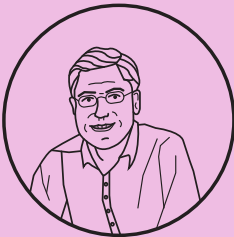


# Boxes and Diamonds

## An Open Introduction to Modal Logic



# **Boxes and Diamonds**

## *The Open Logic Project*

### **Instigator**

Richard Zach, *University of Calgary*

### **Editorial Board**

Aldo Antonelli,<sup>†</sup> *University of California, Davis*

Andrew Arana, *Université de Lorraine*

Jeremy Avigad, *Carnegie Mellon University*

Tim Button, *University College London*

Walter Dean, *University of Warwick*

Gillian Russell, *Dianoia Institute of Philosophy*

Nicole Wyatt, *University of Calgary*

Audrey Yap, *University of Victoria*

### **Contributors**

Samara Burns, *Columbia University*

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**.

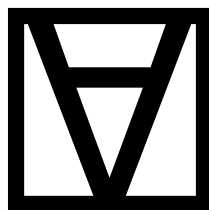
Typeset in Baskervald X and Nimbus Sans by L<sup>A</sup>T<sub>E</sub>X.

This version of *Boxes and Diamonds* is revision (None) ((None)), with content generated from *Open Logic Text* revision unknown (January 23, 2025). Free download at:

<https://bd.openlogicproject.org/>



*Boxes and Diamonds* by Richard Zach is licensed under a **Creative Commons Attribution 4.0 International License**. It is based on *The Open Logic Text* by the Open Logic Project, used under a **Creative Commons Attribution 4.0 International License**.



# *Contents*

<b>Preface</b>	<b>x</b>
<b>Introduction</b>	<b>xi</b>
<b>I Normal Modal Logics</b>	<b>1</b>
<b>1 Syntax and Semantics</b>	<b>2</b>
1.1 Introduction . . . . .	2
1.2 The Language of Basic Modal Logic . . . . .	4
1.3 Simultaneous Substitution . . . . .	5
1.4 Relational Models . . . . .	7
1.5 Truth at a World . . . . .	8
1.6 Truth in a Model . . . . .	10
1.7 Validity . . . . .	10
1.8 Tautological Instances . . . . .	11
1.9 Schemas and Validity . . . . .	14
1.10 Entailment . . . . .	16
Problems . . . . .	18
<b>2 Frame Definability</b>	<b>21</b>
2.1 Introduction . . . . .	21
2.2 Properties of Accessibility Relations . . . . .	22
2.3 Frames . . . . .	25
2.4 Frame Definability . . . . .	26

2.5	First-order Definability . . . . .	29
2.6	Equivalence Relations and <b>S5</b> . . . . .	30
2.7	Second-order Definability . . . . .	33
	Problems . . . . .	36
<b>3</b>	<b>Axiomatic Derivations</b>	<b>38</b>
3.1	Introduction . . . . .	38
3.2	Normal Modal Logics . . . . .	40
3.3	Derivations and Modal Systems . . . . .	42
3.4	Proofs in <b>K</b> . . . . .	44
3.5	Derived Rules . . . . .	46
3.6	More Proofs in <b>K</b> . . . . .	49
3.7	Dual Formulas . . . . .	50
3.8	Proofs in Modal Systems . . . . .	51
3.9	Soundness . . . . .	53
3.10	Showing Systems are Distinct . . . . .	54
3.11	Derivability from a Set of Formulas . . . . .	56
3.12	Properties of Derivability . . . . .	56
3.13	Consistency . . . . .	57
	Problems . . . . .	58
<b>4</b>	<b>Completeness and Canonical Models</b>	<b>60</b>
4.1	Introduction . . . . .	60
4.2	Complete $\Sigma$ -Consistent Sets . . . . .	62
4.3	Lindenbaum's Lemma . . . . .	63
4.4	Modalities and Complete Consistent Sets . . . . .	65
4.5	Canonical Models . . . . .	68
4.6	The Truth Lemma . . . . .	68
4.7	Determination and Completeness for <b>K</b> . . . . .	69
4.8	Frame Completeness . . . . .	70
	Problems . . . . .	74
<b>5</b>	<b>Filtrations and Decidability</b>	<b>75</b>
5.1	Introduction . . . . .	75
5.2	Preliminaries . . . . .	78
5.3	Filtrations . . . . .	79

5.4	Examples of Filtrations . . . . .	82
5.5	Filtrations are Finite . . . . .	84
5.6	<b>K</b> and <b>S5</b> have the Finite Model Property . . . .	85
5.7	<b>S5</b> is Decidable . . . . .	86
5.8	Filtrations and Properties of Accessibility . . . .	87
5.9	Filtrations of Euclidean Models . . . . .	89
	Problems . . . . .	91
 <b>II Counterfactuals</b>		 <b>93</b>
<b>6</b>	<b>Introduction</b>	<b>94</b>
6.1	The Material Conditional . . . . .	94
6.2	Paradoxes of the Material Conditional . . . . .	96
6.3	The Strict Conditional . . . . .	97
6.4	Counterfactuals . . . . .	99
	Problems . . . . .	100
<b>7</b>	<b>Minimal Change Semantics</b>	<b>102</b>
7.1	Introduction . . . . .	102
7.2	Sphere Models . . . . .	104
7.3	Truth and Falsity of Counterfactuals . . . . .	106
7.4	Antecedent Strengthening . . . . .	107
7.5	Transitivity . . . . .	109
7.6	Contraposition . . . . .	111
	Problems . . . . .	111
 <b>III Appendices</b>		 <b>113</b>
<b>A</b>	<b>Sets</b>	<b>114</b>
A.1	Extensionality . . . . .	114
A.2	Subsets and Power Sets . . . . .	116
A.3	Some Important Sets . . . . .	117
A.4	Unions and Intersections . . . . .	119
A.5	Pairs, Tuples, Cartesian Products . . . . .	122
A.6	Russell's Paradox . . . . .	124
	Problems . . . . .	126



<b>B</b>	<b>Relations</b>	<b>127</b>
B.1	Relations as Sets . . . . .	127
B.2	Special Properties of Relations . . . . .	129
B.3	Equivalence Relations . . . . .	131
B.4	Orders . . . . .	132
	Problems . . . . .	135
<b>C</b>	<b>Syntax and Semantics</b>	<b>136</b>
C.1	Introduction . . . . .	136
C.2	Propositional Formulas . . . . .	138
C.3	Preliminaries . . . . .	140
C.4	Formation Sequences . . . . .	142
C.5	Valuations and Satisfaction . . . . .	144
C.6	Semantic Notions . . . . .	146
	Problems . . . . .	147
<b>D</b>	<b>Axiomatic Derivations</b>	<b>150</b>
D.1	Introduction . . . . .	150
D.2	Axiomatic Derivations . . . . .	152
D.3	Rules and Derivations . . . . .	154
D.4	Axiom and Rules for the Propositional Connectives	156
D.5	Examples of Derivations . . . . .	157
D.6	Proof-Theoretic Notions . . . . .	159
D.7	The Deduction Theorem . . . . .	161
D.8	Derivability and Consistency . . . . .	163
D.9	Derivability and the Propositional Connectives .	164
D.10	Soundness . . . . .	166
	Problems . . . . .	167
<b>E</b>	<b>Tableaux</b>	<b>168</b>
E.1	Tableaux . . . . .	168
E.2	Rules and Tableaux . . . . .	170
E.3	Propositional Rules . . . . .	171
E.4	Tableaux . . . . .	172
E.5	Examples of Tableaux . . . . .	173
E.6	Proof-Theoretic Notions . . . . .	178

E.7	Derivability and Consistency . . . . .	181
E.8	Derivability and the Propositional Connectives . . . . .	184
E.9	Soundness . . . . .	187
	Problems . . . . .	190
<b>F</b>	<b>The Completeness Theorem</b>	<b>192</b>
F.1	Introduction . . . . .	192
F.2	Outline of the Proof . . . . .	194
F.3	Complete Consistent Sets of Sentences . . . . .	195
F.4	Lindenbaum’s Lemma . . . . .	197
F.5	Construction of a Model . . . . .	198
F.6	The Completeness Theorem . . . . .	199
	Problems . . . . .	200
	<b>About the Open Logic Project</b>	<b>201</b>

# *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 II. 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 I teach the material.

**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 “ $\Box p \rightarrow p$ ” (if  $p$  is necessary, it is true) or “ $\Diamond p \rightarrow \Box \Diamond p$ ” (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  $M$  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 makes  $\Box A$  satisfied at a world  $w$  iff  $A$  is satisfied at *all* worlds:  $M, w \models \Box A$  iff  $M, v \models A$  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 logically true, and pronounce different inferences as valid. For instance, everything necessary and everything known is true, so  $\Box A \rightarrow A$  is a logical 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 logical 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  $M, w \models \Box A$  iff  $M, v \models A$  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 proposed 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

logics. 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 derivation systems (both axiomatic systems and tableaux). We establish some basic results about these logics, such as the soundness and completeness of the derivation 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 derivation systems. The third part deals with the Lewis-Stalnaker semantics of counterfactual conditionals. The appendices discuss some ideas and results from set theory and the theory of relations that's crucial to the relational semantics, and review syntax, semantics, and derivation 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:  $A \rightarrow B$  is a poor substitute for “ $A$  implies  $B$ .” Instead, they proposed to characterize implication as “Necessarily, if  $A$  then  $B$ ,” symbolized as  $A \rightarrow B$ . In trying to sort out the different properties, Lewis identified 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 gives 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  $\perp$ .
2. A countably infinite set of propositional variables:  $p_0, p_1, p_2, \dots$
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.  $\perp$  is an atomic formula.
2. Every propositional variable  $p_i$  is an (atomic) formula.
3. If  $A$  is a formula, then  $\neg A$  is a 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  $(A \rightarrow B)$  is a formula.
7. If  $A$  is a formula, then  $\Box A$  is a formula.
8. If  $A$  is a formula, then  $\Diamond A$  is a formula.
9. Nothing else is a formula.

**Definition 1.3.** Formulas constructed using the defined operators are to be understood as follows:

1.  $\top$  abbreviates  $\neg \perp$ .
2.  $A \leftrightarrow B$  abbreviates  $(A \rightarrow B) \wedge (B \rightarrow A)$ .

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

1.  $A \equiv \perp$ :  $A[D_1/p_1, \dots, D_n/p_n]$  is  $\perp$ .
2.  $A \equiv q$ :  $A[D_1/p_1, \dots, D_n/p_n]$  is  $q$ , provided  $q \neq p_i$  for  $i = 1, \dots, n$ .
3.  $A \equiv p_i$ :  $A[D_1/p_1, \dots, D_n/p_n]$  is  $D_i$ .
4.  $A \equiv \neg B$ :  $A[D_1/p_1, \dots, D_n/p_n]$  is  $\neg B[D_1/p_1, \dots, D_n/p_n]$ .
5.  $A \equiv (B \wedge C)$ :  $A[D_1/p_1, \dots, D_n/p_n]$  is

$$(B[D_1/p_1, \dots, D_n/p_n] \wedge C[D_1/p_1, \dots, D_n/p_n]).$$

6.  $A \equiv (B \vee C)$ :  $A[D_1/p_1, \dots, D_n/p_n]$  is

$$(B[D_1/p_1, \dots, D_n/p_n] \vee C[D_1/p_1, \dots, D_n/p_n]).$$

7.  $A \equiv (B \rightarrow C)$ :  $A[D_1/p_1, \dots, D_n/p_n]$  is

$$(B[D_1/p_1, \dots, D_n/p_n] \rightarrow C[D_1/p_1, \dots, D_n/p_n]).$$

8.  $A \equiv (B \leftrightarrow C)$ :  $A[D_1/p_1, \dots, D_n/p_n]$  is

$$(B[D_1/p_1, \dots, D_n/p_n] \leftrightarrow C[D_1/p_1, \dots, D_n/p_n]).$$

9.  $A \equiv \Box B$ :  $A[D_1/p_1, \dots, D_n/p_n]$  is  $\Box B[D_1/p_1, \dots, D_n/p_n]$ .
10.  $A \equiv \Diamond B$ :  $A[D_1/p_1, \dots, D_n/p_n]$  is  $\Diamond B[D_1/p_1, \dots, D_n/p_n]$ .

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

**Example 1.5.** 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:

$$\begin{aligned} & \Diamond(p_2 \rightarrow p_3) \rightarrow \Box(\Diamond(p_2 \rightarrow p_3) \wedge p_2)[\neg\Box p_1/p_2], \text{ i.e.,} \\ & \Diamond(\neg\Box p_1 \rightarrow p_3) \rightarrow \Box(\Diamond(\neg\Box p_1 \rightarrow p_3) \wedge \neg\Box p_1) \end{aligned}$$

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

$$\begin{aligned} & p_1 \rightarrow \Box(p_1 \wedge \neg\Box p_1)[\Diamond(p_2 \rightarrow p_3)/p_1], \text{ i.e.,} \\ & \Diamond(p_2 \rightarrow p_3) \rightarrow \Box(\Diamond(p_2 \rightarrow p_3) \wedge \neg\Box\Diamond(p_2 \rightarrow p_3)). \end{aligned}$$

## 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 1.6.** 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.

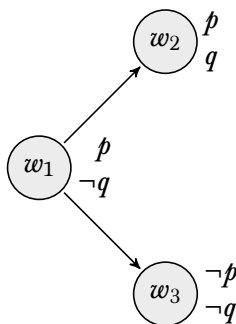


Figure 1.1: A simple model.

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

1.  $A \equiv \perp$ : Never  $M, w \models \perp$ .
2.  $M, w \models p$  iff  $w \in V(p)$ .
3.  $A \equiv \neg B$ :  $M, w \models A$  iff  $M, w \not\models B$ .
4.  $A \equiv (B \wedge C)$ :  $M, w \models A$  iff  $M, w \models B$  and  $M, w \models C$ .
5.  $A \equiv (B \vee C)$ :  $M, w \models A$  iff  $M, w \models B$  or  $M, w \models C$  (or both).
6.  $A \equiv (B \rightarrow C)$ :  $M, w \models A$  iff  $M, w \not\models B$  or  $M, w \models C$ .
7.  $A \equiv \Box B$ :  $M, w \models A$  iff  $M, w' \models B$  for all  $w' \in W$  with  $Rww'$ .
8.  $A \equiv \Diamond B$ :  $M, w \models A$  iff  $M, w' \models B$  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  $Rww'$ . 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  $M, w \not\models \Diamond A$ , for any  $A$ .

**Proposition 1.8.** 1.  $M, w \models \Box A$  iff  $M, w \models \neg \Diamond \neg A$ .

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

*Proof.* 1.  $M, w \models \neg \Diamond \neg A$  iff  $M, w \not\models \Diamond \neg A$  by definition of  $M, w \models$ .  $M, w \models \Diamond \neg A$  iff for some  $w'$  with  $Rww'$ ,  $M, w' \models \neg A$ . Hence,  $M, w \not\models \Diamond \neg A$  iff for all  $w'$  with  $Rww'$ ,  $M, w' \not\models \neg A$ . We also have  $M, w' \not\models \neg A$  iff  $M, w' \models A$ . Together we have  $M, w \models \neg \Diamond \neg A$  iff for all  $w'$  with  $Rww'$ ,  $M, w' \models A$ . Again by definition of  $M, w \models$ , that is the case iff  $M, w \models \Box A$ .

2. Exercise. □



## 1.6 Truth in a Model

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

**Definition 1.9.** A formula  $A$  is *true in a model*  $M = \langle W, R, V \rangle$ , written  $M \models A$ , if and only if  $M, w \models A$  for every  $w \in W$ .

**Proposition 1.10.** 1. If  $M \models A$  then  $M \not\models \neg A$ , but not vice-versa.

2. If  $M \models A \rightarrow B$  then  $M \models A$  only if  $M \models B$ , but not vice-versa.

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

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

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

To show that the converse fails, we need to find a model  $M$  such that  $M \models A$  only if  $M \models B$ , but  $M \not\models A \rightarrow B$ . Consider again the model of Figure 1.1:  $M \not\models p$  and hence (vacuously)  $M \models p$  only if  $M \models q$ . However,  $M \not\models p \rightarrow q$ , as  $p$  is true but  $q$  false at  $w_1$ .  $\square$

## 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,  $\Box(p \wedge q) \rightarrow \Box p$  is valid. Some formulas one might expect to be valid on the basis of the alethic interpretation of  $\Box$ , such as  $\Box p \rightarrow p$ , 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  $\Box p \rightarrow p$  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:  $\Box p \rightarrow p$  is valid in the class of reflexive models.

**Definition 1.11.** A formula  $A$  is *valid* in a class  $\mathcal{C}$  of models if it is true in every model in  $\mathcal{C}$  (i.e., true at every world in every model in  $\mathcal{C}$ ). If  $A$  is valid in  $\mathcal{C}$ , we write  $\mathcal{C} \models A$ , and we write  $\models A$  if  $A$  is valid in the class of *all* models.

**Proposition 1.12.** *If  $A$  is valid in  $\mathcal{C}$  it is also valid in each class  $\mathcal{C}' \subseteq \mathcal{C}$ .*

**Proposition 1.13.** *If  $A$  is valid, then so is  $\Box A$ .*

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

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

**Lemma 1.15.** Suppose  $A$  is a modal-free formula whose propositional variables are  $p_1, \dots, p_n$ , and let  $D_1, \dots, D_n$  be modal formulas. Then for any assignment  $v$ , any model  $M = \langle W, R, V \rangle$ , and any  $w \in W$  such that  $v(p_i) = \mathbb{T}$  if and only if  $M, w \Vdash D_i$  we have that  $v \models A$  if and only if  $M, w \Vdash A[D_1/p_1, \dots, D_n/p_n]$ .

