $D \to D$, since this is what we want to derive. That means A would also have to be D, i.e., we might look at this instance of eq. (D.7):

$$D \to (D \to D)$$

In order to apply MP, we would also need to justify the corresponding second premise, namely A. But in our case, that would be D, and we won't be able to derive D by itself. So we need a different strategy.

The other axiom involving just $\to$ is eq. (D.8), i.e.,

$$(A \to (B \to C)) \to ((A \to B) \to (A \to C))$$

We could get to the last nested conditional by applying MP twice. Again, that would mean that we want an instance of eq. (D.8) where $A \to C$ is $D \to D$, the formula we are aiming for. Then of course, A and C are both D. How should we pick B so that both $A \to (B \to C)$ and $A \to B$, i.e., in our case $D \to (B \to D)$ and $D \to B$, are also derivable? Well, the first of these is already an instance of eq. (D.7), whatever we decide B to be. And $D \to B$ would be another instance of eq. (D.7) if B were $(D \to D)$. So, our derivation is:

1.	$D \to ((D \to D) \to D)$	eq. (D.7)
2.	$(D \to ((D \to D) \to D)) \to$ $((D \to (D \to D)) \to (D \to D))$	eq. (D.8)
3.	$(D \to (D \to D)) \to (D \to D)$	1, 2, MP
4.	$D \to (D \to D)$	eq. (D.7)
5.	$D \to D$	3, 4, MP

Example D.9. Sometimes we want to show that there is a derivation of some formula from some other formulas Γ. For instance, let's show that we can derive $A \to C$ from $\Gamma = \{A \to B, B \to C\}$.

1.	$A \to B$	HYP
2.	$B \to C$	HYP
3.	$(B \to C) \to (A \to (B \to C))$	eq. (D.7)
4.	$A \to (B \to C)$	2, 3, MP
5.	$(A \to (B \to C)) \to$ $((A \to B) \to (A \to C))$	eq. (D.8)
6.	$((A \to B) \to (A \to C))$	4, 5, MP
7.	$A \to C$	1, 6, MP

The lines labelled "Hyp" (for "hypothesis") indicate that the formula on that line is an element of Γ.

Proposition D.10. *If $\Gamma \vdash A \to B$ and $\Gamma \vdash B \to C$, then $\Gamma \vdash A \to C$*

Proof. Suppose $\Gamma \vdash A \to B$ and $\Gamma \vdash B \to C$. Then there is a derivation of $A \to B$ from Γ; and a derivation of $B \to C$ from Γ as well. Combine these into a single derivation by concatenating them. Now add lines 3–7 of the derivation in the preceding example. This is a derivation of $A \to C$—which is the last line of the new derivation—from Γ. Note that the justifications of lines 4 and 7 remain valid if the reference to line number 2 is replaced by reference to the last line of the derivation of $A \to B$, and reference to line number 1 by reference to the last line of the derivation of $B \to C$. □

D.6 Proof-Theoretic Notions

Just as we've defined a number of important semantic notions (tautology, entailment, satisfiabilty), we now define corresponding *proof-theoretic notions.* These are not defined by appeal to satisfaction of sentences in structures, but by appeal to the derivability or non-derivability of certain formulas. It was an important discovery that these notions coincide. That they do is the content of the *soundness* and *completeness theorems.*

Definition D.11 (Derivability). A formula A is *derivable* from Γ, written $\Gamma \vdash A$, if there is a derivation from Γ ending in A.

Definition D.12 (Theorems). A formula A is a *theorem* if there is a derivation of A from the empty set. We write $\vdash A$ if A is a theorem and $\nvdash A$ if it is not.

Definition D.13 (Consistency). A set Γ of formulas is *consistent* if and only if $\Gamma \nvdash \bot$; it is *inconsistent* otherwise.

Proposition D.14 (Reflexivity). *If $A \in \Gamma$, then $\Gamma \vdash A$.*

Proof. The formula A by itself is a derivation of A from Γ. □

Proposition D.15 (Monotony). *If $\Gamma \subseteq \Delta$ and $\Gamma \vdash A$, then $\Delta \vdash A$.*

Proof. Any derivation of A from Γ is also a derivation of A from Δ. □

Proposition D.16 (Transitivity). *If $\Gamma \vdash A$ and $\{A\} \cup \Delta \vdash B$, then $\Gamma \cup \Delta \vdash B$.*

Proof. Suppose $\{A\} \cup \Delta \vdash B$. Then there is a derivation $B_1, \ldots, B_l = B$ from $\{A\} \cup \Delta$. Some of the steps in that derivation will be correct because of a rule which refers to a prior line $B_i = A$. By hypothesis, there is a derivation of A from Γ, i.e., a derivation $A_1, \ldots, A_k = A$ where every A_i is an axiom, an element of Γ, or correct by a rule of inference. Now consider the sequence

$$A_1, \ldots, A_k = A, B_1, \ldots, B_l = B.$$

This is a correct derivation of B from $\Gamma \cup \Delta$ since every $B_i = A$ is now justified by the same rule which justifies $A_k = A$. □

Note that this means that in particular if $\Gamma \vdash A$ and $A \vdash B$, then $\Gamma \vdash B$. It follows also that if $A_1, \ldots, A_n \vdash B$ and $\Gamma \vdash A_i$ for each i, then $\Gamma \vdash B$.

Proposition D.17. *Γ is inconsistent iff $\Gamma \vdash A$ for every A.*

Proof. Exercise. □

Proposition D.18 (Compactness). *1. If $\Gamma \vdash A$ then there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \vdash A$.*

2. If every finite subset of Γ is consistent, then Γ is consistent.

Proof. 1. If $\Gamma \vdash A$, then there is a finite sequence of formulas $A_1, \ldots, A_n$ so that $A \equiv A_n$ and each A_i is either a logical axiom, an element of Γ or follows from previous formulas by modus ponens. Take Γ_0 to be those A_i which are in Γ. Then the derivation is likewise a derivation from Γ_0, and so $\Gamma_0 \vdash A$.

2. This is the contrapositive of (1) for the special case $A \equiv \bot$. □

D.7 The Deduction Theorem

As we've seen, giving derivations in an axiomatic system is cumbersome, and derivations may be hard to find. Rather than actually write out long lists of formulas, it is generally easier to argue that such derivations exist, by making use of a few simple results. We've already established three such results: Proposition D.14 says we can always assert that $\Gamma \vdash A$ when we know that $A \in \Gamma$. Proposition D.15 says that if $\Gamma \vdash A$ then also $\Gamma \cup \{B\} \vdash A$. And Proposition D.16 implies that if $\Gamma \vdash A$ and $A \vdash B$, then $\Gamma \vdash B$. Here's another simple result, a "meta"-version of modus ponens:

Proposition D.19. *If $\Gamma \vdash A$ and $\Gamma \vdash A \to B$, then $\Gamma \vdash B$.*

Proof. We have that $\{A, A \to B\} \vdash B$:

1.	A	Hyp.
2.	$A \to B$	Hyp.
3.	B	1, 2, MP

By Proposition D.16, $\Gamma \vdash B$. □

The most important result we'll use in this context is the deduction theorem:

Theorem D.20 (Deduction Theorem). *$\Gamma \cup \{A\} \vdash B$ if and only if $\Gamma \vdash A \to B$.*

Proof. The "if" direction is immediate. If $\Gamma \vdash A \to B$ then also $\Gamma \cup \{A\} \vdash A \to B$ by Proposition D.15. Also, $\Gamma \cup \{A\} \vdash A$ by Proposition D.14. So, by Proposition D.19, $\Gamma \cup \{A\} \vdash B$.

For the "only if" direction, we proceed by induction on the length of the derivation of B from $\Gamma \cup \{A\}$.

For the induction basis, we prove the claim for every derivation of length 1. A derivation of B from $\Gamma \cup \{A\}$ of length 1 consists of B by itself; and if it is correct B is either $\in \Gamma \cup \{A\}$ or is an axiom. If $B \in \Gamma$ or is an axiom, then $\Gamma \vdash B$. We also have that $\Gamma \vdash B \to (A \to B)$ by eq. (D.7), and Proposition D.19 gives $\Gamma \vdash A \to B$. If $B \in \{A\}$ then $\Gamma \vdash A \to B$ because then last sentence $A \to B$ is the same as $A \to A$, and we have derived that in Example D.8.

For the inductive step, suppose a derivation of B from $\Gamma \cup \{A\}$ ends with a step B which is justified by modus ponens. (If it is not justified by modus ponens, $B \in \Gamma$, $B \equiv A$, or B is an axiom, and the same reasoning as in the induction basis applies.) Then some previous steps in the derivation are $C \to B$ and C, for some formula C, i.e., $\Gamma \cup \{A\} \vdash C \to B$ and $\Gamma \cup \{A\} \vdash C$, and the respective derivations are shorter, so the inductive hypothesis applies to them. We thus have both:

$$\Gamma \vdash A \to (C \to B);$$
$$\Gamma \vdash A \to C.$$

But also

$$\Gamma \vdash (A \to (C \to B)) \to ((A \to C) \to (A \to B)),$$

by eq. (D.8), and two applications of Proposition D.19 give $\Gamma \vdash A \to B$, as required. □

Notice how eq. (D.7) and eq. (D.8) were chosen precisely so that the Deduction Theorem would hold.

The following are some useful facts about derivability, which we leave as exercises.

Proposition D.21. *1.* $\vdash (A \to B) \to ((B \to C) \to (A \to C))$;

2. If $\Gamma \cup \{\neg A\} \vdash \neg B$ *then* $\Gamma \cup \{B\} \vdash A$ *(Contraposition);*

3. $\{A, \neg A\} \vdash B$ *(Ex Falso Quodlibet, Explosion);*

4. $\{\neg\neg A\} \vdash A$ *(Double Negation Elimination);*

5. If $\Gamma \vdash \neg\neg A$ *then* $\Gamma \vdash A$;

D.8 Derivability and Consistency

We will now establish a number of properties of the derivability relation. They are independently interesting, but each will play a role in the proof of the completeness theorem.

Proposition D.22. *If* $\Gamma \vdash A$ *and* $\Gamma \cup \{A\}$ *is inconsistent, then* Γ *is inconsistent.*

Proof. If $\Gamma \cup \{A\}$ is inconsistent, then $\Gamma \cup \{A\} \vdash \bot$. By Proposition D.14, $\Gamma \vdash B$ for every $B \in \Gamma$. Since also $\Gamma \vdash A$ by hypothesis, $\Gamma \vdash B$ for every $B \in \Gamma \cup \{A\}$. By Proposition D.16, $\Gamma \vdash \bot$, i.e., Γ is inconsistent. □

Proposition D.23. $\Gamma \vdash A$ *iff* $\Gamma \cup \{\neg A\}$ *is inconsistent.*

Proof. First suppose $\Gamma \vdash A$. Then $\Gamma \cup \{\neg A\} \vdash A$ by Proposition D.15. $\Gamma \cup \{\neg A\} \vdash \neg A$ by Proposition D.14. We also have $\vdash \neg A \to (A \to \bot)$ by eq. (D.10). So by two applications of Proposition D.19, we have $\Gamma \cup \{\neg A\} \vdash \bot$.

Now assume $\Gamma \cup \{\neg A\}$ is inconsistent, i.e., $\Gamma \cup \{\neg A\} \vdash \bot$. By the deduction theorem, $\Gamma \vdash \neg A \to \bot$. $\Gamma \vdash (\neg A \to \bot) \to \neg\neg A$ by

eq. (D.13), so $\Gamma \vdash \neg\neg A$ by Proposition D.19. Since $\Gamma \vdash \neg\neg A \to A$ (eq. (D.14)), we have $\Gamma \vdash A$ by Proposition D.19 again. □

Proposition D.24. *If $\Gamma \vdash A$ and $\neg A \in \Gamma$, then Γ is inconsistent.*

Proof. $\Gamma \vdash \neg A \to (A \to \bot)$ by eq. (D.10). $\Gamma \vdash \bot$ by two applications of Proposition D.19. □

Proposition D.25. *If $\Gamma \cup \{A\}$ and $\Gamma \cup \{\neg A\}$ are both inconsistent, then Γ is inconsistent.*

Proof. Exercise. □

D.9 Derivability and the Propositional Connectives

Proposition D.26. *1. Both $A \wedge B \vdash A$ and $A \wedge B \vdash B$*

2. $A, B \vdash A \wedge B$.

Proof. 1. From eq. (D.1) and eq. (D.1) by modus ponens.

2. From eq. (D.3) by two applications of modus ponens. □

Proposition D.27. *1. $A \vee B, \neg A, \neg B$ is inconsistent.*

2. Both $A \vdash A \vee B$ and $B \vdash A \vee B$.

Proof. 1. From eq. (D.9) we get $\vdash \neg A \to (A \to \bot)$ and $\vdash \neg A \to (A \to \bot)$. So by the deduction theorem, we have $\{\neg A\} \vdash A \to \bot$ and $\{\neg B\} \vdash B \to \bot$. From eq. (D.6) we get $\{\neg A, \neg B\} \vdash (A \vee B) \to \bot$. By the deduction theorem, $\{A \vee B, \neg A, \neg B\} \vdash \bot$.

2. From eq. (D.4) and eq. (D.5) by modus ponsens. □

Proposition D.28. *1.* $A, A \to B \vdash B$.

2. Both $\neg A \vdash A \to B$ *and* $B \vdash A \to B$.

Proof. 1. We can derive:

1.	A	Hyp
2.	$A \to B$	Hyp
3.	B	1, 2, MP

2. By eq. (D.10) and eq. (D.7) and the deduction theorem, respectively. □

D.10 Soundness

A derivation system, such as axiomatic deduction, is *sound* if it cannot derive things that do not actually hold. Soundness is thus a kind of guaranteed safety property for derivation systems. Depending on which proof theoretic property is in question, we would like to know for instance, that

1. every derivable A is valid;
2. if A is derivable from some others Γ, it is also a consequence of them;
3. if a set of formulas Γ is inconsistent, it is unsatisfiable.

These are important properties of a derivation system. If any of them do not hold, the derivation system is deficient—it would derive too much. Consequently, establishing the soundness of a derivation system is of the utmost importance.

Proposition D.29. *If A is an axiom, then $v \vDash A$ for each valuation v.*

Proof. Do truth tables for each axiom to verify that they are tautologies. □

Theorem D.30 (Soundness). *If $\Gamma \vdash A$ then $\Gamma \vDash A$.*

Proof. By induction on the length of the derivation of A from Γ. If there are no steps justified by inferences, then all formulas in the derivation are either instances of axioms or are in Γ. By the previous proposition, all the axioms are tautologies, and hence if A is an axiom then $\Gamma \vDash A$. If $A \in \Gamma$, then trivially $\Gamma \vDash A$.

If the last step of the derivation of A is justified by modus ponens, then there are formulas B and $B \to A$ in the derivation, and the induction hypothesis applies to the part of the derivation ending in those formulas (since they contain at least one fewer steps justified by an inference). So, by induction hypothesis, $\Gamma \vDash B$ and $\Gamma \vDash B \to A$. Then $\Gamma \vDash A$ by Theorem C.17.

Corollary D.31. *If $\vdash A$, then A is a tautology.*

Corollary D.32. *If Γ is satisfiable, then it is consistent.*

Proof. We prove the contrapositive. Suppose that Γ is not consistent. Then $\Gamma \vdash \bot$, i.e., there is a derivation of $\bot$ from Γ. By Theorem D.30, any valuation v that satisfies Γ must satisfy $\bot$. Since $v \nvDash \bot$ for every valuation v, no v can satisfy Γ, i.e., Γ is not satisfiable. □

Problems

Problem D.1. Show that the following hold by exhibiting derivations from the axioms:

1. $(A \wedge B) \to (B \wedge A)$
2. $((A \wedge B) \to C) \to (A \to (B \to C))$
3. $\neg(A \vee B) \to \neg A$

Problem D.2. Prove Proposition D.17.

Problem D.3. Prove Proposition D.21

Problem D.4. Prove that $\Gamma \vdash \neg A$ iff $\Gamma \cup \{A\}$ is inconsistent.

Problem D.5. Prove Proposition D.25

APPENDIX E

Tableaux

E.1 Tableaux

While many derivation systems operate with arrangements of sentences, tableaux operate with signed formulas. A signed formula is a pair consisting of a truth value sign ($\mathbb{T}$ or $\mathbb{F}$) and a sentence

$$\mathbb{T}\,A \text{ or } \mathbb{F}\,A.$$

A tableau consists of signed formulas arranged in a downward-branching tree. It begins with a number of *assumptions* and continues with signed formulas which result from one of the signed formulas above it by applying one of the rules of inference. Each rule allows us to add one or more signed formulas to the end of a branch, or two signed formulas side by side—in this case a branch splits into two, with the two added signed formulas forming the ends of the two branches.

A rule applied to a complex signed formula results in the addition of signed formulas which are immediate sub-formulas. They come in pairs, one rule for each of the two signs. For instance, the $\wedge\mathbb{T}$ rule applies to $\mathbb{T}\,A \wedge B$, and allows the addition of both the two signed formulas $\mathbb{T}\,A$ and $\mathbb{T}\,B$ to the end of any branch containing $\mathbb{T}\,A \wedge B$, and the rule $A \wedge B\mathbb{F}$ allows a branch to be split by adding $\mathbb{F}\,A$ and $\mathbb{F}\,B$ side-by-side. A tableau is closed if every one of its branches contains a matching pair of signed formulas $\mathbb{T}\,A$ and $\mathbb{F}\,A$.

The $\vdash$ relation based on tableaux is defined as follows: $\Gamma \vdash A$ iff there is some finite set $\Gamma_0 = \{B_1, \ldots, B_n\} \subseteq \Gamma$ such that there is a closed tableau for the assumptions

$$\{\mathbb{F}\, A, \mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n\}$$

For instance, here is a closed tableau that shows that $\vdash (A \wedge B) \rightarrow A$:

1.	$\mathbb{F}\,(A \wedge B) \rightarrow A$	Assumption
2.	$\mathbb{T}\, A \wedge B$	$\rightarrow\mathbb{F}\,1$
3.	$\mathbb{F}\, A$	$\rightarrow\mathbb{F}\,1$
4.	$\mathbb{T}\, A$	$\rightarrow\mathbb{T}\,2$
5.	$\mathbb{T}\, B$	$\rightarrow\mathbb{T}\,2$
	$\otimes$	

A set Γ is inconsistent in the tableau calculus if there is a closed tableau for assumptions

$$\{\mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n\}$$

for some $B_i \in \Gamma$.