*Proof.* By induction on  $A$ .

1.  $A \equiv \perp$ : Both  $v \not\models \perp$  and  $M, w \not\Vdash \perp$ .

2.  $A \equiv p_i$ :

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

3.  $A \equiv \neg B$ :

$$\begin{aligned}
 v \models \neg B &\Leftrightarrow v \not\models B \\
 &\quad \text{by definition of } v \models; \\
 &\Leftrightarrow M, w \not\Vdash B[D_1/p_1, \dots, D_n/p_n]
 \end{aligned}$$

$$\begin{aligned}
& \text{by induction hypothesis} \\
& \Leftrightarrow M, w \Vdash \neg B[D_1/p_1, \dots, D_n/p_n] \\
& \text{by definition of } v \models.
\end{aligned}$$

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

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

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

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

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

$$\begin{aligned}
v \models B \rightarrow C & \Leftrightarrow v \not\models B \text{ or } v \models C \\
& \text{by definition of } v \models \\
& \Leftrightarrow M, w \not\models B[D_1/p_1, \dots, D_n/p_n] \text{ or} \\
& \quad M, w \Vdash C[D_1/p_1, \dots, D_n/p_n] \\
& \text{by induction hypothesis} \\
& \Leftrightarrow M, w \Vdash (B \rightarrow C)[D_1/p_1, \dots, D_n/p_n]
\end{aligned}$$

by definition of  $M, w \models$ .

□

**Proposition 1.16.** *All tautological instances are valid.*

*Proof.* Contrapositively, suppose  $A$  is such that  $M, w \not\models A[D_1/p_1, \dots, D_n/p_n]$ , for some model  $M$  and world  $w$ . Define an assignment  $v$  such that  $v(p_i) = \mathbb{T}$  if and only if  $M, w \models D_i$  (and  $v$  assigns arbitrary values to  $q \notin \{p_1, \dots, p_n\}$ ). Then by Lemma 1.15,  $v \not\models A$ , so  $A$  is not a tautology. □

## 1.9 Schemas and Validity

**Definition 1.17.** A *schema* is a set of formulas comprising all and only the substitution instances of some modal formula  $C$ , i.e.,

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

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 ‘ $A$ ’, ‘ $B$ ’,  $\dots$ , 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 1.18.** 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 1.19.** *The following schema K is valid*

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

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

**Proposition 1.20.** *The following schema DUAL is valid*

$$\Diamond A \leftrightarrow \neg \Box \neg A. \quad (\text{DUAL})$$

*Proof.* Exercise.  $\square$

**Proposition 1.21.** *If  $A$  and  $A \rightarrow B$  are true at a world in a model then so is  $B$ . Hence, the valid formulas are closed under modus ponens.*

**Proposition 1.22.** *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  $M = \langle W, R, V \rangle$  is a modal model, and  $B \equiv A[D_1/p_1, \dots, D_n/p_n]$  is a substitution instance of  $A$ . Define  $M' = \langle W, R, V' \rangle$  by  $V'(p_i) = \{w : M, w \Vdash D_i\}$ . Then  $M, w \Vdash B$  iff  $M', w \Vdash A$ , for any  $w \in W$ . (We leave the proof as an exercise.)

<i>Valid Schemas</i>	<i>Invalid Schemas</i>
$\Box(A \rightarrow B) \rightarrow (\Diamond A \rightarrow \Diamond B)$	$\Box(A \vee B) \rightarrow (\Box A \vee \Box B)$
$\Diamond(A \rightarrow B) \rightarrow (\Box A \rightarrow \Diamond B)$	$(\Diamond A \wedge \Diamond B) \rightarrow \Diamond(A \wedge B)$
$\Box(A \wedge B) \leftrightarrow (\Box A \wedge \Box B)$	$A \rightarrow \Box A$
$\Box A \rightarrow \Box(B \rightarrow A)$	$\Box \Diamond A \rightarrow B$
$\neg \Diamond A \rightarrow \Box(A \rightarrow B)$	$\Box \Box A \rightarrow \Box A$
$\Diamond(A \vee B) \leftrightarrow (\Diamond A \vee \Diamond B)$	$\Box \Diamond A \rightarrow \Diamond \Box A.$

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  $M = \langle W, R, V \rangle$  and some  $w \in W$ ,  $M, w \not\models B$ . But then  $M', w \not\models A$  by the claim, and  $A$  is not valid, a contradiction.  $\square$

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  $M$ ,  $A$  itself is true in  $M$ . But it may happen that  $A$  is true in  $M$  but some instance of  $A$  is false at some world in  $M$ . 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  $\perp$  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.”

**Definition 1.23.** If  $\Gamma$  is a set of formulas and  $A$  a formula, then  $\Gamma$  entails  $A$ , in symbols:  $\Gamma \models A$ , if and only if for every model

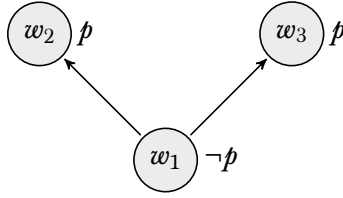


Figure 1.2: Counterexample to  $p \rightarrow \Diamond p \models \Box p \rightarrow p$ .

$M = \langle W, R, V \rangle$  and world  $w \in W$ , if  $M, w \Vdash B$  for every  $B \in \Gamma$ , then  $M, w \Vdash A$ . If  $\Gamma$  contains a single formula  $B$ , then we write  $B \models A$ .

**Example 1.24.** 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  $p \rightarrow \Diamond p \models \Box \neg p \rightarrow \neg p$ , we might argue as follows: Consider a model  $M = \langle W, R, V \rangle$  and  $w \in W$ , and suppose  $M, w \Vdash p \rightarrow \Diamond p$ . We have to show that  $M, w \Vdash \Box \neg p \rightarrow \neg p$ . 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,  $M, w \Vdash p \rightarrow \Diamond p$ , 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  $p \rightarrow \Diamond p \nmodels \Box p \rightarrow p$ . Consider the model in Figure 1.2. We have  $M, w_1 \Vdash \Diamond p$  and hence  $M, w_1 \Vdash p \rightarrow \Diamond p$ . However, since  $M, w_1 \Vdash \Box p$  but  $M, w_1 \nVdash p$ , we have  $M, w_1 \nVdash \Box p \rightarrow p$ .

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  $M', w \nVdash p$ ,  $M', w \Vdash p \rightarrow \Diamond p$ . As no worlds are accessible from  $w$ , we have  $M', w \Vdash \Box p$ , and so  $M', w \nVdash \Box p \rightarrow p$ .



## 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.  $M, w_3 \Vdash \Box \perp$ ;
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 1.8.

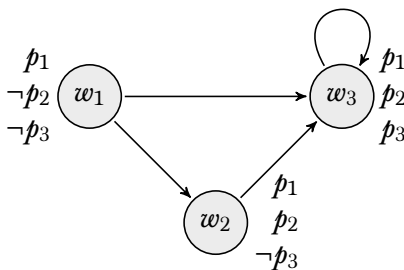
**Problem 1.3.** Let  $M = \langle W, R, V \rangle$  be a model, and suppose  $w_1, w_2 \in W$  are such that:

1.  $w_1 \in V(p)$  if and only if  $w_2 \in V(p)$  (for every propositional variable  $p$ ); 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  $M = \langle W, R, V \rangle$ . 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.  $p \rightarrow \Diamond p$  (for  $p$  atomic);
2.  $A \rightarrow \Diamond A$  (for  $A$  arbitrary);
3.  $\Box p \rightarrow p$  (for  $p$  atomic);
4.  $\neg p \rightarrow \Diamond \Box p$  (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.  $\Box p \rightarrow \Box(q \rightarrow p)$ ;
2.  $\Box \neg \perp$ ;
3.  $\Box p \rightarrow (\Box q \rightarrow \Box p)$ .

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

**Problem 1.8.** Prove Proposition 1.20.

**Problem 1.9.** Prove the claim in the “only if” part of the proof of [Proposition 1.22](#). (Hint: use induction on  $A$ .)

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

$$D: \quad \Box p \rightarrow \Diamond p;$$

$$T: \quad \Box p \rightarrow p;$$

$$B: \quad p \rightarrow \Box \Diamond p;$$

$$4: \quad \Box p \rightarrow \Box \Box p;$$

$$5: \quad \Diamond p \rightarrow \Box \Diamond p.$$

**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:

$$1. (\Diamond A \rightarrow \Box B) \rightarrow (\Box A \rightarrow \Box B);$$

$$2. \Diamond(A \rightarrow B) \vee \Box(B \rightarrow A).$$

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

$$1. p \rightarrow \Diamond \Diamond p;$$

$$2. \Diamond p \rightarrow \Box p.$$

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

**Problem 1.15.** Show that  $\Box(p \rightarrow q) \not\models p \rightarrow \Box q$  and  $p \rightarrow \Box q \not\models \Box(p \rightarrow q)$ .

## 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  $M$  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  $M$ ).

The converse, however, is false. It's not the case, e.g., that if  $\Box p \rightarrow p$  is true in  $M$ , then  $R$  is reflexive. For we can easily find a non-reflexive model  $M$  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  $M$ , 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  $\Box p \rightarrow p$  is false at  $w$ .

This suggests that we should introduce a notation for model structures without a valuation: we call these *frames*. A frame  $F$  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  $F \models A$ , the notion of a formula being *valid* in a frame as:  $M \models A$  for all  $M$  based on  $F$ .

With this notation, we can establish correspondence relations between formulas and classes of frames: e.g.,  $F \models \Box p \rightarrow p$  if, and only if,  $F$  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  $M = \langle W, R, V \rangle$  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  $M$ .*

*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  $A \rightarrow \Box \Diamond A$  be a given instance of B,

<i>If <math>R</math> is ...</i>	<i>then ... is true in <math>M</math>:</i>
<i>serial</i> : $\forall u \exists v Ruv$	$\Box p \rightarrow \Diamond p$ (D)
<i>reflexive</i> : $\forall w Rww$	$\Box p \rightarrow p$ (T)
<i>symmetric</i> : $\forall u \forall v (Ruv \rightarrow Rvu)$	$p \rightarrow \Box \Diamond p$ (B)
<i>transitive</i> : $\forall u \forall v \forall w ((Ruv \wedge Rvw) \rightarrow Ruw)$	$\Box p \rightarrow \Box \Box p$ (4)
<i>euclidean</i> : $\forall w \forall u \forall v ((Rwu \wedge Rwv) \rightarrow Ruv)$	$\Diamond p \rightarrow \Box \Diamond p$ (5)

Table 2.1: Five correspondence facts.

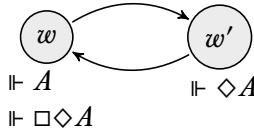


Figure 2.1: The argument from symmetry.

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  $M, w \Vdash A$ , we have  $M, w' \Vdash \Diamond A$ . Since  $w'$  was an arbitrary world such that  $Rww'$ , we have  $M, w \Vdash \Box \Diamond A$ .

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  $M, w \Vdash \Box p$  and  $M, w \not\Vdash p$ . But here we have just a single instance of T that fails in  $M$ ; 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  $M$  and  $M$  is not reflexive. But there are such models, too:

**Proposition 2.2.** *Let  $M = \langle W, R, V \rangle$  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  $A$ :  $M, u \models A$  if and only if  $M, v \models A$  (use induction on  $A$ ).*
2. *Every instance of  $T$  is true in  $M$ .*

*Since  $M$  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.”

<i>If <math>R</math> is ...</i>	<i>then ... is true in <math>M</math>:</i>
<i>partially functional:</i> $\forall w \forall u \forall v ((Rwu \wedge R w v) \rightarrow u = v)$	$\Diamond p \rightarrow \Box p$
<i>functional:</i> $\forall w \exists v \forall u (Rwu \leftrightarrow u = v)$	$\Diamond p \leftrightarrow \Box p$
<i>weakly dense:</i> $\forall u \forall v (Ruv \rightarrow \exists w (Ru w \wedge R w v))$	$\Box \Box p \rightarrow \Box p$
<i>weakly connected:</i> $\forall w \forall u \forall v ((Rwu \wedge R w v) \rightarrow (Ruv \vee u = v \vee R v u))$	$\Box((p \wedge \Box p) \rightarrow q) \vee \Box((q \wedge \Box q) \rightarrow p) \quad (L)$
<i>weakly directed:</i> $\forall w \forall u \forall v ((Rwu \wedge R w v) \rightarrow \exists t (Rut \wedge R vt))$	$\Diamond \Box p \rightarrow \Box \Diamond p \quad (G)$

Table 2.2: Five more correspondence facts.

## 2.3 Frames

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

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

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

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:  $\mathcal{F} \models A$  iff  $\mathcal{C} \models A$  where  $\mathcal{C}$  is the class of all models based on a frame in  $\mathcal{F}$ .

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  $\mathcal{F}$  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  $\mathcal{F}$  is a class of frames, we say  $A$  *defines*  $\mathcal{F}$  iff  $F \models A$  for all and only frames  $F \in \mathcal{F}$ .

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  $F$ , then  $F$  has the property on the left side.*

- Proof.*
1. Suppose  $D$  is valid in  $F = \langle W, R \rangle$ , i.e.,  $F \models \Box p \rightarrow \Diamond p$ . Let  $M = \langle W, R, V \rangle$  be a model based on  $F$ , and  $w \in W$ . We have to show that there is a  $v$  such that  $Rwv$ . Suppose not: then both  $M \models \Box A$  and  $M, w \not\models \Diamond A$  for any  $A$ , including  $p$ . But then  $M, w \not\models \Box p \rightarrow \Diamond p$ , contradicting the assumption that  $F \models \Box p \rightarrow \Diamond p$ .
  2. Suppose  $T$  is valid in  $F$ , i.e.,  $F \models \Box p \rightarrow p$ . 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  $M = \langle W, R, V \rangle$ . By construction, for all

- $u$  such that  $Rwu: M, u \Vdash p$ , and hence  $M, w \Vdash \Box p$ . But by hypothesis  $\Box p \rightarrow p$  is true at  $w$ , so that  $M, w \Vdash p$ , but by definition of  $V$  this is possible only if  $Rww$ .
3. We prove the contrapositive: Suppose  $F$  is not symmetric, we show that B, i.e.,  $p \rightarrow \Box \Diamond p$  is not valid in  $F = \langle W, R \rangle$ . If  $F$  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  $M = \langle W, R, V \rangle$ . Now, by definition of  $V$ ,  $M, w \Vdash p$  for all  $w$  such that not  $Rvw$ , in particular,  $M, u \Vdash p$  since not  $Rvu$ . Also, since  $Rvw$  iff  $w \notin V(p)$ , there is no  $w$  such that  $Rvw$  and  $M, w \Vdash p$ , and hence  $M, v \nVdash \Diamond p$ . Since  $Ruv$ , also  $M, u \nVdash \Box \Diamond p$ . It follows that  $M, u \nVdash p \rightarrow \Box \Diamond p$ , and so B is not valid in  $F$ .
  4. Suppose 4 is valid in  $F = \langle W, R \rangle$ , i.e.,  $F \models \Box p \rightarrow \Box \Box p$ , 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  $M = \langle W, R, V \rangle$ . By definition of  $V$ ,  $M, z \Vdash p$  for all  $z$  such that  $Ruz$ , and hence  $M, u \Vdash \Box p$ . But by hypothesis 4,  $\Box p \rightarrow \Box \Box p$ , is true at  $u$ , so that  $M, u \Vdash \Box \Box p$ . Since  $Ruv$  and  $Rvw$ , we have  $M, w \Vdash p$ , but by definition of  $V$  this is possible only if  $Ruw$ , as desired.
  5. We proceed contrapositively, assuming that the frame  $F = \langle W, R \rangle$  is not euclidean, and show that it falsifies 5, i.e.,  $F \nVdash \Diamond p \rightarrow \Box \Diamond p$ . Suppose there are worlds  $u, v, w \in W$  such that  $Rwu$  and  $Rvw$  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  $M = \langle W, R, V \rangle$ . Then by hypothesis  $M, v \Vdash p$  and since  $Rvw$  also  $M, w \Vdash \Diamond p$ . However, there is no world  $y$  such that  $Ruy$  and  $M, y \Vdash p$  so  $M, u \nVdash \Diamond p$ . Since  $Rwu$ , it follows that  $M, w \nVdash \Box \Diamond p$ , so that 5,  $\Diamond p \rightarrow \Box \Diamond p$ , fails at  $w$ .  $\square$

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  $M \models D$  then  $M$  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  $F$  has the property on the left, the formula on the right is valid in  $F$ . In Theorem 2.6, we proved the converse implications: if a formula on the right is valid in  $F$ ,  $F$  has the property on the left.  $\square$

Theorem 2.6 also shows that the properties can be combined: for instance if both  $B$  and  $4$  are valid in  $F$  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  $M, w \models T$  then  $M, w \models D$ .) 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  $B, p \rightarrow \Box \Diamond p$ . 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  $\forall x \forall y (Q(x, y) \rightarrow Q(y, x))$  in the sense that a first-order structure  $M$  with  $|M| = W$  and  $Q^M = R$  satisfies the preceding formula iff  $R$  is symmetric. This suggests the following definition:

**Definition 2.10.** A class  $\mathcal{F}$  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  $F = \langle W, R \rangle \in \mathcal{F}$  iff  $M \models A$  in the first-order structure  $M$  with  $|M| = W$  and  $Q^M = R$ .