Tableaux were invented in the 1950s independently by Evert Beth and Jaakko Hintikka, and simplified and popularized by Raymond Smullyan. They are very easy to use, since constructing a tableau is a very systematic procedure. Because of the systematic nature of tableaux, they also lend themselves to implementation by computer. However, a tableau is often hard to read and their connection to proofs are sometimes not easy to see. The approach is also quite general, and many different logics have tableau systems. Tableaux also help us to find structures that satisfy given (sets of) sentences: if the set is satisfiable, it won't have a closed tableau, i.e., any tableau will have an open branch. The satisfying structure can be "read off" an open branch, provided every rule it is possible to apply has been applied on that branch. There is also a very close connection to the sequent calculus: essentially, a closed tableau is a condensed derivation in the sequent calculus, written upside-down.

E.2 Rules and Tableaux

A tableau is a systematic survey of the possible ways a sentence can be true or false in a structure. The bulding blocks of a tableau are signed formulas: sentences plus a truth value "sign," either $\mathbb{T}$ or $\mathbb{F}$. These signed formulas are arranged in a (downward growing) tree.

Definition E.1. A *signed formula* is a pair consisting of a truth value and a sentence, i.e., either:

$$\mathbb{T}\,A \text{ or } \mathbb{F}\,A.$$

Intuitively, we might read $\mathbb{T}\,A$ as "A might be true" and $\mathbb{F}\,A$ as "A might be false" (in some structure).

Each signed formula in the tree is either an *assumption* (which are listed at the very top of the tree), or it is obtained from a signed formula above it by one of a number of rules of inference. There are two rules for each possible main operator of the preceding formula, one for the case when the sign is $\mathbb{T}$, and one for the case where the sign is $\mathbb{F}$. Some rules allow the tree to branch, and some only add signed formulas to the branch. A rule may be (and often must be) applied not to the immediately preceding signed formula, but to any signed formula in the branch from the root to the place the rule is applied.

A branch is *closed* when it contains both $\mathbb{T}\,A$ and $\mathbb{F}\,A$. A closed tableau is one where every branch is closed. Under the intuitive interpretation, any branch describes a joint possibility, but $\mathbb{T}\,A$ and $\mathbb{F}\,A$ are not jointly possible. In other words, if a branch is closed, the possibility it describes has been ruled out. In particular, that means that a closed tableau rules out all possibilities of simultaneously making every assumption of the form $\mathbb{T}\,A$ true and every assumption of the form $\mathbb{F}\,A$ false.

A closed tableau *for* A is a closed tableau with root $\mathbb{F}\,A$. If such a closed tableau exists, all possibilities for A being false have been ruled out; i.e., A must be true in every structure.

E.3 Propositional Rules

Rules for ¬

$$\frac{\mathbb{T}\,\neg A}{\mathbb{F}\,A}\ \neg\mathbb{T} \qquad\qquad \frac{\mathbb{F}\,\neg A}{\mathbb{T}\,A}\ \neg\mathbb{F}$$

Rules for ∧

$$\frac{\mathbb{T}\,A \wedge B}{\begin{array}{l}\mathbb{T}\,A\\ \mathbb{T}\,B\end{array}}\ \wedge\mathbb{T} \qquad\qquad \frac{\mathbb{F}\,A \wedge B}{\mathbb{F}\,A \mid \mathbb{F}\,B}\ \wedge\mathbb{F}$$

Rules for ∨

$$\frac{\mathbb{T}\,A \vee B}{\mathbb{T}\,A \mid \mathbb{T}\,B}\ \vee\mathbb{T} \qquad\qquad \frac{\mathbb{F}\,A \vee B}{\begin{array}{l}\mathbb{F}\,A\\ \mathbb{F}\,B\end{array}}\ \vee\mathbb{F}$$

Rules for →

$$\frac{\mathbb{T}\,A \to B}{\mathbb{F}\,A \mid \mathbb{T}\,B}\ \to\mathbb{T} \qquad\qquad \frac{\mathbb{F}\,A \to B}{\begin{array}{l}\mathbb{T}\,A\\ \mathbb{F}\,B\end{array}}\ \to\mathbb{F}$$

The Cut Rule

$$\frac{}{\mathbb{T}\,A \mid \mathbb{F}\,A}\ \text{Cut}$$

The Cut rule is not applied "to" a previous signed formula; rather, it allows every branch in a tableau to be split in two, one branch containing $\mathbb{T}\,A$, the other $\mathbb{F}\,A$. It is not necessary—any set of signed formulas with a closed tableau has one not using Cut—but it allows us to combine tableaux in a convenient way.

E.4 Tableaux

We've said what an assumption is, and we've given the rules of inference. Tableaux are inductively generated from these: each tableau either is a single branch consisting of one or more assumptions, or it results from a tableau by applying one of the rules of inference on a branch.

Definition E.2 (Tableau). A tableau for assumptions $S_1A_1, \ldots, S_nA_n$ (where each S_i is either $\mathbb{T}$ or $\mathbb{F}$) is a tree of signed formulas satisfying the following conditions:

1. The n topmost signed formulas of the tree are S_iA_i, one below the other.
2. Every signed formula in the tree that is not one of the assumptions results from a correct application of an inference rule to a signed formula in the branch above it.

A branch of a tableau is *closed* iff it contains both $\mathbb{T}\,A$ and $\mathbb{F}\,A$, and *open* otherwise. A tableau in which every branch is closed is a *closed tableau* (for its set of assumptions). If a tableau is not closed, i.e., if it contains at least one open branch, it is *open*.

Example E.3. Every set of assumptions on its own is a tableau, but it will generally not be closed. (Obviously, it is closed only if the assumptions already contain a pair of signed formulas $\mathbb{T}\,A$ and $\mathbb{F}\,A$.)

From a tableau (open or closed) we can obtain a new, larger one by applying one of the rules of inference to a signed formula A in it. The rule will append one or more signed formulas to the end of any branch containing the occurrence of A to which we apply the rule.

For instance, consider the assumption $\mathbb{T}\,A \wedge \neg A$. Here is the (open) tableau consisting of just that assumption:

1. $\mathbb{T}\,A \wedge \neg A$ Assumption

We obtain a new tableau from it by applying the $\wedge\mathbb{T}$ rule to the assumption. That rule allows us to add two new lines to the tableau, $\mathbb{T}\,A$ and $\mathbb{T}\,\neg A$:

1.	$\mathbb{T}\,A \wedge \neg A$	Assumption
2.	$\mathbb{T}\,A$	$\wedge\mathbb{T}\,1$
3.	$\mathbb{T}\,\neg A$	$\wedge\mathbb{T}\,1$

When we write down tableaux, we record the rules we've applied on the right (e.g., $\wedge\mathbb{T}1$ means that the signed formula on that line is the result of applying the $\wedge\mathbb{T}$ rule to the signed formula on line 1). This new tableau now contains additional signed formulas, but to only one ($\mathbb{T}\,\neg A$) can we apply a rule (in this case, the $\neg\mathbb{T}$ rule). This results in the closed tableau

1.	$\mathbb{T}\,A \wedge \neg A$	Assumption
2.	$\mathbb{T}\,A$	$\wedge\mathbb{T}\,1$
3.	$\mathbb{T}\,\neg A$	$\wedge\mathbb{T}\,1$
4.	$\mathbb{F}\,A$	$\neg\mathbb{T}\,3$
	$\otimes$	

E.5 Examples of Tableaux

Example E.4. Let's find a closed tableau for the sentence $(A \wedge B) \to A$.

We begin by writing the corresponding assumption at the top of the tableau.

1.	$\mathbb{F}\,(A \wedge B) \to A$	Assumption

There is only one assumption, so only one signed formula to which we can apply a rule. (For every signed formula, there is always at most one rule that can be applied: it's the rule for the corresponding sign and main operator of the sentence.) In this case, this means, we must apply $\to\mathbb{F}$.

1.	$\mathbb{F}\,(A \wedge B) \to A$ ✓	Assumption
2.	$\mathbb{T}\,A \wedge B$	$\to\mathbb{F}\,1$
3.	$\mathbb{F}\,A$	$\to\mathbb{F}\,1$

To keep track of which signed formulas we have applied their corresponding rules to, we write a checkmark next to the sentence. However, *only* write a checkmark if the rule has been applied to all open branches. Once a signed formula has had the corresponding rule applied in every open branch, we will not have to return to it and apply the rule again. In this case, there is only one branch, so the rule only has to be applied once. (Note that checkmarks are only a convenience for constructing tableaux and are not officially part of the syntax of tableaux.)

There is one new signed formula to which we can apply a rule: the $\mathbb{T}\,A \wedge B$ on line 3. Applying the $\wedge\mathbb{T}$ rule results in:

1.	$\mathbb{F}\,(A \wedge B) \to A$ ✓	Assumption
2.	$\mathbb{T}\,A \wedge B$ ✓	$\to\mathbb{F}\,1$
3.	$\mathbb{F}\,A$	$\to\mathbb{F}\,1$
4.	$\mathbb{T}\,A$	$\wedge\mathbb{T}\,2$
5.	$\mathbb{T}\,B$	$\wedge\mathbb{T}\,2$
	$\otimes$	

Since the branch now contains both $\mathbb{T}\,A$ (on line 4) and $\mathbb{F}\,A$ (on line 3), the branch is closed. Since it is the only branch, the tableau is closed. We have found a closed tableau for $(A \wedge B) \to A$.

Example E.5. Now let's find a closed tableau for $(\neg A \vee B) \to (A \to B)$.