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:

$$\Box(\Box p \rightarrow p) \rightarrow \Box p. \quad (W)$$

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

$w_1, w_2, \dots$  such that  $Rw_2w_1, Rw_3w_2, \dots$ . 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  $w_1, w_2, \dots$  such that  $Rw_1w_2, Rw_2w_3, \dots$ .

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

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

Now consider the set of formulas

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

Every finite subset of  $\Gamma$  is satisfiable: Let  $k$  be largest such that  $A_k$  is in the subset,  $|M_k| = \{1, \dots, k\}$ ,  $a_i^{M_k} = i$ , and  $Q^{M_k} = <$ . Since  $<$  on  $\{1, \dots, k\}$  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,  $\Gamma$  is satisfiable in some structure  $M$ . By hypothesis, since  $M \models F$ , the relation  $Q^M$  is converse well-founded. But clearly,  $a_1^M, a_2^M, \dots$  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 3.30**, 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  $F = \langle W, R \rangle$  where  $R$  is an equivalence relation, if and only if it is valid in all frames  $F = \langle W, R \rangle$  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  $M = \langle W, R, V \rangle$  based on a frame  $\langle W, R \rangle$ , where  $R$  is an equivalence on  $W$ . So  $M, w \not\models B$ . Define a model  $M' = \langle W', R', V' \rangle$  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  $M'$  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  $w' \in W'$ :  $M', w' \models A$  if and only if  $M, w' \models A$  for each  $A$  (this makes sense

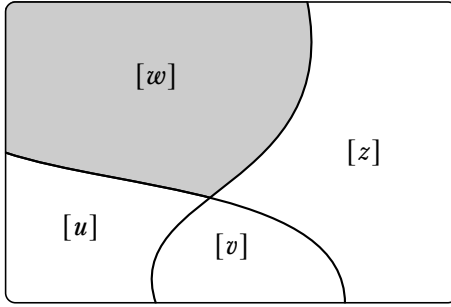


Figure 2.2: A partition of  $W$  in equivalence classes.

since  $W' \subseteq W$ ). In particular,  $M', w \not\models B$ , and  $B$  fails in a model based on a universal frame.  $\square$

## 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*  $ST_x(A)$  is inductively defined as follows:

1.  $A \equiv \perp$ :  $ST_x(A) = \perp$ .
2.  $A \equiv p_i$ :  $ST_x(A) = P_i(x)$ .



3.  $A \equiv \neg B$ :  $ST_x(A) = \neg ST_x(B)$ .
4.  $A \equiv (B \wedge C)$ :  $ST_x(A) = (ST_x(B) \wedge ST_x(C))$ .
5.  $A \equiv (B \vee C)$ :  $ST_x(A) = (ST_x(B) \vee ST_x(C))$ .
6.  $A \equiv (B \rightarrow C)$ :  $ST_x(A) = (ST_x(B) \rightarrow ST_x(C))$ .
7.  $A \equiv \Box B$ :  $ST_x(A) = \forall y (Q(x, y) \rightarrow ST_y(B))$ .
8.  $A \equiv \Diamond B$ :  $ST_x(A) = \exists y (Q(x, y) \wedge ST_y(B))$ .

For instance,  $ST_x(\Box p \rightarrow p)$  is  $\forall y (Q(x, y) \rightarrow P(y)) \rightarrow P(x)$ . Any structure for the language of  $ST_x(A)$  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  $ST_x(A)$  in the corresponding structure agree:

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

$$M, w \Vdash A \text{ iff } M', s \models ST_x(A)$$

*Proof.* By induction on  $A$ . □

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

$$\forall X_1 \dots \forall X_n \forall x ST_x(A)[X_1/P_1, \dots, X_n/P_n],$$

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

$$F \models A \text{ iff } F' \models A'$$

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

**Definition 2.18.** A class  $\mathcal{F}$  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  $F = \langle W, R \rangle \in \mathcal{F}$  iff  $M \models A$  in the structure  $M$  with  $|M| = W$  and  $P^M = R$ .

**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.  $\square$

As an example, consider again the formula  $\Box p \rightarrow p$ . 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

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

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  $M$ . 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  $\{z : Rxx\}$  where  $R = Q^M$ . So, for any  $s$  with  $s(x) \in W$  and  $s(X) = \{z : Rxx\}$  we have

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

Now suppose that  $M \models \forall x Q(x, x)$  and show that  $M \models \forall X \forall x (\forall y (Q(x, y) \rightarrow X(y)) \rightarrow X(x))$ . Pick any assignment  $s$ , and assume  $M, s \models \forall y (Q(x, y) \rightarrow X(y))$ . Let  $s'$  be the  $y$ -variant of  $s$  with  $s'(y) = s(x)$ ; we have  $M, s' \models Q(x, y) \rightarrow X(y)$ , i.e.,  $M, s \models Q(x, x) \rightarrow X(x)$ . Since  $M \models \forall x Q(x, x)$ , the antecedent is true, and we have  $M, s \models X(x)$ , 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  $M = \langle W, R, V \rangle$  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  $M$ .

**Problem 2.4.** Show that if the formula on the right side of [Table 2.2](#) is valid in a frame  $F$ , then  $F$  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

$$\frac{A \quad A \rightarrow B}{B} \text{MP}$$

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

**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  $\Sigma$  is a sequence of formulas  $B_1, B_2, \dots, B_n$ , where each  $B_i$  is either

1. a substitution instance of a tautology, or
2. a substitution instance of a formula in  $\Sigma$ , 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*  $\Sigma$ , 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  $\Sigma$  of modal formulas which

1. contains all tautologies, and
2. is closed under substitution, i.e., if  $A \in \Sigma$ , and  $D_1, \dots, D_n$  are formulas, then

$$A[D_1/p_1, \dots, D_n/p_n] \in \Sigma,$$

3. is closed under *modus ponens*, i.e., if  $A$  and  $A \rightarrow B \in \Sigma$ , 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  $\Sigma$  is *normal* if it contains

$$\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q), \quad (\mathbf{K})$$

$$\Diamond p \leftrightarrow \neg \Box \neg p \quad (\mathbf{DUAL})$$

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**,*

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

*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

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

By the induction hypothesis, we have

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

Since  $\Sigma$  is a normal modal logic, it contains all instances of **K**, in particular

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



Using modus ponens and suitable tautological instances we get

$$\Box A_1 \rightarrow (\Box A_2 \rightarrow \cdots (\Box A_{n-1} \rightarrow \Box A_n) \cdots) \in \Sigma. \quad \square$$

**Proposition 3.7.** *Every normal modal logic  $\Sigma$  contains  $\neg \Diamond \perp$ .*

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

*Proof.* Given  $A_1, \dots, A_n$ , define  $\Sigma$  as the intersection of all normal modal logics containing all instances of  $A_1, \dots, A_n$ . The intersection is non-empty as  $\text{Frm}(\mathcal{L})$ , the set of all formulas, is such a modal logic.  $\square$

**Definition 3.9.** The smallest normal modal logic containing  $A_1, \dots, A_n$  is called a *modal system* and denoted by  $\mathbf{KA}_1 \dots A_n$ . 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,  $\mathbf{K}$ , and  $\text{DUAL}$  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  $\text{MP}$  and necessitation  $\text{NEC}$ .

**Definition 3.10.** Given a modal system  $\mathbf{KA}_1 \dots A_n$  and a formula  $B$  we say that  $B$  is *derivable* in  $\mathbf{KA}_1 \dots A_n$ , written  $\mathbf{KA}_1 \dots A_n \vdash B$ , 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  $\Sigma$ -derivation of  $B$ .

**Proposition 3.11.**  $\mathbf{KA}_1 \dots A_n = \{B : \mathbf{KA}_1 \dots A_n \vdash B\}$ .

*Proof.* We use induction on the length of derivations to show that  $\{B : \mathbf{KA}_1 \dots A_n \vdash B\} \subseteq \mathbf{KA}_1 \dots A_n$ .

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  $\mathbf{KA}_1 \dots A_n$  contains these as well, so  $B \in \mathbf{KA}_1 \dots A_n$ .

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  $C \rightarrow B$  (by MP), then  $C$  and  $C \rightarrow B \in \mathbf{KA}_1 \dots A_n$  by induction hypothesis. But every modal logic is closed under modus ponens, so  $B \in \mathbf{KA}_1 \dots A_n$ . If  $B \equiv \Box C$  follows from  $C$  by NEC, then  $C \in \mathbf{KA}_1 \dots A_n$  by induction hypothesis. But every normal modal logic is closed under NEC, so  $B \in \mathbf{KA}_1 \dots A_n$ .

The converse inclusion follows by showing that  $\Sigma = \{B : \mathbf{KA}_1 \dots A_n \vdash B\}$  is a normal modal logic containing all the instances of  $A_1, \dots, A_n$ , and the observation that  $\mathbf{KA}_1 \dots A_n$  is, by definition, the smallest such logic.

1. Every tautology  $B$  is a tautological instance, so  $\mathbf{KA}_1 \dots A_n \vdash B$ , so  $\Sigma$  contains all tautologies.
2. If  $\mathbf{KA}_1 \dots A_n \vdash C$  and  $\mathbf{KA}_1 \dots A_n \vdash C \rightarrow B$ , then  $\mathbf{KA}_1 \dots A_n \vdash B$ : Combine the derivation of  $C$  with that of  $C \rightarrow B$ , and add the line  $B$ . The last line is justified by MP. So  $\Sigma$  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  $\Sigma$  is closed under uniform substitution. (We have now established that  $\Sigma$  satisfies all conditions of a modal logic.)
4. We have  $\mathbf{K}A_1 \dots A_n \vdash K$ , so  $K \in \Sigma$ .
5. We have  $\mathbf{K}A_1 \dots A_n \vdash \text{DUAL}$ , so  $\text{DUAL} \in \Sigma$ .
6. If  $\mathbf{K}A_1 \dots A_n \vdash C$ , the additional line  $\Box C$  is justified by NEC. Consequently,  $\Sigma$  is closed under NEC. Thus,  $\Sigma$  is normal.  $\square$

### 3.4 Proofs in $\mathbf{K}$

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  $\mathbf{K}$ -proofs.

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

*Proof.*

- |    |  |              |           |
|----|--|--------------|-----------|
| 1. | $A \rightarrow (B \rightarrow A)$  | TAUT         |           |
| 2. | $\Box(A \rightarrow (B \rightarrow A))$  | NEC, 1       |           |
| 3. | $\Box(A \rightarrow (B \rightarrow A)) \rightarrow (\Box A \rightarrow \Box(B \rightarrow A))$ | $\mathbf{K}$ |           |
| 4. | $\Box A \rightarrow \Box(B \rightarrow A)$   | MP, 2, 3     | $\square$ |

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

*Proof.*

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

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

$$(p \rightarrow q) \rightarrow ((p \rightarrow r) \rightarrow (p \rightarrow (q \wedge r))). \quad \square$$

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

*Proof.*

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

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

$$\begin{aligned} & (p \rightarrow q) \rightarrow ((q \rightarrow r) \rightarrow (p \rightarrow r)) \\ & (p \rightarrow (q \rightarrow r)) \rightarrow ((p \wedge q) \rightarrow r) \end{aligned} \quad \square$$

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

*Proof.*

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

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

$$\begin{aligned} & (p \rightarrow q) \rightarrow (\neg q \rightarrow \neg p) \\ & (p \rightarrow q) \rightarrow ((q \rightarrow r) \rightarrow (p \rightarrow r)). \end{aligned} \quad \square$$

### 3.5 Derived Rules

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

applying MP. Passing from  $A \rightarrow B$  and  $B \rightarrow C$  to  $A \rightarrow C$  requires recording the (long) tautological instance

$$(A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C))$$

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

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

$$A_1 \rightarrow (A_2 \rightarrow \dots (A_n \rightarrow B) \dots)$$

is a tautological instance. Applying MP  $n$  times gives a derivation of  $B$ .  $\square$

We will indicate use of this proposition by PL.

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

*Proof.* By induction on  $n$ , just as in the proof of Proposition 3.6.  $\square$

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

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

*Proof.*

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

$\square$

**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:

$$\begin{aligned} &\vdash C(A) \\ &\vdash A \leftrightarrow B \\ &\vdash C(B) \quad \text{by Proposition 3.19} \end{aligned}$$

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

$$\begin{aligned} &\mathbf{K} \vdash \neg\Box\neg\neg p \rightarrow \Diamond\neg p \\ &\mathbf{K} \vdash \neg\neg p \leftrightarrow p \quad \text{TAUT} \\ &\mathbf{K} \vdash \neg\Box p \rightarrow \Diamond\neg p \quad \text{by Proposition 3.19} \end{aligned}$$

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.  $\square$

### 3.6 More Proofs in $\mathbf{K}$

Let's see some more examples of derivability in  $\mathbf{K}$ , now using the simplified method introduced in [section 3.5](#).

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

*Proof.*

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

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

*Proof.*

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



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

*Proof.*

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

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

*Proof.*

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

$$\begin{aligned}
 p &\rightarrow \Diamond p && (\mathbf{T}_\Diamond) \\
 \Diamond \Box p &\rightarrow p && (\mathbf{B}_\Diamond) \\
 \Diamond \Diamond p &\rightarrow \Diamond p && (4_\Diamond) \\
 \Diamond \Box p &\rightarrow \Box p && (5_\Diamond)
 \end{aligned}$$

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.,  $\Box A \rightarrow \Diamond A$  is its own dual in that sense.

**Proposition 3.27.** *For each formula  $A$  in Definition 3.26:  $\mathbf{KA} = \mathbf{KA}_\Diamond$ .*

*Proof.* Exercise. □

### 3.8 Proofs in Modal Systems

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

**Proposition 3.28.** *The following provability results obtain:*

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

*Proof.* We exhibit proofs for each.

1. **KT5**  $\vdash$  B:

1. **KT5**  $\vdash \Diamond A \rightarrow \Box \Diamond A$  5
2. **KT5**  $\vdash A \rightarrow \Diamond A$   $T_\Diamond$
3. **KT5**  $\vdash A \rightarrow \Box \Diamond A$  PL.

2. **KT5**  $\vdash$  4:

1. **KT5**  $\vdash \Diamond \Box A \rightarrow \Box \Diamond \Box A$  5 with  $\Box A$  for  $p$
2. **KT5**  $\vdash \Box A \rightarrow \Diamond \Box A$   $T_{\Diamond}$  with  $\Box A$  for  $p$
3. **KT5**  $\vdash \Box A \rightarrow \Box \Diamond \Box A$  PL, 1, 2
4. **KT5**  $\vdash \Diamond \Box A \rightarrow \Box A$   $5_{\Diamond}$
5. **KT5**  $\vdash \Box \Diamond \Box A \rightarrow \Box \Box A$  RK, 4
6. **KT5**  $\vdash \Box A \rightarrow \Box \Box A$  PL, 3, 5.

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

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

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

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

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

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

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

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

□

**Definition 3.29.** 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 3.30.**  $\mathbf{KTB4} = \mathbf{KT5} = \mathbf{KDB4} = \mathbf{KDB5}$ .

*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  $A_1, \dots, A_n$ , we want every derivable formula to be true in any model in which  $A_1, \dots, A_n$  are true.

**Theorem 3.31 (Soundness Theorem).** *If every instance of  $A_1, \dots, A_n$  is valid in the classes of models  $\mathcal{C}_1, \dots, \mathcal{C}_n$ , respectively, then  $\mathbf{KA}_1 \dots A_n \vdash B$  implies that  $B$  is valid in the class of models  $\mathcal{C}_1 \cap \dots \cap \mathcal{C}_n$ .*

*Proof.* By induction on length of proofs. For brevity, put  $\mathcal{C} = \mathcal{C}_1 \cap \dots \cap \mathcal{C}_n$ .

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  $A_1, \dots, A_n$ . In the first case,  $B$  is valid in  $\mathcal{C}$ , since tautological instance are valid in *any* class of models, by [Proposition 1.16](#). Similarly in the second case, by [Proposition 1.19](#) and [Proposition 1.20](#). Finally in the third case, since  $B$  is valid in  $\mathcal{C}_i$  and  $\mathcal{C} \subseteq \mathcal{C}_i$ , we have that  $B$  is valid in  $\mathcal{C}$  as well by [Proposition 1.12](#).

2. Inductive step: Suppose  $B$  has a proof of length  $k > 1$ . If  $B$  is a tautological instance or an instance of one of  $A_1, \dots, A_n$ , we proceed as in the previous step. So suppose  $B$  is obtained by MP from previous formulas  $C \rightarrow B$  and  $C$ . Then  $C \rightarrow B$  and  $C$  have proofs of length  $< k$ , and by inductive hypothesis they are valid in  $\mathcal{C}$ . By [Proposition 1.21](#),  $B$  is valid in  $\mathcal{C}$  as well. Finally suppose  $B$  is obtained by NEC from  $C$  (so that  $B = \Box C$ ). By inductive hypothesis,  $C$  is valid in  $\mathcal{C}$ , and by [Proposition 1.13](#) so is  $B$ .  $\square$

### 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 3.31](#) allows us to show that two modal systems  $\Sigma$  and  $\Sigma'$  are distinct, by finding a formula  $A$  such that  $\Sigma' \vdash A$  that fails in a model of  $\Sigma$ .

#### Proposition 3.32. $\mathbf{KD} \subsetneq \mathbf{KT}$

*Proof.* This is the syntactic counterpart to the semantic fact that all reflexive relations are serial. To show  $\mathbf{KD} \subseteq \mathbf{KT}$  we need to see that  $\mathbf{KD} \vdash B$  implies  $\mathbf{KT} \vdash B$ , which follows from  $\mathbf{KT} \vdash \mathbf{D}$ , as shown in [Proposition 3.28\(6\)](#). To show that the inclusion is proper, by Soundness ([Theorem 3.31](#)), it suffices to exhibit a model of  $\mathbf{KD}$  where  $\top$ , i.e.,  $\Box p \rightarrow p$ , fails (an easy task left as an exercise), for then by Soundness  $\mathbf{KD} \not\vdash \Box p \rightarrow p$ .  $\square$

#### Proposition 3.33. $\mathbf{KB} \neq \mathbf{K4}$ .

*Proof.* We construct a symmetric model where some instance of  $\mathbf{4}$  fails; since obviously the instance is derivable for  $\mathbf{K4}$  but not in  $\mathbf{KB}$ , it will follow  $\mathbf{K4} \not\subseteq \mathbf{KB}$ . Consider the symmetric model  $M$  of [Figure 3.1](#). Since the model is symmetric,  $\mathbf{K}$  and  $\mathbf{B}$  are true in  $M$  (by [Proposition 1.19](#) and [Theorem 2.1](#), respectively). However,  $M, w_1 \not\models \Box p \rightarrow \Box \Box p$ .  $\square$

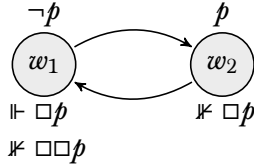


Figure 3.1: A symmetric model falsifying an instance of 4.

**Theorem 3.34.**  $\text{KTB} \not\models 4$  and  $\text{KTB} \not\models 5$ .

*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  $M, w_1 \not\models \Box p \rightarrow \Box\Box p$ , so 4 fails at  $w_1$ . Similarly,  $M, w_2 \not\models \Diamond\neg p \rightarrow \Box\Diamond\neg p$ , so the instance of 5 with  $A = \neg p$  fails at  $w_2$ .  $\square$

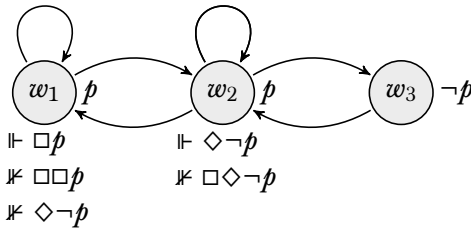


Figure 3.2: The model for Theorem 3.34.

**Theorem 3.35.**  $\text{KD5} \neq \text{KT4} = \text{S4}$ .

*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  $M, w_1 \not\models \Box p \rightarrow \Box\Box p$ .  $\square$

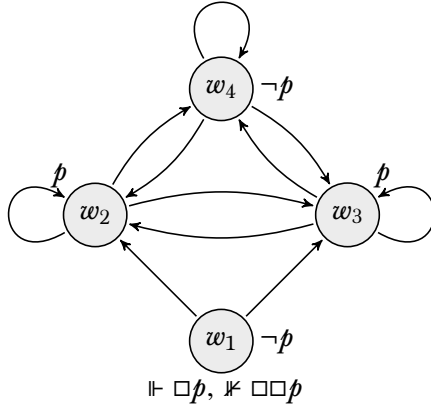


Figure 3.3: The model for Theorem 3.35.

### 3.11 Derivability from a Set of Formulas

In section 3.8 we defined a notion of provability of a formula in a system  $\Sigma$ . We now extend this notion to provability in  $\Sigma$  from formulas in a set  $\Gamma$ .

**Definition 3.36.** A formula  $A$  is derivable in a system  $\Sigma$  from a set of formulas  $\Gamma$ , written  $\Gamma \vdash_{\Sigma} A$  if and only if there are  $B_1, \dots, B_n \in \Gamma$  such that  $\Sigma \vdash B_1 \rightarrow (B_2 \rightarrow \dots (B_n \rightarrow A) \dots)$ .

### 3.12 Properties of Derivability

**Proposition 3.37.** Let  $\Sigma$  be a modal system and  $\Gamma$  a set of modal formulas. The following properties hold:

1. Monotonicity: 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  $\Gamma \vdash_{\Sigma} B \rightarrow A$ ;
5.  $\Gamma \vdash_{\Sigma} A_1$  and ... and  $\Gamma \vdash_{\Sigma} A_n$  and  $A_1 \rightarrow (A_2 \rightarrow \cdots (A_n \rightarrow B) \cdots)$  is a tautological instance, then  $\Gamma \vdash_{\Sigma} B$ .

The proof is an easy exercise. Part (5) of Proposition 3.37 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 3.38.** A set  $\Gamma$  is *deductively closed* relatively to a system  $\Sigma$  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  $\perp$ , 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 3.39.** A set  $\Gamma$  is *consistent* relatively to a system  $\Sigma$  or, as we will say,  $\Sigma$ -consistent, if and only if  $\Gamma \not\vdash_{\Sigma} \perp$ .

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

**Proposition 3.40.** Let  $\Gamma$  be a set of formulas. Then:

1.  $\Gamma$  is  $\Sigma$ -consistent if and only if there is some formula  $A$  such that  $\Gamma \not\vdash_{\Sigma} A$ .
2.  $\Gamma \vdash_{\Sigma} A$  if and only if  $\Gamma \cup \{\neg A\}$  is not  $\Sigma$ -consistent.



3. If  $\Gamma$  is  $\Sigma$ -consistent, then for any formula  $A$ , either  $\Gamma \cup \{A\}$  is  $\Sigma$ -consistent or  $\Gamma \cup \{\neg A\}$  is  $\Sigma$ -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  $\Sigma$ -consistent. Then by (2), both  $\Gamma, A \vdash_{\Sigma} \perp$  and  $\Gamma, \neg A \vdash_{\Sigma} \perp$ . By the deduction theorem  $\Gamma \vdash_{\Sigma} A \rightarrow \perp$  and  $\Gamma \vdash_{\Sigma} \neg A \rightarrow \perp$ . But  $(A \rightarrow \perp) \rightarrow ((\neg A \rightarrow \perp) \rightarrow \perp)$  is a tautological instance, hence by Proposition 3.37(5),  $\Gamma \vdash_{\Sigma} \perp$ .  $\square$

## Problems

**Problem 3.1.** Prove Proposition 3.7.

**Problem 3.2.** Find derivations in  $\mathbf{K}$  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 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.  $\mathbf{K} \vdash \Diamond \neg \perp \rightarrow (\Box A \rightarrow \Diamond A)$ ;
2.  $\mathbf{K} \vdash \Box(A \vee B) \rightarrow (\Diamond A \vee \Box B)$ ;
3.  $\mathbf{K} \vdash (\Diamond A \rightarrow \Box B) \rightarrow \Box(A \rightarrow B)$ .

**Problem 3.5.** Prove Proposition 3.27.

**Problem 3.6.** Prove Proposition 3.30.

**Problem 3.7.** Give an alternative proof of [Theorem 3.35](#) using a model with 3 worlds.

**Problem 3.8.** Provide a single reflexive transitive model showing that both  $\mathbf{KT4} \not\models B$  and  $\mathbf{KT4} \not\models 5$ .

## CHAPTER 4

# *Completeness and Canonical Models*

### 4.1 Introduction

If  $\Sigma$  is a modal system, then the soundness theorem establishes that if  $\Sigma \vdash A$ , then  $A$  is valid in any class  $\mathcal{C}$  of models in which all instances of all formulas in  $\Sigma$  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  $\not\vdash A$ , there is a countermodel, i.e., a model  $M$  such that  $M \not\models A$ . Equivalently (negating  $A$ ), we could prove that whenever  $\not\vdash \neg A$ , 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  $\not\vdash_{\Sigma} \neg A$ . This is a minimal requirement, since if  $\vdash_{\Sigma} \neg A$ , by soundness, there is no model for  $A$  (in which  $\Sigma$  is true). Now note that  $\not\vdash_{\Sigma} \neg A$  iff  $A$  is  $\Sigma$ -consistent. (Recall that  $\Sigma \not\vdash_{\Sigma} \neg A$  and  $A \not\vdash_{\Sigma} \perp$  are equivalent.) So our task is to construct a model for every  $\Sigma$ -consistent formula.

The trick we'll use is to find a  $\Sigma$ -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*  $\Sigma$ -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  $\Sigma$ -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  $\Sigma$ -consistent sets as the basis for the set of worlds. A tricky part will be to figure out when a complete  $\Sigma$ -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  $\Sigma$ -consistent set—iff  $A$  is an element of that set. If  $A$  is  $\Sigma$ -consistent, it will be an element of at least one complete  $\Sigma$ -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  $\Sigma$ -consistent formula  $A$  is true at some world. This single model is the *canonical* model for  $\Sigma$ .

## 4.2 Complete $\Sigma$ -Consistent Sets

Suppose  $\Sigma$  is a set of modal formulas—think of them as the axioms or defining principles of a normal modal logic. A set  $\Gamma$  is  $\Sigma$ -consistent iff  $\Gamma \not\vdash_{\Sigma} \perp$ , i.e., if there is no derivation of  $A_1 \rightarrow (A_2 \rightarrow \cdots (A_n \rightarrow \perp) \cdots)$  from  $\Sigma$ , where each  $A_i \in \Gamma$ . We will construct a “canonical” model in which each world is taken to be a special kind of  $\Sigma$ -consistent set: one which is not just  $\Sigma$ -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  $\Gamma$  is *complete  $\Sigma$ -consistent* if and only if it is  $\Sigma$ -consistent and for every  $A$ , either  $A \in \Gamma$  or  $\neg A \in \Gamma$ .

Complete  $\Sigma$ -consistent sets  $\Gamma$  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  $\Gamma$  mirrors the truth conditions for the propositional connectives. This will be important when we define the “canonical model.”

**Proposition 4.2.** *Suppose  $\Gamma$  is complete  $\Sigma$ -consistent. Then:*

1.  $\Gamma$  is deductively closed in  $\Sigma$ .
2.  $\Sigma \subseteq \Gamma$ .
3.  $\perp \notin \Gamma$
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.  $A \rightarrow B \in \Gamma$  iff  $A \notin \Gamma$  or  $B \in \Gamma$

- Proof.*
1. Suppose  $\Gamma \vdash_{\Sigma} A$  but  $A \notin \Gamma$ . Then since  $\Gamma$  is complete  $\Sigma$ -consistent,  $\neg A \in \Gamma$ . This would make  $\Gamma$  inconsistent, since  $A, \neg A \vdash_{\Sigma} \perp$ .
  2. If  $A \in \Sigma$  then  $\Gamma \vdash_{\Sigma} A$ , and  $A \in \Gamma$  by deductive closure, i.e., case (1).
  3. If  $\perp \in \Gamma$ , then  $\Gamma \vdash_{\Sigma} \perp$ , so  $\Gamma$  would be  $\Sigma$ -inconsistent.
  4. If  $\neg A \in \Gamma$ , then by consistency  $A \notin \Gamma$ ; and if  $A \notin \Gamma$  then  $A \in \Gamma$  since  $\Gamma$  is complete  $\Sigma$ -consistent.
  5. Exercise.
  6. Suppose  $A \vee B \in \Gamma$ , and  $A \notin \Gamma$  and  $B \notin \Gamma$ . Since  $\Gamma$  is complete  $\Sigma$ -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  $\Gamma$  is  $\Sigma$ -inconsistent, a contradiction.
  7. Exercise.

### 4.3 Lindenbaum's Lemma

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

**Theorem 4.3 (Lindenbaum's Lemma).** *If  $\Gamma$  is  $\Sigma$ -consistent then there is a complete  $\Sigma$ -consistent set  $\Delta$  extending  $\Gamma$ .*

*Proof.* Let  $A_0, A_1, \dots$  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  $p_0, \dots, p_n$ . We define sets of formulas  $\Delta_n$  by induction on  $n$ , and we then set  $\Delta = \bigcup_n \Delta_n$ . We first put  $\Delta_0 = \Gamma$ . Supposing that  $\Delta_n$  has been defined, we define  $\Delta_{n+1}$  by:

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

Now let  $\Delta = \bigcup_{n=0}^{\infty} \Delta_n$ .

We have to show that this definition actually yields a set  $\Delta$  with the required properties, i.e.,  $\Gamma \subseteq \Delta$  and  $\Delta$  is complete  $\Sigma$ -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  $\Delta$  is the union of all  $\Delta_n$ . (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  $\Delta$  is complete by definition.

Finally, we have to show, that  $\Delta$  is  $\Sigma$ -consistent. To do this, we show that (a) if  $\Delta$  were  $\Sigma$ -inconsistent, then some  $\Delta_n$  would be  $\Sigma$ -inconsistent, and (b) all  $\Delta_n$  are  $\Sigma$ -consistent.

So suppose  $\Delta$  were  $\Sigma$ -inconsistent. Then  $\Delta \vdash_{\Sigma} \perp$ , i.e., there are  $A_1, \dots, A_k \in \Delta$  such that  $\Sigma \vdash A_1 \rightarrow (A_2 \rightarrow \dots (A_k \rightarrow \perp) \dots)$ . Since  $\Delta = \bigcup_{n=0}^{\infty} \Delta_n$ , 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  $\Delta_n$ . This would mean  $\Delta_n \vdash_{\Sigma} \perp$ , i.e.,  $\Delta_n$  is  $\Sigma$ -inconsistent.

To show that each  $\Delta_n$  is  $\Sigma$ -consistent, we use a simple induction on  $n$ .  $\Delta_0 = \Gamma$ , and we assumed  $\Gamma$  was  $\Sigma$ -consistent. So the claim holds for  $n = 0$ . Now suppose it holds for  $n$ , i.e.,  $\Delta_n$  is  $\Sigma$ -consistent.  $\Delta_{n+1}$  is either  $\Delta_n \cup \{A_n\}$  if that is  $\Sigma$ -consistent, otherwise it is  $\Delta_n \cup \{\neg A_n\}$ . In the first case,  $\Delta_{n+1}$  is clearly  $\Sigma$ -consistent. However, by [Proposition 3.40\(3\)](#), either  $\Delta_n \cup \{A_n\}$  or  $\Delta_n \cup \{\neg A_n\}$  is consistent, so  $\Delta_{n+1}$  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  $\Sigma$ -consistent set  $\Delta$  extending  $\Gamma$  (including when  $\Gamma = \emptyset$ , in which case we get another characterization of the modal system  $\Sigma$ .)*

*Proof.* Suppose  $\Gamma \vdash_{\Sigma} A$ , and let  $\Delta$  be any complete  $\Sigma$ -consistent set extending  $\Gamma$ . If  $A \notin \Delta$  then by maximality  $\neg A \in \Delta$  and so  $\Delta \vdash_{\Sigma} A$  (by monotonicity) and  $\Delta \vdash_{\Sigma} \neg A$  (by reflexivity), and so  $\Delta$  is inconsistent. Conversely if  $\Gamma \not\vdash_{\Sigma} A$ , then  $\Gamma \cup \{\neg A\}$  is  $\Sigma$ -consistent, and by Lindenbaum's Lemma there is a complete consistent set  $\Delta$  extending  $\Gamma \cup \{\neg A\}$ . By consistency,  $A \notin \Delta$ .  $\square$

## 4.4 Modalities and Complete Consistent Sets

When we construct a model  $M^{\Sigma}$  whose set of worlds is given by the complete  $\Sigma$ -consistent sets  $\Delta$  in some normal modal logic  $\Sigma$ , we will also need to define an accessibility relation  $R^{\Sigma}$  between such “worlds.” We want it to be the case that the accessibility relation (and the assignment  $V^{\Sigma}$ ) 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  $\Delta'$  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  $\Delta'$  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  $R^\Sigma$  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  $\Gamma$  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  $\Gamma$  with  $\Box$  in front of every formula in  $\Gamma$ ;  $\Box^{-1}\Gamma$  is all the  $\Box$ 'ed formulas of  $\Gamma$  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  $\Gamma$  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  $B_1, \dots, B_k \in \Gamma$  such that  $\Sigma \vdash B_1 \rightarrow (B_2 \rightarrow \dots (B_n \rightarrow A) \dots)$ . Since  $\Sigma$  is normal, by rule RK,  $\Sigma \vdash \Box B_1 \rightarrow (\Box B_2 \rightarrow \dots (\Box B_n \rightarrow \Box A) \dots)$ , where obviously  $\Box B_1, \dots, \Box B_k \in \Box\Gamma$ . Hence, by definition,  $\Box\Gamma \vdash_\Sigma \Box A$ .  $\square$

**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 monotonicity.  $\square$

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

*Proof.* Suppose  $\Gamma$  is complete  $\Sigma$ -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  $\Gamma$  is complete  $\Sigma$ -consistent, it is deductively closed, and hence  $\Gamma \not\models_{\Sigma} \Box A$ . By Lemma 4.7,  $\Box^{-1}\Gamma \not\models_{\Sigma} A$ . By Proposition 3.40(2),  $\Box^{-1}\Gamma \cup \{\neg A\}$  is  $\Sigma$ -consistent. By Lindenbaum’s Lemma, there is a complete  $\Sigma$ -consistent set  $\Delta$  such that  $\Box^{-1}\Gamma \cup \{\neg A\} \subseteq \Delta$ . By consistency,  $A \notin \Delta$ .  $\square$

**Lemma 4.9.** *Suppose  $\Gamma$  and  $\Delta$  are complete  $\Sigma$ -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  $\Delta$  (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  $\Gamma$ , by consistency  $\Box A \notin \Gamma$ , as required.  $\square$

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

*Proof.* Suppose  $\Gamma$  is complete  $\Sigma$ -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  $\Gamma$  is complete  $\Sigma$ -consistent. By Proposition 4.8,  $\Box\neg A \notin \Gamma$

iff, for some complete  $\Sigma$ -consistent  $\Delta$  with  $\Box^{-1}\Gamma \subseteq \Delta$ ,  $\neg A \notin \Delta$ . Now consider any such  $\Delta$ . By Lemma 4.9,  $\Box^{-1}\Gamma \subseteq \Delta$  iff  $\Diamond\Gamma \subseteq \Delta$ . Also,  $\neg A \notin \Delta$  iff  $A \in \Delta$  by Proposition 4.2(4). So  $\Diamond A \in \Gamma$  iff, for some complete  $\Sigma$ -consistent  $\Delta$  with  $\Diamond\Gamma \subseteq \Delta$ ,  $A \in \Delta$ .  $\square$

## 4.5 Canonical Models

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

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

1.  $W^\Sigma = \{\Delta : \Delta \text{ is complete } \Sigma\text{-consistent}\}$ .
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  $M^\Sigma$  is defined in such a way that  $M^\Sigma, \Delta \Vdash A$  iff  $A \in \Delta$ . For propositional variables, the definition of  $V^\Sigma$  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.  $A \equiv \perp$ :  $M^\Sigma, \Delta \Vdash \perp$  by **Definition 1.7**, and  $\perp \notin \Delta$  by **Proposition 4.2(3)**.
2.  $A \equiv p$ :  $M^\Sigma, \Delta \Vdash p$  iff  $\Delta \in V^\Sigma(p)$  by **Definition 1.7**. Also,  $\Delta \in V^\Sigma(p)$  iff  $p \in \Delta$  by definition of  $V^\Sigma$ .
3.  $A \equiv \neg B$ :  $M^\Sigma, \Delta \Vdash \neg B$  iff  $M^\Sigma, \Delta \nVdash B$  (**Definition 1.7**) 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 1.7**) iff  $B \in \Delta$  or  $C \in \Delta$  (by inductive hypothesis) iff  $B \vee C \in \Delta$  (by **Proposition 4.2(6)**).
6.  $A \equiv B \rightarrow C$ : Exercise.
7.  $A \equiv \Box B$ : First suppose that  $M^\Sigma, \Delta \Vdash \Box B$ . By **Definition 1.7**, for every  $\Delta'$  such that  $R^\Sigma \Delta \Delta'$ ,  $M^\Sigma, \Delta' \Vdash B$ . By inductive hypothesis, for every  $\Delta'$  such that  $R^\Sigma \Delta \Delta'$ ,  $B \in \Delta'$ . By definition of  $R^\Sigma$ , for every  $\Delta'$  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  $\Delta'$  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 1.7**,  $M^\Sigma, \Delta \Vdash \Box B$ .
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 model for  $\Sigma$  are exactly the  $\Sigma$ -derivable ones. Models with this property are said to *determine*  $\Sigma$ .