We begin with the corresponding assumption:

1.	$\mathbb{F}\,(\neg A \vee B) \to (A \to B)$	Assumption

The one signed formula in this tableau has main operator $\to$ and sign $\mathbb{F}$, so we apply the $\to\mathbb{F}$ rule to it to obtain:

1.	$\mathbb{F}\,(\neg A \vee B) \to (A \to B)$ ✓	Assumption
2.	$\mathbb{T}\,\neg A \vee B$	$\to\mathbb{F}\,1$
3.	$\mathbb{F}\,(A \to B)$	$\to\mathbb{F}\,1$

We now have a choice as to whether to apply $\vee\mathbb{T}$ to line 2 or $\to\mathbb{F}$ to line 3. It actually doesn't matter which order we pick, as long as each signed formula has its corresponding rule applied in every branch. So let's pick the first one. The $\vee\mathbb{T}$ rule allows the tableau to branch, and the two conclusions of the rule will be the new signed formulas added to the two new branches. This results in:

1.	$\mathbb{F}\,(\neg A \vee B) \to (A \to B)$ ✓	Assumption
2.	$\mathbb{T}\,\neg A \vee B$ ✓	$\to\mathbb{F}\,1$
3.	$\mathbb{F}\,(A \to B)$	$\to\mathbb{F}\,1$
4.	$\mathbb{T}\,\neg A$ | $\mathbb{T}\,B$	$\vee\mathbb{T}\,2$

We have not applied the $\to\mathbb{F}$ rule to line 3 yet: let's do that now. To save time, we apply it to both branches. Recall that we write a checkmark next to a signed formula only if we have applied the corresponding rule in every open branch. So it's a good idea to apply a rule at the end of every branch that contains the signed formula the rule applies to. That way we won't have to return to that signed formula lower down in the various branches.

1.	$\mathbb{F}\,(\neg A \vee B) \to (A \to B)$ ✓	Assumption
2.	$\mathbb{T}\,\neg A \vee B$ ✓	$\to\mathbb{F}\,1$
3.	$\mathbb{F}\,(A \to B)$ ✓	$\to\mathbb{F}\,1$
4.	$\mathbb{T}\,\neg A$ | $\mathbb{T}\,B$	$\vee\mathbb{T}\,2$
5.	$\mathbb{T}\,A$ | $\mathbb{T}\,A$	$\to\mathbb{F}\,3$
6.	$\mathbb{F}\,B$ | $\mathbb{F}\,B$	$\to\mathbb{F}\,3$
	$\otimes$	

The right branch is now closed. On the left branch, we can still apply the $\neg\mathbb{T}$ rule to line 4. This results in $\mathbb{F}\,A$ and closes the left branch:

1.	$\mathbb{F}\,(\neg A \vee B) \to (A \to B)$ ✓		Assumption
2.	$\mathbb{T}\,\neg A \vee B$ ✓		$\to\mathbb{F}\,1$
3.	$\mathbb{F}\,(A \to B)$ ✓		$\to\mathbb{F}\,1$
4.	$\mathbb{T}\,\neg A$	$\mathbb{T}\,B$	$\vee\mathbb{T}\,2$
5.	$\mathbb{T}\,A$	$\mathbb{T}\,A$	$\to\mathbb{F}\,3$
6.	$\mathbb{F}\,B$	$\mathbb{F}\,B$	$\to\mathbb{F}\,3$
7.	$\mathbb{F}\,A$	$\otimes$	$\neg\mathbb{T}\,4$
	$\otimes$		

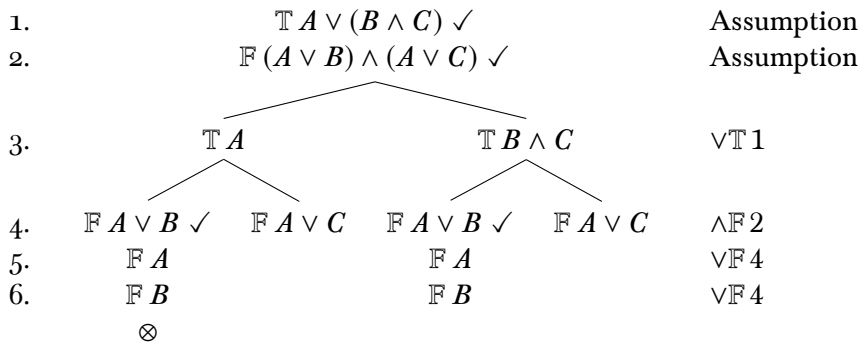
Example E.6. We can give tableaux for any number of signed formulas as assumptions. Often it is also necessary to apply more than one rule that allows branching; and in general a tableau can have any number of branches. For instance, consider a tableau for $\{\mathbb{T}\,A \vee (B \wedge C), \mathbb{F}\,(A \vee B) \wedge (A \vee C)\}$. We start by applying the $\vee\mathbb{T}$ to the first assumption:

1.	$\mathbb{T}\,A \vee (B \wedge C)$ ✓		Assumption
2.	$\mathbb{F}\,(A \vee B) \wedge (A \vee C)$		Assumption
3.	$\mathbb{T}\,A$	$\mathbb{T}\,B \wedge C$	$\vee\mathbb{T}\,1$

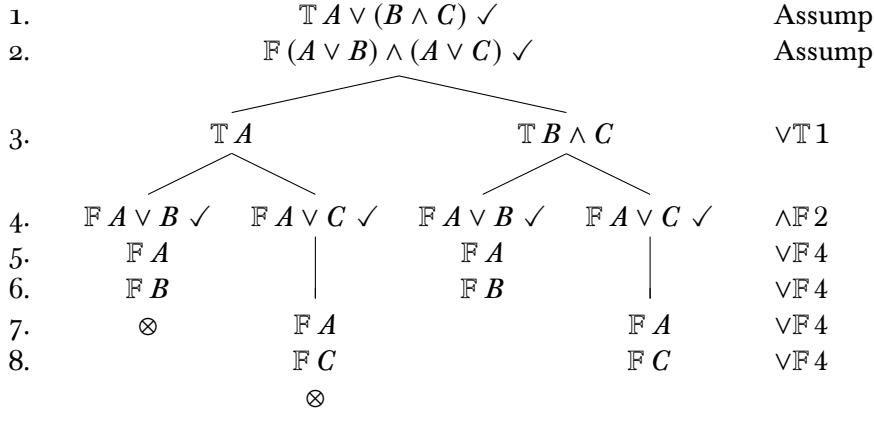
Now we can apply the $\wedge\mathbb{F}$ rule to line 2. We do this on both branches simultaneously, and can therefore check off line 2:

1.	$\mathbb{T}\,A \vee (B \wedge C)$ ✓				Assumption
2.	$\mathbb{F}\,(A \vee B) \wedge (A \vee C)$ ✓				Assumption
3.	$\mathbb{T}\,A$		$\mathbb{T}\,B \wedge C$		$\vee\mathbb{T}\,1$
4.	$\mathbb{F}\,A \vee B$	$\mathbb{F}\,A \vee C$	$\mathbb{F}\,A \vee B$	$\mathbb{F}\,A \vee C$	$\wedge\mathbb{F}\,2$

Now we can apply $\vee\mathbb{F}$ to all the branches containing $A \vee B$:

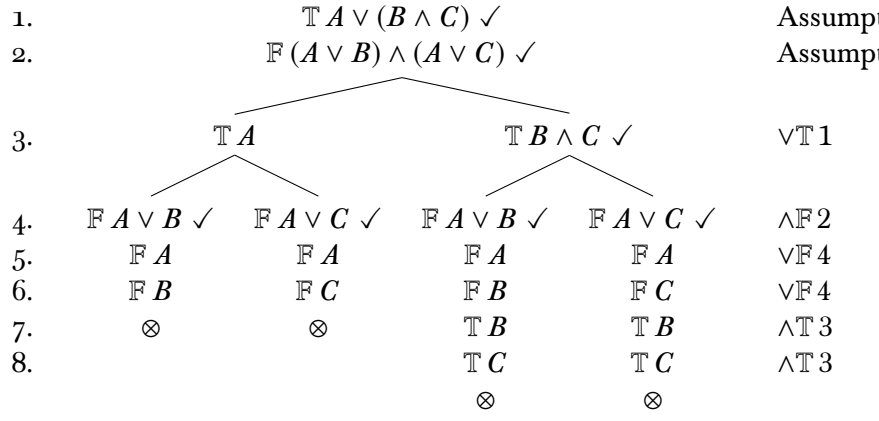


The leftmost branch is now closed. Let's now apply $\vee\mathbb{F}$ to $A \vee C$:

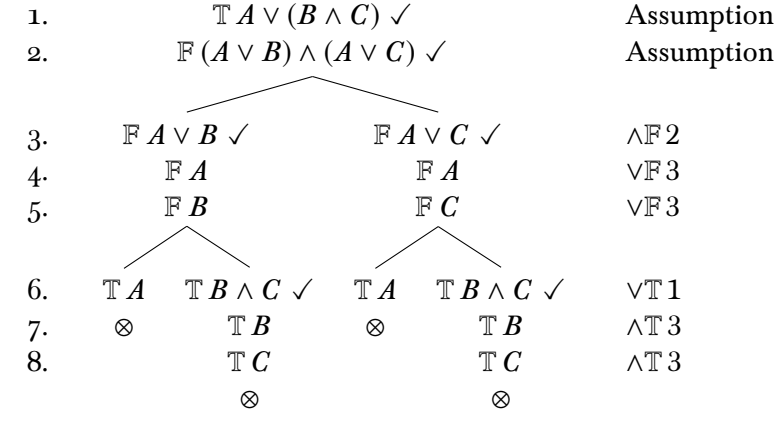


Note that we moved the result of applying $\vee\mathbb{F}$ a second time below for clarity. In this instance it would not have been needed, since the justifications would have been the same.