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

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

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

Conversely, if  $\Sigma \vdash A$  then by [Proposition 4.2\(1\)](#), every complete  $\Sigma$ -consistent  $\Delta$  contains  $A$ , and hence by the Truth Lemma,  $M^\Sigma, \Delta \models A$  for every  $\Delta \in W^\Sigma$ , i.e.,  $M^\Sigma \models A$ .  $\square$

Since the canonical model for  $\mathbf{K}$  determines  $\mathbf{K}$ , we immediately have completeness of  $\mathbf{K}$  as a corollary:

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

*Proof.* Contrapositively, if  $\mathbf{K} \not\vdash A$  then by Determination  $M^\mathbf{K} \not\models A$  and hence  $A$  is not valid.  $\square$

For the general case of completeness of a system  $\Sigma$  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  $\Sigma$  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  $\mathbf{K}$  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  $\Sigma$  contains one of the formulas on the left-hand side of Table 4.1, then the canonical model for  $\Sigma$  has the corresponding property on the right-hand side.*

<i>If <math>\Sigma</math> contains ...</i>	<i>... the canonical model for <math>\Sigma</math> is:</i>
D: $\Box A \rightarrow \Diamond A$	serial;
T: $\Box A \rightarrow A$	reflexive;
B: $A \rightarrow \Box \Diamond A$	symmetric;
4: $\Box A \rightarrow \Box \Box A$	transitive;
5: $\Diamond A \rightarrow \Box \Diamond A$	euclidean.

Table 4.1: Basic correspondence facts.

*Proof.* We take each of these up in turn.

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

Now suppose  $\Sigma$  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  $\Sigma$  contains B, and suppose  $R^\Sigma \Delta \Delta'$  for  $\Delta, \Delta' \in W^\Sigma$ . 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  $\Sigma$  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  $\Sigma$  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.  $\square$

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  $\mathcal{C}_D, \mathcal{C}_T, \mathcal{C}_B, \mathcal{C}_4$ , and  $\mathcal{C}_5$  be the class of all serial, reflexive, symmetric, transitive, and euclidean models (respectively). Then for any schemas  $A_1, \dots, A_n$  among D, T, B, 4, and 5, the system  $\mathbf{KA}_1 \dots \mathbf{A}_n$  is determined by the class of models  $\mathcal{C} = \mathcal{C}_{A_1} \cap \dots \cap \mathcal{C}_{A_n}$ .*

**Proposition 4.18.** *Let  $\Sigma$  be a normal modal logic; then:*

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

(see [Table 2.2](#) for definitions of these frame properties).

*Proof.* 1. Suppose that  $\Sigma$  contains the schema  $\Diamond A \rightarrow \Box A$ , to show that  $R^\Sigma$  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  $\mathcal{A}_2 \subseteq \mathcal{A}_3$  and  $\mathcal{A}_3 \subseteq \mathcal{A}_2$ . For the first inclusion, let  $A \in \mathcal{A}_2$ ; then  $\Diamond A \in \mathcal{A}_1$ , and by the schema and deductive closure of  $\mathcal{A}_1$  also  $\Box A \in \mathcal{A}_1$ , whence by the hypothesis that  $R^\Sigma \mathcal{A}_1 \mathcal{A}_3$ ,  $A \in \mathcal{A}_3$ . The second inclusion is similar.

2. This follows immediately from part (1) and the seriality proof in [Theorem 4.16](#).
3. Suppose  $\Sigma$  contains the schema  $\Box\Box A \rightarrow \Box A$  and to show that  $R^\Sigma$  is weakly dense, let  $R^\Sigma \mathcal{A}_1 \mathcal{A}_2$ . We need to show that there is a complete  $\Sigma$ -consistent set  $\mathcal{A}_3$  such that  $R^\Sigma \mathcal{A}_1 \mathcal{A}_3$  and  $R^\Sigma \mathcal{A}_3 \mathcal{A}_2$ . Let:

$$\Gamma = \Box^{-1} \mathcal{A}_1 \cup \Diamond \mathcal{A}_2.$$

It suffices to show that  $\Gamma$  is  $\Sigma$ -consistent, for then by Lindenbaum's Lemma it can be extended to a complete  $\Sigma$ -consistent set  $\mathcal{A}_3$  such that  $\Box^{-1} \mathcal{A}_1 \subseteq \mathcal{A}_3$  and  $\Diamond \mathcal{A}_2 \subseteq \mathcal{A}_3$ , i.e.,  $R^\Sigma \mathcal{A}_1 \mathcal{A}_3$  and  $R^\Sigma \mathcal{A}_3 \mathcal{A}_2$  (by [Lemma 4.9](#)).

Suppose for contradiction that  $\Gamma$  is not consistent. Then there are formulas  $\Box A_1, \dots, \Box A_n \in \mathcal{A}_1$  and  $B_1, \dots, B_m \in \mathcal{A}_2$  such that

$$A_1, \dots, A_n, \Diamond B_1, \dots, \Diamond B_m \vdash_\Sigma \perp.$$

Since  $\Diamond(B_1 \wedge \dots \wedge B_m) \rightarrow (\Diamond B_1 \wedge \dots \wedge \Diamond B_m)$  is derivable in every normal modal logic, we argue as follows, contradicting the consistency of  $\mathcal{A}_2$ :

$$\begin{aligned} & A_1, \dots, A_n, \Diamond B_1, \dots, \Diamond B_m \vdash_\Sigma \perp \\ & A_1, \dots, A_n \vdash_\Sigma (\Diamond B_1 \wedge \dots \wedge \Diamond B_m) \rightarrow \perp \\ & \quad \text{by the deduction theorem} \\ & \quad \text{Proposition 3.37(4), and TAUT} \\ & A_1, \dots, A_n \vdash_\Sigma \Diamond(B_1 \wedge \dots \wedge B_m) \rightarrow \perp \\ & \quad \text{since } \Sigma \text{ is normal} \end{aligned}$$



$$\begin{aligned}
& A_1, \dots, A_n \vdash_{\Sigma} \neg \Diamond (B_1 \wedge \dots \wedge B_m) \\
& \quad \text{by PL} \\
& A_1, \dots, A_n \vdash_{\Sigma} \Box \neg (B_1 \wedge \dots \wedge B_m) \\
& \quad \Box \neg \text{ for } \neg \Diamond \\
& \Box A_1, \dots, \Box A_n \vdash_{\Sigma} \Box \Box \neg (B_1 \wedge \dots \wedge B_m) \\
& \quad \text{by Lemma 4.6} \\
& \Box A_1, \dots, \Box A_n \vdash_{\Sigma} \Box \neg (B_1 \wedge \dots \wedge B_m) \\
& \quad \text{by schema } \Box \Box A \rightarrow \Box A \\
& \Delta_1 \vdash_{\Sigma} \Box \neg (B_1 \wedge \dots \wedge B_m) \\
& \quad \text{by monotonicity, Proposition 3.37(1)} \\
& \Box \neg (B_1 \wedge \dots \wedge B_m) \in \Delta_1 \\
& \quad \text{by deductive closure;} \\
& \neg (B_1 \wedge \dots \wedge B_m) \in \Delta_2 \\
& \quad \text{since } R^{\Sigma} \Delta_1 \Delta_2. \quad \square
\end{aligned}$$

On the strength of these examples, one might think that every system  $\Sigma$  of modal logic is *complete*, in the sense that it proves every formula which is valid in every frame in which every theorem of  $\Sigma$  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  $\Gamma$  is complete  $\Sigma$ -consistent, then  $\Diamond A \in \Gamma$  if and only if there is a complete  $\Sigma$ -consistent  $\Delta$  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  $\mathbf{K}$ , 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 may 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 may be able to collect *all* worlds in finitely many (possibly infinite) “classes” of such worlds. In other words, we should partition the set of worlds in the right way, i.e., in such a way that 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  $\Gamma$  of formulas is *closed under subformulas* if it contains every subformula of a formula in  $\Gamma$ . Further,  $\Gamma$  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  $\Gamma$  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  $\Gamma$  is closed under sub-formulas. Define a relation  $\equiv$  on  $W$  to hold of any two worlds that make the same formulas from  $\Gamma$  true, i.e.:

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

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  $M$  and  $\Gamma$ ,  $\equiv$  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  $\Gamma$  true as itself. It is symmetric since if  $u$  makes the same formulas from  $\Gamma$  true as  $v$ , the same holds for  $v$  and  $u$ . It is also transitive, since if  $u$  makes the same formulas from  $\Gamma$  true as  $v$ , and  $v$  as  $w$ , then  $u$  makes the same formulas from  $\Gamma$  true as  $w$ .  $\square$

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  $M$  through  $\Gamma$ , we define when a model  $M^*$  counts as a filtration of  $M$ . 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  $\Gamma$  be closed under subformulas and  $M = \langle W, R, V \rangle$ . A *filtration of  $M$  through  $\Gamma$*  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  $\Gamma$ , 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  $\Gamma$  is closed under subformulas. Since  $A \in \Gamma$  and  $\Gamma$  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.  $A \equiv \perp$ : 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  $\Gamma$  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,  $M, w \Vdash A$ .
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  $\Gamma$  be closed under subformulas. Then:*

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

## 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  $\Gamma$ . 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  $\Gamma$  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  $M, u \models \Box A$ . By definition of  $R^*$ , there are  $u' \equiv u$  and  $v' \equiv v$  such that  $Ru'v'$ . Since  $u$  and  $u'$  agree on  $\Gamma$ , also  $M, u' \models \Box A$ , so that  $M, v' \models A$ . By closure of  $\Gamma$  under sub-formulas,  $v$  and  $v'$  agree on  $A$ , so  $M, v \models A$ , as desired.

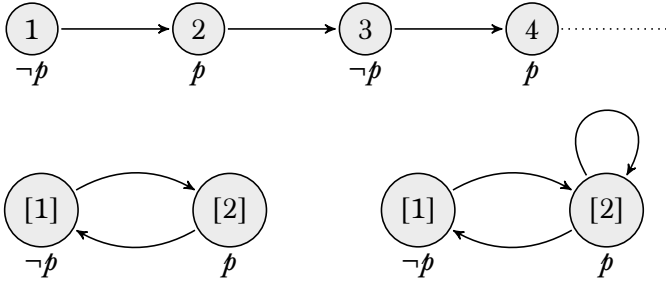


Figure 5.1: An infinite model and its filtrations.

We leave the verification of (2c) as an exercise.  $\square$

**Definition 5.9.** Where  $\Gamma$  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.  $\square$

**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  $\Gamma$  be the set of sub-formulas of  $\Box p \rightarrow p$ , i.e.,  $\{p, \Box p, \Box p \rightarrow p\}$ .  $p$  is true at all and only the even numbers,  $\Box p$  is true at all and only the odd numbers, so  $\Box p \rightarrow p$  is true at all and only the even numbers. In other words, every odd number makes  $\Box p$  true but  $p$  and  $\Box p \rightarrow p$  false; every even number makes  $p$  and  $\Box p \rightarrow p$  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  $R^*[1][1]$ ,  $M, 1 \models \Box p$  but  $M, 1 \not\models p$ , contradicting [\(2b\)](#). 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  $\Box p \rightarrow p$  are false and  $\Box p$  is true at  $[1]$ ;  $p$  and  $\Box p \rightarrow p$  are true and  $\Box p$  is false at  $[2]$ .

## 5.5 Filtrations are Finite

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

**Proposition 5.12.** *If  $\Gamma$  is finite then any filtration  $M^*$  of a model  $M$  through  $\Gamma$  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  $\Gamma$ ,  $A \in \Gamma$  either  $A$  is true at both  $u$  and  $v$ , or at neither. So each class  $[w]$  corresponds to subset of  $\Gamma$ , 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  $\Gamma$ . 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  $\Gamma$ , 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  $\Gamma$  contains  $n$  sentences, the cardinality of  $W^*$  is no greater than  $2^n$ .  $\square$

## 5.6 K and S5 have the Finite Model Property

**Definition 5.13.** A system  $\Sigma$  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  $\Sigma$  then  $A$  is true at a world in a *finite* model of  $\Sigma$ .

**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  $\Gamma$  of sub-formulas of  $A$ . Any formula only has finitely many sub-formulas, so  $\Gamma$  is finite. By [Proposition 5.12](#),  $|W^*| \leq 2^n$ , where  $n$  is the number of formulas in  $\Gamma$ . And since **K** imposes no restriction on models,  $M^*$  is a **K**-model.  $\square$

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  $\Gamma$  contain  $A$  as well as all of its subformulas; clearly  $\Gamma$  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  $R^*[u][v]$ .  $\square$

**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.  $\square$

## 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 formula 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  $p_1, \dots, p_k$ . Since for each  $n$  there are only finitely many models with  $n$  worlds assigning a value to  $p_1, \dots, p_k$ , we can enumerate, *in parallel*, all the theorems of S5 by

generating proofs in some systematic way; and all the models containing 1, 2, ... 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.  $\square$

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  $\Gamma$  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.

**Theorem 5.18.** *Let  $M = \langle W, R, V \rangle$  be a model,  $\Gamma$  closed under sub-formulas. 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  $M^* = \langle W^*, R^*, V^* \rangle$  is a filtration if  $M$  is symmetric.*

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

*Table 5.1:* Conditions on possible worlds for defining filtrations.

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

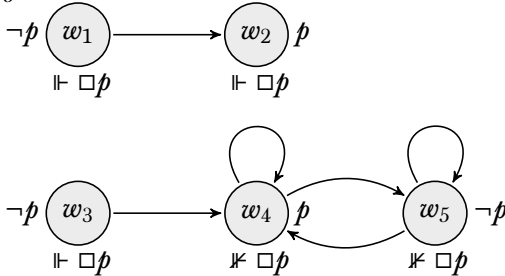
2. Exercise.

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  $\Gamma$ , then  $[w_1] = [w_3]$  since  $w_1$  and  $w_3$  are the only worlds that agree on  $\Gamma$ . 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  $w_5$ .



*Figure 5.2:* A serial and euclidean model.

In particular, to obtain a euclidean filtration it is not enough to consider filtrations through arbitrary  $\Gamma$ 's closed under subformulas. Instead we need to consider sets  $\Gamma$  that are *modally closed* (see [Definition 5.1](#)). Such sets of sentences are infinite, and



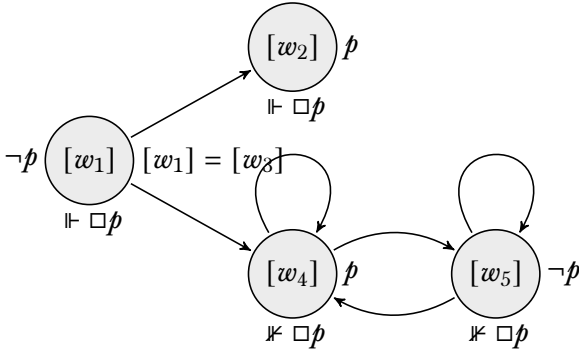


Figure 5.3: The filtration of the model in Figure 5.2.

therefore do not immediately yield a finite model property or the decidability of the corresponding system.

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

1. *If  $M$  is symmetric, so is  $M^*$ .*
2. *If  $M$  is transitive, so is  $M^*$ .*
3. *If  $M$  is euclidean, so is  $M^*$ .*

*Proof.* 1. If  $M^*$  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  $M, u \Vdash \Box A$ ; then  $M, u \Vdash \Box \Box A$  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.

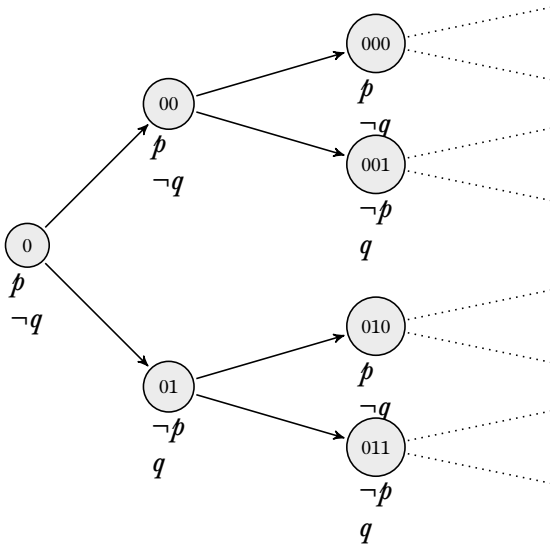
3. Exercise. Use the fact that  $B$  and  $B_{\Diamond}$  are valid in all symmetric models.  $\square$

## 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  $M, w \not\models \Box(p \vee q) \rightarrow (\Box p \vee \Box q)$  for every  $w$ .

Let  $\Gamma$  be the set of sub-formulas of  $\Box(p \vee q) \rightarrow (\Box p \vee \Box q)$ . 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.** Complete the proof of [Theorem 5.18](#).

**Problem 5.7.** Complete the proof of [Theorem 5.19](#).

## PART II

# *Counter- factuals*

## CHAPTER 6

# *Introduction*

### 6.1 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 learn 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  $v \models A \rightarrow B$  iff  $v \not\models A$  or  $v \models B$ . 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:

$$\text{If } \Gamma, A \models B \text{ then } \Gamma \models A \rightarrow B \quad (6.1)$$

It is truth-functional:  $A \rightarrow B$  and  $\neg A \vee B$  are equivalent:

$$A \rightarrow B \models \neg A \vee B \quad (6.2)$$

$$\neg A \vee B \models A \rightarrow B \quad (6.3)$$

A material conditional is entailed by its consequent and by the negation of its antecedent:

$$B \models A \rightarrow B \quad (6.4)$$

$$\neg A \models A \rightarrow B \quad (6.5)$$

A false material conditional is equivalent to the conjunction of its antecedent and the negation of its consequent: if  $A \rightarrow B$  is false,  $A \wedge \neg B$  is true, and vice versa:

$$\neg(A \rightarrow B) \models A \wedge \neg B \quad (6.6)$$

$$A \wedge \neg B \models \neg(A \rightarrow B) \quad (6.7)$$

The material conditional supports modus ponens:

$$A, A \rightarrow B \models B \quad (6.8)$$

The material conditional agglomerates:

$$A \rightarrow B, A \rightarrow C \models A \rightarrow (B \wedge C) \quad (6.9)$$

We can always strengthen the antecedent, i.e., the conditional is *monotonic*:

$$A \rightarrow B \models (A \wedge C) \rightarrow B \quad (6.10)$$

The material conditional is transitive, i.e., the chain rule is valid:

$$A \rightarrow B, B \rightarrow C \models A \rightarrow C \quad (6.11)$$

The material conditional is equivalent to its contrapositive:

$$A \rightarrow B \models \neg B \rightarrow \neg A \quad (6.12)$$

$$\neg B \rightarrow \neg A \models A \rightarrow B \quad (6.13)$$

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.

## 6.2 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.<sup>1</sup> 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  $p \models q$  we must consider *all* valuations, and  $p \not\models q$  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

---

<sup>1</sup>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.

## 6.3 The Strict Conditional

Lewis introduced the *strict conditional*  $\rightarrow$  and argued that it, not the material conditional, corresponds to implication. In alethic modal logic,  $A \rightarrow B$  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,  $(A \rightarrow B) \vee (B \rightarrow A)$  is not valid. The strict conditional  $A \rightarrow B$  is also not equivalent to  $\neg A \vee B$ , so it is not truth functional.

We have:

$$A \rightarrow B \models \neg A \vee B \text{ but:} \quad (6.14)$$

$$\neg A \vee B \not\models A \rightarrow B \quad (6.15)$$

$$B \not\models A \rightarrow B \quad (6.16)$$

$$\neg A \not\models A \rightarrow B \quad (6.17)$$

$$\neg(A \rightarrow B) \not\models A \wedge \neg B \text{ but:} \quad (6.18)$$

$$A \wedge \neg B \models \neg(A \rightarrow B) \quad (6.19)$$



However, the strict conditional still supports modus ponens:

$$A, A \rightarrow B \models B \quad (6.20)$$

The strict conditional agglomerates:

$$A \rightarrow B, A \rightarrow C \models A \rightarrow (B \wedge C) \quad (6.21)$$

Antecedent strengthening holds for the strict conditional:

$$A \rightarrow B \models (A \wedge C) \rightarrow B \quad (6.22)$$

The strict conditional is also transitive:

$$A \rightarrow B, B \rightarrow C \models A \rightarrow C \quad (6.23)$$

Finally, the strict conditional is equivalent to its contrapositive:

$$A \rightarrow B \models \neg B \rightarrow \neg A \quad (6.24)$$

$$\neg B \rightarrow \neg A \models A \rightarrow B \quad (6.25)$$

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

$$\Box \neg A \models A \rightarrow B \quad (6.26)$$

$$\Box B \models A \rightarrow B \quad (6.27)$$

$$A \rightarrow B \models \Box(A \rightarrow B) \quad (6.28)$$

$$\neg(A \rightarrow B) \models \Box \neg(A \rightarrow B) \quad (6.29)$$

These are not problems if you think of  $\rightarrow$  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  $\rightarrow$  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.

## 6.4 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 President John F. Kennedy was killed by *someone*, and if that someone wasn’t (contrary to the Warren Report) Lee Harvey Oswald, then someone else killed Kennedy. 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 6.1.** Give S5-counterexamples to the entailment relations which do not hold for the strict conditional, i.e., for:

1.  $\neg p \not\models \Box(p \rightarrow q)$

2.  $q \not\models \Box(p \rightarrow q)$
3.  $\neg\Box(p \rightarrow q) \not\models p \wedge \neg q$
4.  $\not\models \Box(p \rightarrow q) \vee \Box(q \rightarrow p)$

**Problem 6.2.** Show that the valid entailment relations hold for the strict conditional by giving **S5**-proofs of:

1.  $\Box(A \rightarrow B) \models \neg A \vee B$
2.  $A \wedge \neg B \models \neg\Box(A \rightarrow B)$
3.  $A, \Box(A \rightarrow B) \models B$
4.  $\Box(A \rightarrow B), \Box(A \rightarrow C) \models \Box(A \rightarrow (B \wedge C))$
5.  $\Box(A \rightarrow B) \models \Box((A \wedge C) \rightarrow B)$
6.  $\Box(A \rightarrow B), \Box(B \rightarrow C) \models \Box(A \rightarrow C)$
7.  $\Box(A \rightarrow B) \models \Box(\neg B \rightarrow \neg A)$
8.  $\Box(\neg B \rightarrow \neg A) \models \Box(A \rightarrow B)$

**Problem 6.3.** Give proofs in **S5** of:

1.  $\Box\neg A \models A \rightarrow B$
2.  $A \rightarrow B \models \Box(A \rightarrow B)$
3.  $\neg(A \rightarrow B) \models \Box\neg(A \rightarrow B)$

Use the definition of  $\rightarrow$  to do so.

## CHAPTER 7

# *Minimal Change Semantics*

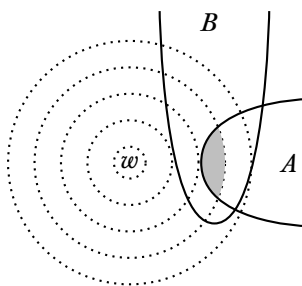
### 7.1 Introduction

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  $A \Box\rightarrow B$ . 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  $M, w \Vdash A \Box\rightarrow B$  is given in terms of  $M, w' \Vdash A$  and  $M, w' \Vdash B$  for some (other) worlds  $w'$ . Which  $w'$ ? Intuitively, the one(s) closest to  $w$  for which it holds that  $M, w' \Vdash A$ . 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,  $A \Box \rightarrow B$  is satisfied at  $w$  if  $B$  is true at all closest  $A$ -worlds.

## 7.2 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 7.1.** A *sphere model* is a triple  $M = \langle W, O, V \rangle$  where  $W$  is a non-empty set of worlds,  $V: \text{At}_0 \rightarrow \wp(W)$  is a valuation, and  $O: W \rightarrow \wp(\wp(W))$  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  $S \subseteq S'$ , 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 7.1](#) corresponds to the sphere model with  $W = \{w, w_1, \dots, w_7\}$ ,  $V(p) = \{w_5, w_6, w_7\}$ . The innermost sphere  $S_1 = \{w\}$ . The closest worlds to  $w$  are  $w_1, w_2, w_3$ , so the next larger sphere is  $S_2 = \{w, w_1, w_2, w_3\}$ . The worlds further out

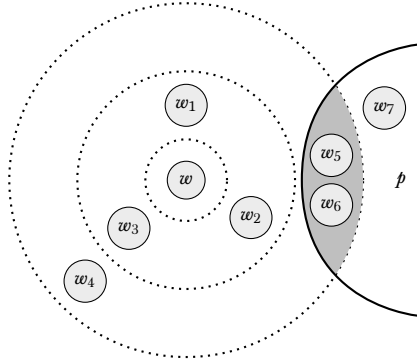


Figure 7.1: Diagram of a sphere model

are  $w_4, w_5, w_6$ , so the outermost sphere is  $S_3 = \{w, w_1, \dots, w_6\}$ . 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  $M$ ,  $M, w \models A$ , we expand the definition for modal formulas to include a clause for  $B \Box \rightarrow C$ :

**Definition 7.2.**  $M, w \models B \Box \rightarrow C$  iff either

1. For all  $u \in \bigcup O_w$ ,  $M, u \not\models B$ , or
2. For some  $S \in O_w$ ,
  - a)  $M, u \models B$  for some  $u \in S$ , and
  - b) for all  $v \in S$ , either  $M, v \not\models B$  or  $M, v \models C$ .

According to this definition,  $M, w \models B \Box \rightarrow C$  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



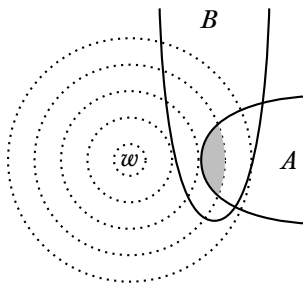


Figure 7.2: Non-vacuously true counterfactual

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  $S_1 \supsetneq S_2 \supsetneq \dots \supsetneq \{w\}$  of  $B$ -admitting spheres, and hence no innermost  $B$ -admitting spheres. In that case,  $M, w \models B \Box \rightarrow C$  iff  $B \rightarrow C$  holds throughout the spheres  $S_i$ ,  $S_{i+1}$ ,  $\dots$ , for some  $i$ .

### 7.3 Truth and Falsity of Counterfactuals

A counterfactual  $A \Box \rightarrow B$  is (non-vacuously) true if the closest  $A$ -worlds are all  $B$ -worlds, as depicted in Figure 7.2. 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 7.3).

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,  $A \Box \rightarrow \neg B$  is also false (see Figure 7.4). If the closest  $A$ -worlds do not overlap with the  $B$ -worlds at all, then  $A \Box \rightarrow B$ . But, in this case

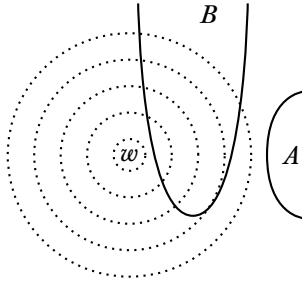


Figure 7.3: Vacuously true counterfactual

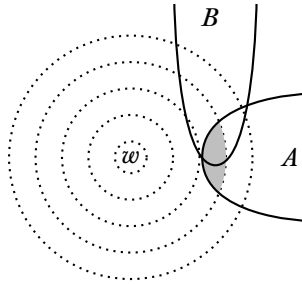


Figure 7.4: False counterfactual, false opposite

all the closest  $A$ -worlds are  $\neg B$ -worlds, and so  $A \Box \rightarrow \neg B$  is true (see Figure 7.5).

In contrast to the strict conditional, counterfactuals may be contingent. Consider the sphere model in Figure 7.6. The  $A$ -worlds closest to  $u$  are all  $B$ -worlds, so  $M, u \models A \Box \rightarrow B$ . But there are  $A$ -worlds closest to  $v$  which are not  $B$ -worlds, so  $M, v \not\models A \Box \rightarrow B$ .

## 7.4 Antecedent Strengthening

“Strengthening the antecedent” refers to the inference  $A \rightarrow C \models (A \wedge B) \rightarrow C$ . It is valid for the material conditional, but invalid

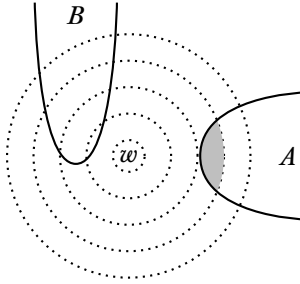


Figure 7.5: False counterfactual, true opposite

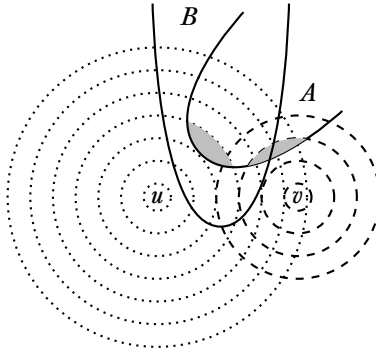


Figure 7.6: Contingent counterfactual

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:

If 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:

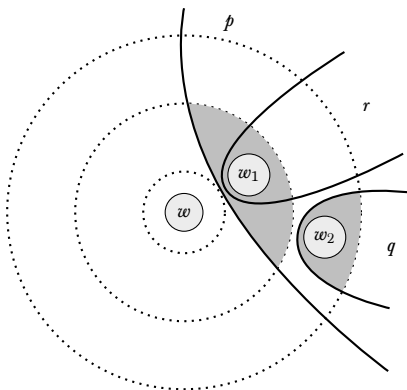


Figure 7.7: Counterexample to antecedent strengthening

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 should be.

**Example 7.3.** The sphere semantics invalidates the inference, i.e., we have  $p \Box \rightarrow r \not\models (p \wedge q) \Box \rightarrow r$ . Consider the model  $M = \langle W, O, V \rangle$  where  $W = \{w, w_1, w_2\}$ ,  $O_w = \{\{w\}, \{w, w_1\}, \{w, w_1, w_2\}\}$ ,  $V(p) = \{w_1, w_2\}$ ,  $V(q) = \{w_2\}$ , 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  $M, w \models p \Box \rightarrow r$ . There is also a  $(p \wedge q)$ -admitting sphere  $S' = \{w, w_1, w_2\}$  but  $M, w_2 \not\models (p \wedge q) \rightarrow r$ , so  $M, w \not\models (p \wedge q) \Box \rightarrow r$  (see Figure 7.7).

## 7.5 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.<sup>1</sup>

**Example 7.4.** The sphere semantics invalidates the inference, i.e., we have  $p \Box \rightarrow q, q \Box \rightarrow r \not\models p \Box \rightarrow r$ . 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  $p \rightarrow q$  is true at all worlds in it, so  $M, w \models p \Box \rightarrow q$ . There is also

---

<sup>1</sup>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  $M \vDash q \rightarrow r$  is true at all worlds in it, so  $M, w \Vdash q \Box \rightarrow r$ . However, the  $p$ -admitting sphere  $\{w, w_1, w_2\}$  contains a world, namely  $w_2$ , where  $M, w_2 \not\vDash p \rightarrow r$ .

## 7.6 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 7.5.** The sphere semantics invalidates contraposition, i.e., we have  $p \Box \rightarrow q \not\vDash \neg q \Box \rightarrow \neg p$ . 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  $p \rightarrow q$  is true at all worlds in it, so  $M, w \Vdash p \Box \rightarrow q$ . 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  $M, w_2 \not\vDash \neg q \rightarrow \neg p$ .

## Problems

**Problem 7.1.** Find a convincing, intuitive example for the failure of transitivity of counterfactuals.

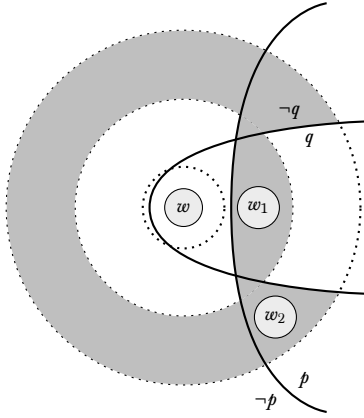


Figure 7.8: Counterexample to contraposition

**Problem 7.2.** Draw the sphere diagram corresponding to the counterexample in Example 7.4.

**Problem 7.3.** In Example 7.4, 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 III

# *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  $a_1, \dots, a_n$ , then  $\{a_1, \dots, a_n\}$  is *the* set whose elements are  $a_1, \dots, a_n$ . 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  $A \not\subseteq B$ . 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:

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}$$

the set of natural numbers

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$$

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.