Two branches remain open, and $\mathbb{T}\, B \wedge C$ on line 3 remains unchecked. We apply $\wedge\mathbb{T}$ to it to obtain a closed tableau:



For comparison, here's a closed tableau for the same set of assumptions in which the rules are applied in a different order:



E.6 Proof-Theoretic Notions

Just as we've defined a number of important semantic notions (validity, entailment, satisfiabilty), we now define corresponding *proof-theoretic notions*. These are not defined by appeal to satisfaction of sentences in structures, but by appeal to the existence of

certain closed tableaux. It was an important discovery that these notions coincide. That they do is the content of the *soundness* and *completeness theorems.*

Definition E.7 (Theorems). A sentence A is a *theorem* if there is a closed tableau for $\mathbb{F}\, A$. We write $\vdash A$ if A is a theorem and $\nvdash A$ if it is not.

Definition E.8 (Derivability). A sentence A is *derivable from* a set of sentences Γ, $\Gamma \vdash A$, iff there is a finite set $\{B_1, \ldots, B_n\} \subseteq \Gamma$ and a closed tableau for the set

$$\{\mathbb{F}\, A, \mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n, \}$$

If A is not derivable from Γ we write $\Gamma \nvdash A$.

Definition E.9 (Consistency). A set of sentences Γ is *inconsistent* iff there is a finite set $\{B_1, \ldots, B_n\} \subseteq \Gamma$ and a closed tableau for the set

$$\{\mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n, \}.$$

If Γ is not inconsistent, we say it is *consistent.*

Proposition E.10 (Reflexivity). *If $A \in \Gamma$, then $\Gamma \vdash A$.*

Proof. If $A \in \Gamma$, $\{A\}$ is a finite subset of Γ and the tableau

1.	$\mathbb{F}\, A$	Assumption
2.	$\mathbb{T}\, A$	Assumption
	$\otimes$	

is closed. □

Proposition E.11 (Monotony). *If $\Gamma \subseteq \Delta$ and $\Gamma \vdash A$, then $\Delta \vdash A$.*

Proof. Any finite subset of Γ is also a finite subset of Δ. □

Proposition E.12 (Transitivity). *If $\Gamma \vdash A$ and $\{A\} \cup \Delta \vdash B$, then $\Gamma \cup \Delta \vdash B$.*

Proof. If $\{A\} \cup \Delta \vdash B$, then there is a finite subset $\Delta_0 = \{C_1, \ldots, C_n\} \subseteq \Delta$ such that

$$\{\mathbb{F}\, B, \mathbb{T}\, A, \mathbb{T}\, C_1, \ldots, \mathbb{T}\, C_n\}$$

has a closed tableau. If $\Gamma \vdash A$ then there are D_1, …, D_m such that

$$\{\mathbb{F}\, A, \mathbb{T}\, D_1, \ldots, \mathbb{T}\, D_m\}$$

has a closed tableau.

Now consider the tableau with assumptions

$$\mathbb{F}\, B, \mathbb{T}\, C_1, \ldots, \mathbb{T}\, C_n, \mathbb{T}\, D_1, \ldots, \mathbb{T}\, D_m.$$

Apply the Cut rule on A. This generates two branches, one has $\mathbb{T}\, A$ in it, the other $\mathbb{F}\, A$. Thus, on the one branch, all of

$$\{\mathbb{F}\, B, \mathbb{T}\, A, \mathbb{T}\, C_1, \ldots, \mathbb{T}\, C_n\}$$

are available. Since there is a closed tableau for these assumptions, we can attach it to that branch; every branch through $\mathbb{T}\, A_1$ closes. On the other branch, all of

$$\{\mathbb{F}\, A, \mathbb{T}\, D_1, \ldots, \mathbb{T}\, D_m\}$$

are available, so we can also complete the other side to obtain a closed tableau. This shows $\Gamma \cup \Delta \vdash B$. □

Note that this means that in particular if $\Gamma \vdash A$ and $A \vdash B$, then $\Gamma \vdash B$. It follows also that if $A_1, \ldots, A_n \vdash B$ and $\Gamma \vdash A_i$ for each i, then $\Gamma \vdash B$.

Proposition E.13. *Γ is inconsistent iff $\Gamma \vdash A$ for every sentence A.*

Proof. Exercise. □

Proposition E.14 (Compactness). 1. *If $\Gamma \vdash A$ then there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \vdash A$.*

2. *If every finite subset of Γ is consistent, then Γ is consistent.*

Proof. 1. If $\Gamma \vdash A$, then there is a finite subset $\Gamma_0 = \{B_1, \ldots, B_n\}$ and a closed tableau for

$$\mathbb{F}\, A, \mathbb{T}\, B_1, \cdots \mathbb{T}\, B_n$$

This tableau also shows $\Gamma_0 \vdash A$.

2. If Γ is inconsistent, then for some finite subset $\Gamma_0 = \{B_1, \ldots, B_n\}$ there is a closed tableau for

$$\mathbb{T}\, B_1, \cdots \mathbb{T}\, B_n$$

This closed tableau shows that Γ_0 is inconsistent. □

E.7 Derivability and Consistency

We will now establish a number of properties of the derivability relation. They are independently interesting, but each will play a role in the proof of the completeness theorem.

Proposition E.15. *If $\Gamma \vdash A$ and $\Gamma \cup \{A\}$ is inconsistent, then Γ is inconsistent.*

Proof. There are finite $\Gamma_0 = \{B_1, \ldots, B_n\}$ and $\Gamma_1 = \{C_1, \ldots, C_n\} \subseteq \Gamma$ such that

$$\{\mathbb{F}\, A, \mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n\}$$
$$\{\mathbb{T}\, \neg A, \mathbb{T}\, C_1, \ldots, \mathbb{T}\, C_m\}$$

have closed tableaux. Using the Cut rule on A we can combine these into a single closed tableau that shows $\Gamma_0 \cup \Gamma_1$ is inconsistent. Since $\Gamma_0 \subseteq \Gamma$ and $\Gamma_1 \subseteq \Gamma$, $\Gamma_0 \cup \Gamma_1 \subseteq \Gamma$, hence Γ is inconsistent. □

Proposition E.16. *$\Gamma \vdash A$ iff $\Gamma \cup \{\neg A\}$ is inconsistent.*

Proof. First suppose $\Gamma \vdash A$, i.e., there is a closed tableau for

$$\{\mathbb{F}\, A, \mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n\}$$

Using the $\neg\mathbb{T}$ rule, this can be turned into a closed tableau for

$$\{\mathbb{T}\, \neg A, \mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n\}.$$

On the other hand, if there is a closed tableau for the latter, we can turn it into a closed tableau of the former by removing every formula that results from $\neg\mathbb{T}$ applied to the first assumption $\mathbb{T}\, \neg A$ as well as that assumption, and adding the assumption $\mathbb{F}\, A$. For if a branch was closed before because it contained the conclusion of $\neg\mathbb{T}$ applied to $\mathbb{T}\, \neg A$, i.e., $\mathbb{F}\, A$, the corresponding branch in the new tableau is also closed. If a branch in the old tableau was closed because it contained the assumption $\mathbb{T}\, \neg A$ as well as $\mathbb{F}\, \neg A$ we can turn it into a closed branch by applying $\neg\mathbb{F}$ to $\mathbb{F}\, \neg A$ to obtain $\mathbb{T}\, A$. This closes the branch since we added $\mathbb{F}\, A$ as an assumption. □

Proposition E.17. *If $\Gamma \vdash A$ and $\neg A \in \Gamma$, then Γ is inconsistent.*

Proof. Suppose $\Gamma \vdash A$ and $\neg A \in \Gamma$. Then there are $B_1, \ldots, B_n \in \Gamma$ such that

$$\{\mathbb{F}\, A, \mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n\}$$

has a closed tableau. Replace the assumption $\mathbb{F}\, A$ by $\mathbb{T}\, \neg A$, and insert the conclusion of $\neg\mathbb{T}$ applied to $\mathbb{F}\, A$ after the assumptions. Any sentence in the tableau justified by appeal to line 1 in the old tableau is now justified by appeal to line $n + 1$. So if the old tableau was closed, the new one is. It shows that Γ is inconsistent, since all assumptions are in Γ. □

Proposition E.18. *If $\Gamma \cup \{A\}$ and $\Gamma \cup \{\neg A\}$ are both inconsistent, then Γ is inconsistent.*

Proof. If there are $B_1, \ldots, B_n \in \Gamma$ and $C_1, \ldots, C_m \in \Gamma$ such that

$$\{\mathbb{T}\, A, \mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n\}$$
$$\{\mathbb{T}\, \neg A, \mathbb{T}\, C_1, \ldots, \mathbb{T}\, C_m\}$$

both have closed tableaux, we can construct a tableau that shows that Γ is inconsistent by using as assumptions $\mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n$ together with $\mathbb{T}\, C_1, \ldots, \mathbb{T}\, C_m$, followed by an application of the Cut rule, yielding two branches, one starting with $\mathbb{T}\, A$, the other with $\mathbb{F}\, A$. Add on the part below the assumptions of the first tableau on the left side. Here, every rule application is still correct, and every branch closes. On the right side, add the part below the assumptions of the seond tableau, with the results of any applications of $\neg\mathbb{T}$ to $\mathbb{T}\, \neg A$ removed.

For if a branch was closed before because it contained the conclusion of $\neg\mathbb{T}$ applied to $\mathbb{T}\, \neg A$, i.e., $\mathbb{F}\, A$, as well as $\mathbb{F}\, A$, the corresponding branch in the new tableau is also closed. If a branch in the old tableau was closed because it contained the assumption $\mathbb{T}\, \neg A$ as well as $\mathbb{F}\, \neg A$ we can turn it into a closed branch by applying $\neg\mathbb{F}$ to $\mathbb{F}\, \neg A$ to obtain $\mathbb{T}\, A$. □

E.8 Derivability and the Propositional Connectives

Proposition E.19. *1. Both $A \wedge B \vdash A$ and $A \wedge B \vdash B$.*

2. $A, B \vdash A \wedge B$.