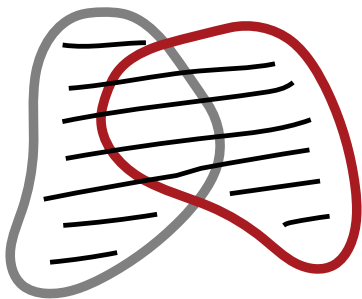
We'll sometimes also use the set of positive integers  $\mathbb{Z}^+ = \{1, 2, 3, \dots\}$  and the set containing just the first two natural numbers  $\mathbb{B} = \{0, 1\}$ .

**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*  $\Lambda$  among the strings over  $A$ , for every alphabet  $A$ . For instance,

$$\mathbb{B}^* = \{\Lambda, 0, 1, 00, 01, 10, 11, \\ 000, 001, 010, 011, 100, 101, 110, 111, 0000, \dots\}.$$

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  $A^\omega$  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  $A$ .



*Figure A.1:* The union  $A \cup B$  of two sets is set of elements of  $A$  together with those of  $B$ .

## 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  $\varphi$ , 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.

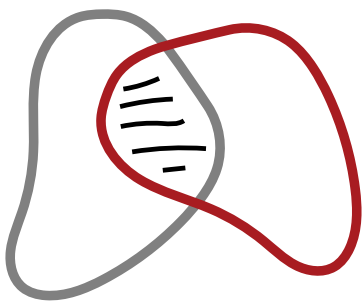
This operation on sets—combining them—is very useful and common, and so we give it a formal name and a symbol.

**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:  $\{a, b, c\} \cup \{a\} = \{a, b, c\}$ .



*Figure A.2:* The intersection  $A \cap B$  of two sets is the set of elements they have in common.

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:

$$\begin{aligned}\bigcap A &= \{x : x \text{ belongs to every element of } A\}, \text{ i.e.,} \\ &= \{x : \text{for all } B \in A, x \in B\}\end{aligned}$$

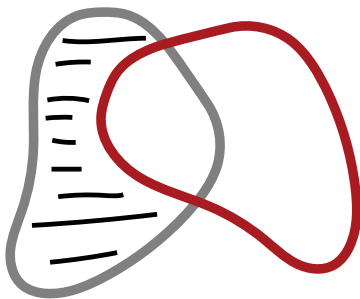
**Example A.21.** Suppose  $A = \{\{a, b\}, \{a, d, e\}, \{a, d\}\}$ . Then  $\bigcup A = \{a, b, d, e\}$  and  $\bigcap A = \{a\}$ .

We could also do the same for a sequence of sets  $A_1, A_2, \dots$

$$\begin{aligned}\bigcup_i A_i &= \{x : x \text{ belongs to one of the } A_i\} \\ \bigcap_i A_i &= \{x : x \text{ belongs to every } A_i\}.\end{aligned}$$

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





*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  $B$ .

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*  $\langle x_1, \dots, x_n \rangle$ .

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:

$$\begin{aligned} A^1 &= A \\ A^{k+1} &= A^k \times A \end{aligned}$$

**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  $A = \{x_1, \dots, x_n\}$ , then  $A \times B = B_{x_1} \cup \dots \cup B_{x_n}$ , and so has  $n \cdot m$  elements.

To visualize this, arrange the elements of  $A \times B$  in a grid:

$$\begin{array}{cccc} B_{x_1} = & \{\langle x_1, y_1 \rangle & \langle x_1, y_2 \rangle & \dots & \langle x_1, y_m \rangle\} \\ B_{x_2} = & \{\langle x_2, y_1 \rangle & \langle x_2, y_2 \rangle & \dots & \langle x_2, y_m \rangle\} \\ & \vdots & \vdots & & \vdots \\ B_{x_n} = & \{\langle x_n, y_1 \rangle & \langle x_n, y_2 \rangle & \dots & \langle x_n, y_m \rangle\} \end{array}$$

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.  $\square$

**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  $\varphi$ ’s, *then* there is only one such set. Otherwise put: having fixed some  $\varphi$ , 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.* If  $R = \{x : x \notin x\}$  exists, then  $R \in R$  iff  $R \notin R$ , which is a contradiction.  $\square$

Let's run through this proof more slowly. If  $R$  exists, it makes sense to ask whether  $R \in R$  or not. Suppose that indeed  $R \in R$ . Now,  $R$  was defined as the set of all sets that are not elements of themselves. So, if  $R \in R$ , then  $R$  does not itself have  $R$ 's defining property. 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  $R \notin R$ , then  $R$  itself does have  $R$ 's defining property, and so  $R$  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:

$$\begin{array}{ccccc} \langle \mathbf{0}, \mathbf{0} \rangle & \langle 0, 1 \rangle & \langle 0, 2 \rangle & \langle 0, 3 \rangle & \dots \\ \langle 1, 0 \rangle & \langle \mathbf{1}, \mathbf{1} \rangle & \langle 1, 2 \rangle & \langle 1, 3 \rangle & \dots \\ \langle 2, 0 \rangle & \langle 2, 1 \rangle & \langle \mathbf{2}, \mathbf{2} \rangle & \langle 2, 3 \rangle & \dots \\ \langle 3, 0 \rangle & \langle 3, 1 \rangle & \langle 3, 2 \rangle & \langle \mathbf{3}, \mathbf{3} \rangle & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

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  $\text{Id}_A = \{\langle x, x \rangle : x \in A\}$  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.,  $Ln m$  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:  $Kn m$  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  $x, y \in A$ , if  $x \neq y$ , then either  $Rxy$  or  $Ryx$ .

**Definition B.8 (Irreflexivity).** A relation  $R \subseteq A^2$  is called *irreflexive* if, for all  $x \in A$ , not  $Rxx$ .

**Definition B.9 (Asymmetry).** A relation  $R \subseteq A^2$  is called *asymmetric* if for no pair  $x, y \in A$  we have both  $Rxy$  and  $Ryx$ .

Note that if  $A \neq \emptyset$ , then no irreflexive relation on  $A$  is reflexive and every asymmetric relation on  $A$  is also anti-symmetric. However, there are  $R \subseteq A^2$  that are not reflexive and also not irreflexive, and there are anti-symmetric relations that are not asymmetric.

## 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  $Rxy$ .  $\square$

**Example B.13.** A nice example of equivalence relations comes from modular arithmetic. For any  $a$ ,  $b$ , and  $n \in \mathbb{Z}^+$ , say that  $a \equiv_n b$  iff dividing  $a$  by  $n$  gives the same remainder as dividing  $b$  by  $n$ . (Somewhat more symbolically:  $a \equiv_n b$  iff, for some  $k \in \mathbb{Z}$ ,  $a - b = kn$ .) 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*.

**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  $\preceq$  on  $\mathbb{B}^*$ :  $x \preceq y$  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,  $01 \preceq 10$  and  $10 \preceq 01$ , 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 total order* or *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  $\subsetneq$ .

Any strict order  $R$  on  $A$  can be turned into a partial order by adding the diagonal  $\text{Id}_A$ , 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  $\text{Id}_A$ . These next two results make this precise.

**Proposition B.24.** If  $R$  is a strict order on  $A$ , then  $R^+ = R \cup \text{Id}_A$  is a partial order. Moreover, if  $R$  is a strict linear order, 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  $R^+ = R \cup \text{Id}_A$ . We have to show that  $R^+$  is reflexive, anti-symmetric, and transitive.

$R^+$  is clearly reflexive, since  $\langle x, x \rangle \in \text{Id}_A \subseteq R^+$  for all  $x \in A$ .

To show  $R^+$  is anti-symmetric, suppose for reductio that  $R^+xy$  and  $R^+yx$  but  $x \neq y$ . Since  $\langle x, y \rangle \in R \cup \text{Id}_A$ , but  $\langle x, y \rangle \notin \text{Id}_A$ , 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  $\langle x, y \rangle \in \text{Id}_A$ , i.e.,  $x = y$ , or  $\langle y, z \rangle \in \text{Id}_A$ , 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 that  $R$  is also 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.  $\square$

**Proposition B.25.** *If  $R$  is a partial order on  $A$ , then  $R^- = R \setminus \text{Id}_A$  is a strict order. Moreover, if  $R$  is a linear order, then  $R^-$  is a strict linear order.*

*Proof.* This is left as an exercise.  $\square$

The following simple result establishes that strict linear orders satisfy an extensionality-like property:

**Proposition B.26.** *If  $<$  is a strict linear order on  $A$ , 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.  $\square$

## 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  $n \in \mathbb{Z}^+$ , and that  $\mathbb{N}/\equiv_n$  has exactly  $n$  members.

**Problem B.4.** Give a proof of [Proposition B.25](#).

## 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 choose different symbols, will select different sets of connectives as primitive, and 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 propositional 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  $\bar{v}(A)$  for any formula  $A$ . A formula is satisfied in a valuation  $v$  iff  $\bar{v}(A) = \mathbb{T}$ —we write this as  $v \models A$ . 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  $v \models A$  for sentences we can then define the basic semantic notions of tautology, entailment, and satisfiability. A formula is a tautology,  $\models A$ , if every valuation satisfies it, i.e.,  $\bar{v}(A) = \mathbb{T}$  for any  $v$ . It is entailed by a set of formulas,  $\Gamma \models A$ , if every valuation that satisfies all the formulas in  $\Gamma$  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  $\perp$  using *logical connectives*.

1. A countably infinite set  $At_0$  of propositional variables  $p_0, p_1, \dots$
2. The propositional constant for falsity  $\perp$ .
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  $\mathcal{L}_0$ .

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  $\perp$  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  $\text{Frm}(\mathcal{L}_0)$  of *formulas* of propositional logic is defined inductively as follows:

1.  $\perp$  is an atomic formula.
2. Every propositional variable  $p_i$  is an atomic formula.
3. If  $A$  is a formula, then  $\neg A$  is a 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  $(A \rightarrow B)$  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  $\perp$ ,  $p_i$ . “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.

When writing a formula  $(B * C)$  constructed from  $B$ ,  $C$  using a two-place connective  $*$ , we will often leave out the outermost pair of parentheses and write simply  $B * C$ .

**Definition C.2.** Formulas constructed using the defined operators are to be understood as follows:

1.  $\top$  abbreviates  $\neg\perp$ .
2.  $A \leftrightarrow B$  abbreviates  $(A \rightarrow B) \wedge (B \rightarrow A)$ .

**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  $(A \rightarrow B)$  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  $S \subseteq \text{Frm}(\mathcal{L}_0)$ .  $S$  satisfies all the conditions of **Definition C.1**: it contains all atomic formulas and is closed under

the logical operators.  $\text{Frm}(\mathcal{L}_0)$  is the smallest such class, so  $\text{Frm}(\mathcal{L}_0) \subseteq S$ . So  $\text{Frm}(\mathcal{L}_0) = S$ , and every formula has property  $P$ .  $\square$

**Proposition C.5.** *Any formula in  $\text{Frm}(\mathcal{L}_0)$  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  $\text{Frm}(\mathcal{L}_0)$  has exactly one parsing as one of the following*

1.  $\perp$ .
2.  $p_n$  for some  $p_n \in \text{At}_0$ .
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.  $(B \rightarrow C)$  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  $(B \rightarrow C)$  and  $(B' \rightarrow C')$ . 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.  $\square$

**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  $p_1, \dots, p_n$  by formulas  $B_1, \dots, B_n$  is denoted by  $A[B_1/p_1, \dots, B_n/p_n]$ .

## C.4 Formation Sequences

Defining formulas via an inductive definition, and the complementary technique of proving properties of formulas via induction, is an elegant and efficient approach. However, it can also be useful to consider a more bottom-up, step-by-step approach to the construction of formulas, which we do here using the notion of a *formation sequence*.

**Definition C.9 (Formation sequences for formulas).** A finite sequence  $\langle A_0, \dots, A_n \rangle$  of strings of symbols from the language  $\mathcal{L}_0$  is a *formation sequence* for  $A$  if  $A \equiv A_n$  and for all  $i \leq n$ , either  $A_i$  is an atomic formula or there exist  $j, k < i$  such that one of the following holds:

1.  $A_i \equiv \neg A_j$ .
2.  $A_i \equiv (A_j \wedge A_k)$ .
3.  $A_i \equiv (A_j \vee A_k)$ .
4.  $A_i \equiv (A_j \rightarrow A_k)$ .

**Example C.10.**

$$\langle p_0, p_1, (p_1 \wedge p_0), \neg(p_1 \wedge p_0) \rangle$$

is a formation sequence of  $\neg(p_1 \wedge p_0)$ , as is

$$\langle p_0, p_1, p_0, (p_1 \wedge p_0), (p_0 \rightarrow p_1), \neg(p_1 \wedge p_0) \rangle.$$

As can be seen from the second example, formation sequences may contain ‘junk’: formulas which are redundant or do not contribute to the construction.

**Proposition C.11.** *Every formula  $A$  in  $\text{Frm}(\mathcal{L}_0)$  has a formation sequence.*

*Proof.* Suppose  $A$  is atomic. Then the sequence  $\langle A \rangle$  is a formation sequence for  $A$ . Now suppose that  $B$  and  $C$  have formation sequences  $\langle B_0, \dots, B_n \rangle$  and  $\langle C_0, \dots, C_m \rangle$  respectively.

1. If  $A \equiv \neg B$ , then  $\langle B_0, \dots, B_n, \neg B_n \rangle$  is a formation sequence for  $A$ .
2. If  $A \equiv (B \wedge C)$ , then  $\langle B_0, \dots, B_n, C_0, \dots, C_m, (B_n \wedge C_m) \rangle$  is a formation sequence for  $A$ .
3. If  $A \equiv (B \vee C)$ , then  $\langle B_0, \dots, B_n, C_0, \dots, C_m, (B_n \vee C_m) \rangle$  is a formation sequence for  $A$ .
4. If  $A \equiv (B \rightarrow C)$ , then  $\langle B_0, \dots, B_n, C_0, \dots, C_m, (B_n \rightarrow C_m) \rangle$  is a formation sequence for  $A$ .

By the principle of induction on formulas, every formula has a formation sequence.  $\square$

We can also prove the converse. This is important because it shows that our two ways of defining formulas are equivalent: they give the same results. It also means that we can prove theorems about formulas by using ordinary induction on the length of formation sequences.

**Lemma C.12.** *Suppose that  $\langle A_0, \dots, A_n \rangle$  is a formation sequence for  $A_n$ , and that  $k \leq n$ . Then  $\langle A_0, \dots, A_k \rangle$  is a formation sequence for  $A_k$ .*

*Proof.* Exercise.  $\square$

**Theorem C.13.**  $\text{Frm}(\mathcal{L}_0)$  is the set of all strings of symbols in the language  $\mathcal{L}_0$  with a formation sequence.

*Proof.* Let  $F$  be the set of all strings of symbols in the language  $\mathcal{L}_0$  that have a formation sequence. We have seen in [Proposition C.11](#) that  $\text{Frm}(\mathcal{L}_0) \subseteq F$ , so now we prove the converse.

Suppose  $A$  has a formation sequence  $\langle A_0, \dots, A_n \rangle$ . We prove that  $A \in \text{Frm}(\mathcal{L}_0)$  by strong induction on  $n$ . Our induction hypothesis is that every string of symbols with a formation sequence of length  $m < n$  is in  $\text{Frm}(\mathcal{L}_0)$ . By the definition of a formation sequence, either  $A_n$  is atomic or there must exist  $j, k < n$  such that one of the following is the case:

1.  $A_n \equiv \neg A_j$ .
2.  $A_n \equiv (A_j \wedge A_k)$ .
3.  $A_n \equiv (A_j \vee A_k)$ .
4.  $A_n \equiv (A_j \rightarrow A_k)$ .

Now we reason by cases. If  $A_n$  is atomic then  $A_n \in \text{Frm}(\mathcal{L}_0)$ . Suppose instead that  $A \equiv (A_j \wedge A_k)$ . By [Lemma C.12](#),  $\langle A_0, \dots, A_j \rangle$  and  $\langle A_0, \dots, A_k \rangle$  are formation sequences for  $A_j$  and  $A_k$  respectively. Since these are proper initial subsequences of the formation sequence for  $A$ , they both have length less than  $n$ . Therefore by the induction hypothesis,  $A_j$  and  $A_k$  are in  $\text{Frm}(\mathcal{L}_0)$ , and so by the definition of a formula, so is  $(A_j \wedge A_k)$ . The other cases follow by parallel reasoning.  $\square$

## C.5 Valuations and Satisfaction

**Definition C.14 (Valuations).** Let  $\{\mathbb{T}, \mathbb{F}\}$  be the set of the two truth values, “true” and “false.” A *valuation* for  $\mathcal{L}_0$  is a function  $v$  assigning either  $\mathbb{T}$  or  $\mathbb{F}$  to the propositional variables of the language, i.e.,  $v: \text{At}_0 \rightarrow \{\mathbb{T}, \mathbb{F}\}$ .