Proof. 1. Both $\{\mathbb{F}\,A, \mathbb{T}\,A \wedge B\}$ and $\{\mathbb{F}\,B, \mathbb{T}\,A \wedge B\}$ have closed tableaux

1.	$\mathbb{F}\,A$	Assumption
2.	$\mathbb{T}\,A \wedge B$	Assumption
3.	$\mathbb{T}A$	$\wedge\mathbb{T}\,2$
4.	$\mathbb{T}B$	$\wedge\mathbb{T}\,2$
	$\otimes$	

1.	$\mathbb{F}\,B$	Assumption
2.	$\mathbb{T}\,A \wedge B$	Assumption
3.	$\mathbb{T}A$	$\wedge\mathbb{T}\,2$
4.	$\mathbb{T}B$	$\wedge\mathbb{T}\,2$
	$\otimes$	

2. Here is a closed tableau for $\{\mathbb{T}\,A, \mathbb{T}\,B, \mathbb{F}\,A \wedge B\}$:

1.	$\mathbb{F}\,A \wedge B$	Assumption
2.	$\mathbb{T}\,A$	Assumption
3.	$\mathbb{T}\,B$	Assumption
4.	$\mathbb{F}\,A$ $\quad$ $\mathbb{F}\,B$	$\wedge\mathbb{F}\,1$
	$\otimes$ $\quad$ $\otimes$	

Proposition E.20. *1. $A \vee B, \neg A, \neg B$ is inconsistent.*

2. Both $A \vdash A \vee B$ and $B \vdash A \vee B$.

Proof. 1. We give a closed tableau of $\{\mathbb{T}\, A \vee B, \mathbb{T}\, \neg A, \mathbb{T}\, \neg B\}$:

1.	$\mathbb{T}\, A \vee B$	Assumption
2.	$\mathbb{T}\, \neg A$	Assumption
3.	$\mathbb{T}\, \neg B$	Assumption
4.	$\mathbb{F}\, A$	$\neg\mathbb{T}\, 2$
5.	$\mathbb{F}\, B$	$\neg\mathbb{T}\, 3$
6.	$\mathbb{T}\, A$ | $\mathbb{T}\, B$	$\vee\mathbb{T}\, 1$
	$\otimes$ $\otimes$	

2. Both $\{\mathbb{F}\, A \vee B, \mathbb{T}\, A\}$ and $\{\mathbb{F}\, A \vee B, \mathbb{T}\, B\}$ have closed tableaux:

1.	$\mathbb{F}\, A \vee B$	Assumption
2.	$\mathbb{T}\, A$	Assumption
3.	$\mathbb{F}A$	$\vee\mathbb{F}\, 1$
4.	$\mathbb{F}B$	$\vee\mathbb{F}\, 1$
	$\otimes$	

1.	$\mathbb{F}\, A \vee B$	Assumption
2.	$\mathbb{T}\, B$	Assumption
3.	$\mathbb{F}A$	$\vee\mathbb{F}\, 1$
4.	$\mathbb{F}B$	$\vee\mathbb{F}\, 1$
	$\otimes$	

Proposition E.21. *1.* $A, A \to B \vdash B$.

2. Both $\neg A \vdash A \to B$ *and* $B \vdash A \to B$.

Proof. 1. $\{\mathbb{F}\, B, \mathbb{T}\, A \to B, \mathbb{T}\, A\}$ has a closed tableau:

1.	$\mathbb{F}\, B$	Assumption
2.	$\mathbb{T}\, A \to B$	Assumption
3.	$\mathbb{T}\, A$	Assumption
4.	$\mathbb{F}\, A$ $\quad$ $\mathbb{T}\, B$	$\to\mathbb{T}\, 2$
	$\otimes$ $\quad$ $\otimes$	

2. Both s$\{\mathbb{F}\, A \to B, \mathbb{T}\, \neg A\}$ and $\{\mathbb{F}\, A \to B, \mathbb{T}\, \neg B\}$ have closed tableaux:

1.	$\mathbb{F}\, A \to B$	Assumption
2.	$\mathbb{T}\, \neg A$	Assumption
3.	$\mathbb{T}\, A$	$\to\mathbb{F}\, 1$
4.	$\mathbb{F}\, B$	$\to\mathbb{F}\, 1$
5.	$\mathbb{F}\, A$	$\neg\mathbb{T}\, 2$
	$\otimes$	

1.	$\mathbb{F}\, A \to B$	Assumption
2.	$\mathbb{T}\, \neg B$	Assumption
3.	$\mathbb{T}\, A$	$\to\mathbb{F}\, 1$
4.	$\mathbb{F}\, B$	$\to\mathbb{F}\, 1$
5.	$\mathbb{F}\, B$	$\neg\mathbb{T}\, 2$
	$\otimes$	

E.9 Soundness

A derivation system, such as tableaux, is *sound* if it cannot derive things that do not actually hold. Soundness is thus a kind of guaranteed safety property for derivation systems. Depending on which proof theoretic property is in question, we would like to know for instance, that

1. every derivable A is a tautology;
2. if a sentence is derivable from some others, it is also a consequence of them;
3. if a set of sentences is inconsistent, it is unsatisfiable.

These are important properties of a derivation system. If any of them do not hold, the derivation system is deficient—it would derive too much. Consequently, establishing the soundness of a derivation system is of the utmost importance.

Because all these proof-theoretic properties are defined via closed tableaux of some kind or other, proving (1)–(3) above requires proving something about the semantic properties of closed tableaux. We will first define what it means for a signed formula to be satisfied in a structure, and then show that if a tableau is closed, no structure satisfies all its assumptions. (1)–(3) then follow as corollaries from this result.

Definition E.22. A valuation v *satisfies* a signed formula $\mathbb{T}\,A$ iff $v \vDash A$, and it satisfies $\mathbb{F}\,A$ iff $v \nvDash A$. v satisfies a set of signed formulas Γ iff it satisfies every $S\,A \in \Gamma$. Γ is *satisfiable* if there is a valuation that satisfies it, and *unsatisfiable* otherwise.

Theorem E.23 (Soundness). *If Γ has a closed tableau, Γ is unsatisfiable.*

Proof. Let's call a branch of a tableau satisfiable iff the set of signed formulas on it is satisfiable, and let's call a tableau satisfiable if it contains at least one satisfiable branch.

We show the following: Extending a satisfiable tableau by one of the rules of inference always results in a satisfiable tableau. This will prove the theorem: any closed tableau results by applying rules of inference to the tableau consisting only of assumptions from Γ. So if Γ were satisfiable, any tableau for it would be satisfiable. A closed tableau, however, is clearly not satisfiable: every branch contains both $\mathbb{T}\,A$ and $\mathbb{F}\,A$, and no structure can both satisfy and not satisfy A.

Suppose we have a satisfiable tableau, i.e., a tableau with at least one satisfiable branch. Applying a rule of inference either adds signed formulas to a branch, or splits a branch in two. If the tableau has a satisfiable branch which is not extended by the rule application in question, it remains a satisfiable branch in the extended tableau, so the extended tableau is satisfiable. So we only have to consider the case where a rule is applied to a satisfiable branch.

Let Γ be the set of signed formulas on that branch, and let $S\,A \in \Gamma$ be the signed formula to which the rule is applied. If the rule does not result in a split branch, we have to show that the extended branch, i.e., Γ together with the conclusions of the rule, is still satisfiable. If the rule results in split branch, we have to show that at least one of the two resulting branches is satisfiable.

First, we consider the possible inferences with only one premise.

1. The branch is expanded by applying $\neg\mathbb{T}$ to $\mathbb{T}\,\neg B \in \Gamma$. Then the extended branch contains the signed formulas $\Gamma \cup \{\mathbb{F}\,B\}$. Suppose $v \vDash \Gamma$. In particular, $v \vDash \neg B$. Thus, $v \nvDash B$, i.e., v satisfies $\mathbb{F}\,B$.

2. The branch is expanded by applying $\neg\mathbb{F}$ to $\mathbb{F}\,\neg B \in \Gamma$: Exercise.

3. The branch is expanded by applying $\wedge\mathbb{T}$ to $\mathbb{T}\,B \wedge C \in \Gamma$, which results in two new signed formulas on the branch: $\mathbb{T}\,B$ and $\mathbb{T}\,C$. Suppose $v \vDash \Gamma$, in particular $v \vDash B \wedge C$. Then $v \vDash B$ and $v \vDash C$. This means that v satisfies both $\mathbb{T}\,B$ and $\mathbb{T}\,C$.

4. The branch is expanded by applying $\vee\mathbb{F}$ to $\mathbb{T}\,B \vee C \in \Gamma$: Exercise.

5. The branch is expanded by applying $\rightarrow\mathbb{F}$ to $\mathbb{T}\,B \rightarrow C \in \Gamma$: This results in two new signed formulas on the branch: $\mathbb{T}\,B$ and $\mathbb{F}\,C$. Suppose $v \vDash \Gamma$, in particular $v \nvDash B \rightarrow C$. Then $v \vDash B$ and $v \nvDash C$. This means that v satisfies both $\mathbb{T}\,B$ and $\mathbb{F}\,C$.

Now let's consider the possible inferences with two premises.

1. The branch is expanded by applying $\wedge\mathbb{F}$ to $\mathbb{F}\,B \wedge C \in \Gamma$, which results in two branches, a left one continuing through $\mathbb{F}\,B$ and a right one through $\mathbb{F}\,C$. Suppose $v \vDash \Gamma$, in particular $v \nvDash B \wedge C$. Then $v \nvDash B$ or $v \nvDash C$. In the former case, v satisfies $\mathbb{F}\,B$, i.e., v satisfies the formulas on the left branch. In the latter, v satisfies $\mathbb{F}\,C$, i.e., v satisfies the formulas on the right branch.

2. The branch is expanded by applying $\vee\mathbb{T}$ to $\mathbb{T}\,B \vee C \in \Gamma$: Exercise.

3. The branch is expanded by applying $\rightarrow\mathbb{T}$ to $\mathbb{T}\,B \rightarrow C \in \Gamma$: Exercise.

4. The branch is expanded by Cut: This results in two branches, one containing $\mathbb{T}\,B$, the other containing $\mathbb{F}\,B$. Since $v \vDash \Gamma$ and either $v \vDash B$ or $v \nvDash B$, v satisfies either the left or the right branch. □

Corollary E.24. *If $\vdash A$ then A is a tautology.*

Corollary E.25. *If $\Gamma \vdash A$ then $\Gamma \vDash A$.*

Proof. If $\Gamma \vdash A$ then for some $B_1, \ldots, B_n \in \Gamma$, $\{\mathbb{F}\, A, \mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n\}$ has a closed tableau. By Theorem E.23, every valuation v either makes some B_i false or makes A true. Hence, if $v \vDash \Gamma$ then also $v \vDash A$. □

Corollary E.26. *If Γ is satisfiable, then it is consistent.*

Proof. We prove the contrapositive. Suppose that Γ is not consistent. Then there are $B_1, \ldots, B_n \in \Gamma$ and a closed tableau for $\{\mathbb{T}\, B, \ldots, \mathbb{T}\, B\}$. By Theorem E.23, there is no v such that $v \vDash B_i$ for all $i = 1, \ldots, n$. But then Γ is not satisfiable. □

Problems

Problem E.1. Give closed tableaux of the following:

1. $\mathbb{F}\, \neg(A \to B) \to (A \wedge \neg B)$
2. $\mathbb{F}\, (A \to C) \vee (B \to C), \mathbb{T}\, (A \wedge B) \to C$

Problem E.2. Prove Proposition E.13

Problem E.3. Prove that $\Gamma \vdash \neg A$ iff $\Gamma \cup \{A\}$ is inconsistent.

Problem E.4. Complete the proof of Theorem E.23.

APPENDIX F

The Completeness Theorem

F.1 Introduction

The completeness theorem is one of the most fundamental results about logic. It comes in two formulations, the equivalence of which we'll prove. In its first formulation it says something fundamental about the relationship between semantic consequence and our proof system: if a sentence A follows from some sentences Γ, then there is also a derivation that establishes $\Gamma \vdash A$. Thus, the proof system is as strong as it can possibly be without proving things that don't actually follow.

In its second formulation, it can be stated as a model existence result: every consistent set of sentences is satisfiable. Consistency is a proof-theoretic notion: it says that our proof system is unable to produce certain derivations. But who's to say that just because there are no derivations of a certain sort from Γ, it's guaranteed that there is valuation v with $v \vDash \Gamma$? Before the completeness theorem was first proved—in fact before we had

the proof systems we now do—the great German mathematician David Hilbert held the view that consistency of mathematical theories guarantees the existence of the objects they are about. He put it as follows in a letter to Gottlob Frege:

> If the arbitrarily given axioms do not contradict one another with all their consequences, then they are true and the things defined by the axioms exist. This is for me the criterion of truth and existence.

Frege vehemently disagreed. The second formulation of the completeness theorem shows that Hilbert was right in at least the sense that if the axioms are consistent, then *some* valuation exists that makes them all true.

These aren't the only reasons the completeness theorem—or rather, its proof—is important. It has a number of important consequences, some of which we'll discuss separately. For instance, since any derivation that shows $\Gamma \vdash A$ is finite and so can only use finitely many of the sentences in Γ, it follows by the completeness theorem that if A is a consequence of Γ, it is already a consequence of a finite subset of Γ. This is called *compactness*. Equivalently, if every finite subset of Γ is consistent, then Γ itself must be consistent.

Although the compactness theorem follows from the completeness theorem via the detour through derivations, it is also possible to use the *the proof of* the completeness theorem to establish it directly. For what the proof does is take a set of sentences with a certain property—consistency—and constructs a structure out of this set that has certain properties (in this case, that it satisfies the set). Almost the very same construction can be used to directly establish compactness, by starting from "finitely satisfiable" sets of sentences instead of consistent ones.

F.2 Outline of the Proof

The proof of the completeness theorem is a bit complex, and upon first reading it, it is easy to get lost. So let us outline the proof. The first step is a shift of perspective, that allows us to see a route to a proof. When completeness is thought of as "whenever $\Gamma \models A$ then $\Gamma \vdash A$," it may be hard to even come up with an idea: for to show that $\Gamma \vdash A$ we have to find a derivation, and it does not look like the hypothesis that $\Gamma \models A$ helps us for this in any way. For some proof systems it is possible to directly construct a derivation, but we will take a slightly different approach. The shift in perspective required is this: completeness can also be formulated as: "if Γ is consistent, it is satisfiable." Perhaps we can use the information in Γ together with the hypothesis that it is consistent to construct a valuation that satisfies every formula in Γ. After all, we know what kind of valuation we are looking for: one that is as Γ describes it!

If Γ contains only propositional variables, it is easy to construct a model for it. All we have to do is come up with a valuation v such that $v \models p$ for all $p \in \Gamma$. Well, let $v(p) = \mathbb{T}$ iff $p \in \Gamma$.

Now suppose Γ contains some formula $\neg B$, with B atomic. We might worry that the construction of v interferes with the possibility of making $\neg B$ true. But here's where the consistency of Γ comes in: if $\neg B \in \Gamma$, then $B \notin \Gamma$, or else Γ would be inconsistent. And if $B \notin \Gamma$, then according to our construction of v, $v \not\models B$, so $v \models \neg B$. So far so good.

What if Γ contains complex, non-atomic formulas? Say it contains $A \wedge B$. To make that true, we should proceed as if both A and B were in Γ. And if $A \vee B \in \Gamma$, then we will have to make at least one of them true, i.e., proceed as if one of them was in Γ.

This suggests the following idea: we add additional formulas to Γ so as to (a) keep the resulting set consistent and (b) make sure that for every possible atomic sentence A, either A is in the resulting set, or $\neg A$ is, and (c) such that, whenever $A \wedge B$ is in the set, so are both A and B, if $A \vee B$ is in the set, at least one of

A or B is also, etc. We keep doing this (potentially forever). Call the set of all formulas so added Γ^*. Then our construction above would provide us with a valuation v for which we could prove, by induction, that it satisfies all sentences in Γ^*, and hence also all sentence in Γ since $\Gamma \subseteq \Gamma^*$. It turns out that guaranteeing (a) and (b) is enough. A set of sentences for which (b) holds is called *complete*. So our task will be to extend the consistent set Γ to a consistent and complete set Γ^*.

So here's what we'll do. First we investigate the properties of complete consistent sets, in particular we prove that a complete consistent set contains $A \wedge B$ iff it contains both A and B, $A \vee B$ iff it contains at least one of them, etc. (Proposition F.2). We'll then take the consistent set Γ and show that it can be extended to a consistent and complete set Γ^* (Lemma F.3). This set Γ^* is what we'll use to define our valuation $v(\Gamma^*)$. The valuation is determined by the propositional variables in Γ^* (Definition F.4). We'll use the properties of complete consistent sets to show that indeed $v(\Gamma^*) \vDash A$ iff $A \in \Gamma^*$ (Lemma F.5), and thus in particular, $v(\Gamma^*) \vDash \Gamma$.

F.3 Complete Consistent Sets of Sentences

Definition F.1 (Complete set). A set Γ of sentences is *complete* iff for any sentence A, either $A \in \Gamma$ or $\neg A \in \Gamma$.

Complete sets of sentences leave no questions unanswered. For any sentence A, Γ "says" if A is true or false. The importance of complete sets extends beyond the proof of the completeness theorem. A theory which is complete and axiomatizable, for instance, is always decidable.

Complete consistent sets are important in the completeness proof since we can guarantee that every consistent set of sentences Γ is contained in a complete consistent set Γ^*. A complete consistent set contains, for each sentence A, either A or its negation $\neg A$, but not both. This is true in particular for propositional

variables, so from a complete consistent set, we can construct a valuation where the truth value assigned to propositional variables is defined according to which propositional variables are in Γ^*. This valuation can then be shown to make all sentences in Γ^* (and hence also all those in Γ) true. The proof of this latter fact requires that $\neg A \in \Gamma^*$ iff $A \notin \Gamma^*$, $(A \vee B) \in \Gamma^*$ iff $A \in \Gamma^*$ or $B \in \Gamma^*$, etc.

In what follows, we will often tacitly use the properties of reflexivity, monotonicity, and transitivity of $\vdash$ (see appendices D.6 and E.6).

Proposition F.2. *Suppose Γ is complete and consistent. Then:*

1. *If $\Gamma \vdash A$, then $A \in \Gamma$.*
2. *$A \wedge B \in \Gamma$ iff both $A \in \Gamma$ and $B \in \Gamma$.*
3. *$A \vee B \in \Gamma$ iff either $A \in \Gamma$ or $B \in \Gamma$.*
4. *$A \to B \in \Gamma$ iff either $A \notin \Gamma$ or $B \in \Gamma$.*

Proof. Let us suppose for all of the following that Γ is complete and consistent.

1. If $\Gamma \vdash A$, then $A \in \Gamma$.

   Suppose that $\Gamma \vdash A$. Suppose to the contrary that $A \notin \Gamma$. Since Γ is complete, $\neg A \in \Gamma$. By Propositions E.17 and D.24, Γ is inconsistent. This contradicts the assumption that Γ is consistent. Hence, it cannot be the case that $A \notin \Gamma$, so $A \in \Gamma$.

2. Exercise.

3. First we show that if $A \vee B \in \Gamma$, then either $A \in \Gamma$ or $B \in \Gamma$. Suppose $A \vee B \in \Gamma$ but $A \notin \Gamma$ and $B \notin \Gamma$. Since Γ is complete, $\neg A \in \Gamma$ and $\neg B \in \Gamma$. By Propositions E.20 and D.27, item (1), Γ is inconsistent, a contradiction. Hence, either $A \in \Gamma$ or $B \in \Gamma$.

For the reverse direction, suppose that $A \in \Gamma$ or $B \in \Gamma$. By Propositions E.20 and D.27, item (2), $\Gamma \vdash A \vee B$. By (1), $A \vee B \in \Gamma$, as required.

4. Exercise. □

F.4 Lindenbaum's Lemma

We now prove a lemma that shows that any consistent set of sentences is contained in some set of sentences which is not just consistent, but also complete. The proof works by adding one sentence at a time, guaranteeing at each step that the set remains consistent. We do this so that for every A, either A or $\neg A$ gets added at some stage. The union of all stages in that construction then contains either A or its negation $\neg A$ and is thus complete. It is also consistent, since we made sure at each stage not to introduce an inconsistency.

Lemma F.3 (Lindenbaum's Lemma). *Every consistent set Γ in a language $\mathcal{L}$ can be extended to a complete and consistent set Γ^*.*

Proof. Let Γ be consistent. Let $A_0, A_1, \ldots$ be an enumeration of all the sentences of $\mathcal{L}$. Define $\Gamma_0 = \Gamma$, and

$$\Gamma_{n+1} = \begin{cases} \Gamma_n \cup \{A_n\} & \text{if } \Gamma_n \cup \{A_n\} \text{ is consistent;} \\ \Gamma_n \cup \{\neg A_n\} & \text{otherwise.} \end{cases}$$

Let $\Gamma^* = \bigcup_{n \geq 0} \Gamma_n$.

Each Γ_n is consistent: Γ_0 is consistent by definition. If $\Gamma_{n+1} = \Gamma_n \cup \{A_n\}$, this is because the latter is consistent. If it isn't, $\Gamma_{n+1} = \Gamma_n \cup \{\neg A_n\}$. We have to verify that $\Gamma_n \cup \{\neg A_n\}$ is consistent. Suppose it's not. Then *both* $\Gamma_n \cup \{A_n\}$ and $\Gamma_n \cup \{\neg A_n\}$ are inconsistent. This means that Γ_n would be inconsistent by Propositions E.17 and D.24, contrary to the induction hypothesis.

For every n and every $i < n$, $\Gamma_i \subseteq \Gamma_n$. This follows by a simple induction on n. For $n = 0$, there are no $i < 0$, so the claim holds automatically. For the inductive step, suppose it is true for n. We have $\Gamma_{n+1} = \Gamma_n \cup \{A_n\}$ or $= \Gamma_n \cup \{\neg A_n\}$ by construction. So $\Gamma_n \subseteq \Gamma_{n+1}$. If $i < n$, then $\Gamma_i \subseteq \Gamma_n$ by inductive hypothesis, and so $\subseteq \Gamma_{n+1}$ by transitivity of $\subseteq$.

From this it follows that every finite subset of Γ^* is a subset of Γ_n for some n, since each $B \in \Gamma^*$ not already in Γ_0 is added at some stage i. If n is the last one of these, then all B in the finite subset are in Γ_n. So, every finite subset of Γ^* is consistent. By Propositions E.14 and D.18, Γ^* is consistent.

Every sentence of $\mathrm{Frm}(\mathscr{L})$ appears on the list used to define Γ^*. If $A_n \notin \Gamma^*$, then that is because $\Gamma_n \cup \{A_n\}$ was inconsistent. But then $\neg A_n \in \Gamma^*$, so Γ^* is complete. □

F.5 Construction of a Model

We are now ready to define a valuation that makes all $A \in \Gamma$ true. To do this, we first apply Lindenbaum's Lemma: we get a complete consistent $\Gamma^* \supseteq \Gamma$. We let the propositional variables in Γ^* determine $v(\Gamma^*)$.

Definition F.4. Suppose Γ^* is a complete consistent set of formulas. Then we let

$$v(\Gamma^*)(p) = \begin{cases} \mathbb{T} & \text{if } p \in \Gamma^* \\ \mathbb{F} & \text{if } p \notin \Gamma^* \end{cases}$$

Lemma F.5 (Truth Lemma). *$v(\Gamma^*) \vDash A$ iff $A \in \Gamma^*$.*

Proof. We prove both directions simultaneously, and by induction on A.

1. $A \equiv \bot$: $v(\Gamma^*) \nvDash \bot$ by definition of satisfaction. On the other hand, $\bot \notin \Gamma^*$ since Γ^* is consistent.

2. $A \equiv p$: $v(\Gamma^*) \vDash p$ iff $v(\Gamma^*)(p) = \mathbb{T}$ (by the definition of satisfaction) iff $p \in \Gamma^*$ (by the construction of $v(\Gamma^*)$).

3. $A \equiv \neg B$: $v(\Gamma^*) \vDash A$ iff $v(\Gamma^*) \vDash B$ (by definition of satisfaction). By induction hypothesis, $v(\Gamma^*) \vDash B$ iff $B \notin \Gamma^*$. Since Γ^* is consistent and complete, $B \notin \Gamma^*$ iff $\neg B \in \Gamma^*$.

4. $A \equiv B \wedge C$: exercise.

5. $A \equiv B \vee C$: $v(\Gamma^*) \vDash A$ iff $v(\Gamma^*) \vDash B$ or $v(\Gamma^*) \vDash C$ (by definition of satisfaction) iff $B \in \Gamma^*$ or $C \in \Gamma^*$ (by induction hypothesis). This is the case iff $(B \vee C) \in \Gamma^*$ (by Proposition F.2(3)).

6. $A \equiv B \to C$: exercise.

F.6 The Completeness Theorem

Let's combine our results: we arrive at the completeness theorem.

Theorem F.6 (Completeness Theorem). *Let Γ be a set of sentences. If Γ is consistent, it is satisfiable.*

Proof. Suppose Γ is consistent. By Lemma F.3, there is a $\Gamma^* \supseteq \Gamma$ which is consistent and complete. By Lemma F.5, $v(\Gamma^*) \vDash A$ iff $A \in \Gamma^*$. From this it follows in particular that for all $A \in \Gamma$, $v(\Gamma^*) \vDash A$, so Γ is satisfiable. □

Corollary F.7 (Completeness Theorem, Second Version). *For all Γ and sentences A: if $\Gamma \vDash A$ then $\Gamma \vdash A$.*

Proof. Note that the Γ's in Corollary F.7 and Theorem F.6 are universally quantified. To make sure we do not confuse ourselves, let us restate Theorem F.6 using a different variable: for any set of sentences Δ, if Δ is consistent, it is satisfiable. By contraposition, if Δ is not satisfiable, then Δ is inconsistent. We will use this to prove the corollary.

Suppose that $\Gamma \vDash A$. Then $\Gamma \cup \{\neg A\}$ is unsatisfiable by Proposition C.16. Taking $\Gamma \cup \{\neg A\}$ as our Δ, the previous version of Theorem F.6 gives us that $\Gamma \cup \{\neg A\}$ is inconsistent. By Propositions E.16 and D.23, $\Gamma \vdash A$. □

Problems

Problem F.1. Complete the proof of Proposition F.2.

Problem F.2. Complete the proof of Lemma F.5.

Problem F.3. Use Corollary F.7 to prove Theorem F.6, thus showing that the two formulations of the completeness theorem are equivalent.

Problem F.4. In order for a derivation system to be complete, its rules must be strong enough to prove every unsatisfiable set inconsistent. Which of the rules of derivation were necessary to prove completeness? Are any of these rules not used anywhere in the proof? In order to answer these questions, make a list or diagram that shows which of the rules of derivation were used in which results that lead up to the proof of Theorem F.6. Be sure to note any tacit uses of rules in these proofs.

About the Open Logic Project

The *Open Logic Text* is an open-source, collaborative textbook of formal meta-logic and formal methods, starting at an intermediate level (i.e., after an introductory formal logic course). Though aimed at a non-mathematical audience (in particular, students of philosophy and computer science), it is rigorous.

Coverage of some topics currently included may not yet be complete, and many sections still require substantial revision. We plan to expand the text to cover more topics in the future. We also plan to add features to the text, such as a glossary, a list of further reading, historical notes, pictures, better explanations, sections explaining the relevance of results to philosophy, computer science, and mathematics, and more problems and examples. If you find an error, or have a suggestion, please let the project team know.

The project operates in the spirit of open source. Not only is the text freely available, we provide the LaTeX source under the Creative Commons Attribution license, which gives anyone the right to download, use, modify, re-arrange, convert, and re-distribute our work, as long as they give appropriate credit. Please see the Open Logic Project website at openlogicproject.org for additional information.