**Definition C.15.** Given a valuation  $v$ , define the evaluation function  $\bar{v}: \text{Frm}(\mathcal{L}_0) \rightarrow \{\mathbb{T}, \mathbb{F}\}$  inductively by:

$$\begin{aligned}\bar{v}(\perp) &= \mathbb{F}; \\ \bar{v}(\rho_n) &= v(\rho_n); \\ \bar{v}(\neg A) &= \begin{cases} \mathbb{T} & \text{if } \bar{v}(A) = \mathbb{F}; \\ \mathbb{F} & \text{otherwise.} \end{cases} \\ \bar{v}(A \wedge B) &= \begin{cases} \mathbb{T} & \text{if } \bar{v}(A) = \mathbb{T} \text{ and } \bar{v}(B) = \mathbb{T}; \\ \mathbb{F} & \text{if } \bar{v}(A) = \mathbb{F} \text{ or } \bar{v}(B) = \mathbb{F}. \end{cases} \\ \bar{v}(A \vee B) &= \begin{cases} \mathbb{T} & \text{if } \bar{v}(A) = \mathbb{T} \text{ or } \bar{v}(B) = \mathbb{T}; \\ \mathbb{F} & \text{if } \bar{v}(A) = \mathbb{F} \text{ and } \bar{v}(B) = \mathbb{F}. \end{cases} \\ \bar{v}(A \rightarrow B) &= \begin{cases} \mathbb{T} & \text{if } \bar{v}(A) = \mathbb{F} \text{ or } \bar{v}(B) = \mathbb{T}; \\ \mathbb{F} & \text{if } \bar{v}(A) = \mathbb{T} \text{ and } \bar{v}(B) = \mathbb{F}. \end{cases}\end{aligned}$$

The clauses correspond to the following truth tables:

$A$	$\neg A$
$\mathbb{T}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$

$A$	$B$	$A \wedge B$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{F}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$

$A$	$B$	$A \vee B$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$

$A$	$B$	$A \rightarrow B$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{F}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{T}$

**Theorem C.16 (Local Determination).** Suppose that  $v_1$  and  $v_2$  are valuations that agree on the propositional variables 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.17 (Satisfaction).** We can inductively define the notion of *satisfaction of a formula  $A$  by a valuation  $v$* ,  $v \models A$ , as follows. (We write  $v \not\models A$  to mean “not  $v \models A$ .”)

1.  $A \equiv \perp$ :  $v \not\models A$ .
2.  $A \equiv p_i$ :  $v \models A$  iff  $v(p_i) = \mathbb{T}$ .
3.  $A \equiv \neg B$ :  $v \models A$  iff  $v \not\models B$ .
4.  $A \equiv (B \wedge C)$ :  $v \models A$  iff  $v \models B$  and  $v \models C$ .
5.  $A \equiv (B \vee C)$ :  $v \models A$  iff  $v \models B$  or  $v \models C$  (or both).
6.  $A \equiv (B \rightarrow C)$ :  $v \models A$  iff  $v \not\models B$  or  $v \models C$  (or both).

If  $\Gamma$  is a set of formulas,  $v \models \Gamma$  iff  $v \models A$  for every  $A \in \Gamma$ .

**Proposition C.18.**  $v \models A$  iff  $\overline{v}(A) = \mathbb{T}$ .

*Proof.* By induction on  $A$ . □

## C.6 Semantic Notions

We define the following semantic notions:

- Definition C.19.**
1. A formula  $A$  is *satisfiable* if for some  $v$ ,  $v \models A$ ; it is *unsatisfiable* if for no  $v$ ,  $v \models A$ ;
  2. A formula  $A$  is a *tautology* if  $v \models A$  for all valuations  $v$ ;

3. A formula  $A$  is *contingent* if it is satisfiable but not a tautology;
4. If  $\Gamma$  is a set of formulas,  $\Gamma \models A$  (“ $\Gamma$  entails  $A$ ”) if and only if  $v \models A$  for every valuation  $v$  for which  $v \models \Gamma$ .
5. If  $\Gamma$  is a set of formulas,  $\Gamma$  is *satisfiable* if there is a valuation  $v$  for which  $v \models \Gamma$ , and  $\Gamma$  is *unsatisfiable* otherwise.

**Proposition C.20.**     1.  $A$  is a tautology if and only if  $\emptyset \models A$ ;

2. If  $\Gamma \models A$  and  $\Gamma \models A \rightarrow B$  then  $\Gamma \models B$ ;

3. If  $\Gamma$  is satisfiable then every finite subset of  $\Gamma$  is also satisfiable;

4. Monotonicity: if  $\Gamma \subseteq \Delta$  and  $\Gamma \models A$  then also  $\Delta \models A$ ;

5. Transitivity: if  $\Gamma \models A$  and  $\Delta \cup \{A\} \models B$  then  $\Gamma \cup \Delta \models B$ .

*Proof.* Exercise. □

**Proposition C.21.**  $\Gamma \models A$  if and only if  $\Gamma \cup \{\neg A\}$  is unsatisfiable.

*Proof.* Exercise. □

**Theorem C.22 (Semantic Deduction Theorem).**  $\Gamma \models A \rightarrow B$  if and only if  $\Gamma \cup \{A\} \models B$ .

*Proof.* Exercise. □

## Problems

**Problem C.1.** Prove Proposition C.5

**Problem C.2.** Prove Proposition C.6

**Problem C.3.** For each of the five formulas below determine whether the formula can be expressed as a substitution  $A[B/p_i]$  where  $A$  is (i)  $p_0$ ; (ii)  $(\neg p_0 \wedge p_1)$ ; and (iii)  $((\neg p_0 \rightarrow p_1) \wedge p_2)$ . In each case specify the relevant substitution.

1.  $p_1$
2.  $(\neg p_0 \wedge p_0)$
3.  $((p_0 \vee p_1) \wedge p_2)$
4.  $\neg((p_0 \rightarrow p_1) \wedge p_2)$
5.  $((\neg(p_0 \rightarrow p_1) \rightarrow (p_0 \vee p_1)) \wedge \neg(p_0 \wedge p_1))$

**Problem C.4.** Give a mathematically rigorous definition of  $A[B/p]$  by induction.

**Problem C.5.** Consider adding to  $\mathcal{L}_0$  a ternary connective  $\diamond$  with evaluation given by

$$\bar{v}(\diamond(A, B, C)) = \begin{cases} \bar{v}(B) & \text{if } \bar{v}(A) = \mathbb{T}; \\ \bar{v}(C) & \text{if } \bar{v}(A) = \mathbb{F}. \end{cases}$$

Write down the truth table for this connective.

**Problem C.6.** Prove [Proposition C.18](#)

**Problem C.7.** For each of the following four formulas determine whether it is (a) satisfiable, (b) tautology, and (c) contingent.

1.  $(p_0 \rightarrow (\neg p_1 \rightarrow \neg p_0))$ .
2.  $((p_0 \wedge \neg p_1) \rightarrow (\neg p_0 \wedge p_2)) \leftrightarrow ((p_2 \rightarrow p_0) \rightarrow (p_0 \rightarrow p_1))$ .
3.  $(p_0 \leftrightarrow p_1) \rightarrow (p_2 \leftrightarrow \neg p_1)$ .
4.  $((p_0 \leftrightarrow (\neg p_1 \wedge p_2)) \vee (p_2 \rightarrow (p_0 \leftrightarrow p_1)))$ .

**Problem C.8.** Prove [Proposition C.20](#)

**Problem C.9.** Prove Proposition C.21

**Problem C.10.** Prove Theorem C.22

## 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 derivation 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 derivation 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  $\Gamma$ .” 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 unsatisfiability: 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  $\Gamma$  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 counts 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  $\Gamma$  of sentences, or
3.  $A$  is justified by a rule of inference.

To be an axiom,  $A$  has to have the form of one 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

$$A \rightarrow (B \rightarrow A) \quad B \rightarrow (B \vee C) \quad (B \wedge C) \rightarrow B$$

are common axioms that govern  $\rightarrow$ ,  $\vee$  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  $A \rightarrow B$  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  $A \rightarrow B$  (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  $\Gamma$  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  $\Gamma$  is empty, i.e., every sentence in the derivation is justified either by (1) or (3). For instance, here is a derivation that shows that  $\vdash A \rightarrow (B \rightarrow (B \vee A))$ :

1.  $B \rightarrow (B \vee A)$
2.  $(B \rightarrow (B \vee A)) \rightarrow (A \rightarrow (B \rightarrow (B \vee A)))$
3.  $A \rightarrow (B \rightarrow (B \vee A))$

The sentence on line 1 is of the form of the axiom  $A \rightarrow (A \vee B)$  (with the roles of  $A$  and  $B$  reversed). The sentence on line 2 is of the form of the axiom  $A \rightarrow (B \rightarrow A)$ . Thus, both lines are justified. Line 3 is justified by modus ponens: if we abbreviate it as  $D$ , then line 2 has the form  $C \rightarrow D$ , where  $C$  is  $B \rightarrow (B \vee A)$ , i.e., line 1.



A set  $\Gamma$  is inconsistent if  $\Gamma \vdash \perp$ . A complete axiom system will also prove that  $\perp \rightarrow A$  for any  $A$ , and so if  $\Gamma$  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 derivation 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  $\Gamma$  is a set of formulas of  $\mathcal{L}$  then a *derivation* from  $\Gamma$  is a finite sequence  $A_1, \dots, A_n$  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 a derivation 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  $\Gamma$ .

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  $\Gamma$ .

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  $B \rightarrow A$  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:  $A_1, \dots, A_n$  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  $B \rightarrow A_i$ , and for some  $k < i$ ,  $A_k$  is  $B$ .

The last clause says that  $A_i$  follows from  $A_j$  ( $B$ ) and  $A_k$  ( $B \rightarrow A_i$ ) by modus ponens. If we can go from 1 to  $n$ , and each time we find a formula  $A_i$  that is either in  $\Gamma$ , 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  $\Gamma$ , written  $\Gamma \vdash A$ , if there is a derivation from  $\Gamma$  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  $\text{Ax}_0$  of *axioms* for the propositional connectives comprises all formulas of the following forms:

$$(A \wedge B) \rightarrow A \quad (\text{D.1})$$

$$(A \wedge B) \rightarrow B \quad (\text{D.2})$$

$$A \rightarrow (B \rightarrow (A \wedge B)) \quad (\text{D.3})$$

$$A \rightarrow (A \vee B) \quad (\text{D.4})$$

$$A \rightarrow (B \vee A) \quad (\text{D.5})$$

$$(A \rightarrow C) \rightarrow ((B \rightarrow C) \rightarrow ((A \vee B) \rightarrow C)) \quad (\text{D.6})$$

$$A \rightarrow (B \rightarrow A) \quad (\text{D.7})$$

$$(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C)) \quad (\text{D.8})$$

$$(A \rightarrow B) \rightarrow ((A \rightarrow \neg B) \rightarrow \neg A) \quad (\text{D.9})$$

$$\neg A \rightarrow (A \rightarrow B) \quad (\text{D.10})$$

$$\top \quad (\text{D.11})$$

$$\perp \rightarrow A \quad (\text{D.12})$$

$$(A \rightarrow \perp) \rightarrow \neg A \quad (\text{D.13})$$

$$\neg\neg A \rightarrow A \quad (\text{D.14})$$

**Definition D.6 (Modus ponens).** If  $B$  and  $B \rightarrow A$  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  $(\neg D \vee E) \rightarrow (D \rightarrow E)$ . 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  $A \rightarrow B$  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  $D \rightarrow E$ , i.e., the instance

$$(\neg D \rightarrow (D \rightarrow E)) \rightarrow ((E \rightarrow (D \rightarrow E)) \rightarrow ((\neg D \vee E) \rightarrow (D \rightarrow E))).$$

Why? Two applications of MP yield the last part, which is what we want. And we easily see that  $\neg D \rightarrow (D \rightarrow E)$  is an instance of [eq. \(D.10\)](#), and  $E \rightarrow (D \rightarrow E)$  is an instance of [eq. \(D.7\)](#). So our derivation is:

1.  $\neg D \rightarrow (D \rightarrow E)$  [eq. \(D.10\)](#)
2.  $(\neg D \rightarrow (D \rightarrow E)) \rightarrow$   
 $((E \rightarrow (D \rightarrow E)) \rightarrow ((\neg D \vee E) \rightarrow (D \rightarrow E)))$  [eq. \(D.6\)](#)
3.  $((E \rightarrow (D \rightarrow E)) \rightarrow ((\neg D \vee E) \rightarrow (D \rightarrow E)))$  1, 2, MP
4.  $E \rightarrow (D \rightarrow E)$  [eq. \(D.7\)](#)
5.  $(\neg D \vee E) \rightarrow (D \rightarrow E)$  3, 4, MP

**Example D.8.** Let's try to find a derivation of  $D \rightarrow D$ . 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  $A \rightarrow B$  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  $D \rightarrow D$ , 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):

$$D \rightarrow (D \rightarrow D)$$

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  $\rightarrow$  is eq. (D.8), i.e.,

$$(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$$

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  $A \rightarrow C$  is  $D \rightarrow D$ , the formula we are aiming for. Then of course,  $A$  and  $C$  are both  $D$ . How should we pick  $B$  so that both  $A \rightarrow (B \rightarrow C)$  and  $A \rightarrow B$ , i.e., in our case  $D \rightarrow (B \rightarrow D)$  and  $D \rightarrow B$ , are also derivable? Well, the first of these is already an instance of eq. (D.7), whatever we decide  $B$  to be. And  $D \rightarrow B$  would be another instance of eq. (D.7) if  $B$  were  $(D \rightarrow D)$ . So, our derivation is:

- |    |  |           |
|----|--|-----------|
| 1. | $D \rightarrow ((D \rightarrow D) \rightarrow D)$  | eq. (D.7) |
| 2. | $(D \rightarrow ((D \rightarrow D) \rightarrow D)) \rightarrow$<br>$((D \rightarrow (D \rightarrow D)) \rightarrow (D \rightarrow D))$ | eq. (D.8) |
| 3. | $(D \rightarrow (D \rightarrow D)) \rightarrow (D \rightarrow D)$  | 1, 2, MP  |
| 4. | $D \rightarrow (D \rightarrow D)$  | eq. (D.7) |
| 5. | $D \rightarrow D$  | 3, 4, MP  |

**Example D.9.** Sometimes we want to show that there is a derivation of some formula from some other formulas  $\Gamma$ . For instance, let's show that we can derive  $A \rightarrow C$  from  $\Gamma = \{A \rightarrow B, B \rightarrow C\}$ .

1.	$A \rightarrow B$	HYP
2.	$B \rightarrow C$	HYP
3.	$(B \rightarrow C) \rightarrow (A \rightarrow (B \rightarrow C))$	eq. (D.7)
4.	$A \rightarrow (B \rightarrow C)$	2, 3, MP
5.	$(A \rightarrow (B \rightarrow C)) \rightarrow$ $((A \rightarrow B) \rightarrow (A \rightarrow C))$	eq. (D.8)
6.	$((A \rightarrow B) \rightarrow (A \rightarrow C))$	4, 5, MP
7.	$A \rightarrow C$	1, 6, MP

The lines labelled “HYP” (for “hypothesis”) indicate that the formula on that line is an element of  $\Gamma$ .

**Proposition D.10.** *If  $\Gamma \vdash A \rightarrow B$  and  $\Gamma \vdash B \rightarrow C$ , then  $\Gamma \vdash A \rightarrow C$*

*Proof.* Suppose  $\Gamma \vdash A \rightarrow B$  and  $\Gamma \vdash B \rightarrow C$ . Then there is a derivation of  $A \rightarrow B$  from  $\Gamma$ ; and a derivation of  $B \rightarrow C$  from  $\Gamma$  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  $A \rightarrow C$ —which is the last line of the new derivation—from  $\Gamma$ . 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  $A \rightarrow B$ , and reference to line number 1 by reference to the last line of the derivation of  $B \rightarrow C$ .  $\square$

## D.6 Proof-Theoretic Notions

Just as we’ve defined a number of important semantic notions (tautology, entailment, satisfiability), 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  $\Gamma$ , written  $\Gamma \vdash A$ , if there is a derivation from  $\Gamma$  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  $\Gamma$  of formulas is *consistent* if and only if  $\Gamma \nvdash \perp$ ; 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  $\Gamma$ .  $\square$

**Proposition D.15 (Monotonicity).** If  $\Gamma \subseteq \Delta$  and  $\Gamma \vdash A$ , then  $\Delta \vdash A$ .

*Proof.* Any derivation of  $A$  from  $\Gamma$  is also a derivation of  $A$  from  $\Delta$ .  $\square$

**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  $B_1, \dots, B_l = B$  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  $\Gamma$ , i.e., a derivation  $A_1, \dots, A_k = A$  where every  $A_i$  is an axiom, an element of  $\Gamma$ , or correct by a rule of inference. Now consider the sequence

$$A_1, \dots, A_k = A, B_1, \dots, B_l = B.$$

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  $A_k = A$ .  $\square$

Note that this means that in particular if  $\Gamma \vdash A$  and  $A \vdash B$ , then  $\Gamma \vdash B$ . It follows also that if  $A_1, \dots, A_n \vdash B$  and  $\Gamma \vdash A_i$  for each  $i$ , then  $\Gamma \vdash B$ .

**Proposition D.17.**  *$\Gamma$  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  $\Gamma$  is consistent, then  $\Gamma$  is consistent.*

*Proof.* 1. If  $\Gamma \vdash A$ , then there is a finite sequence of formulas  $A_1, \dots, A_n$  so that  $A \equiv A_n$  and each  $A_i$  is either a logical axiom, an element of  $\Gamma$  or follows from previous formulas by modus ponens. Take  $\Gamma_0$  to be those  $A_i$  which are in  $\Gamma$ . Then the derivation is likewise a derivation from  $\Gamma_0$ , and so  $\Gamma_0 \vdash A$ .

2. This is the contrapositive of (1) for the special case  $A \equiv \perp$ .

□

## 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  $\Gamma \vdash A \rightarrow B$ , then  $\Gamma \vdash B$ .*

*Proof.* We have that  $\{A, A \rightarrow B\} \vdash B$ :

1.  $A$  Hyp.
2.  $A \rightarrow B$  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  $\Gamma \vdash A \rightarrow B$ .*

*Proof.* The “if” direction is immediate. If  $\Gamma \vdash A \rightarrow B$  then also  $\Gamma \cup \{A\} \vdash A \rightarrow B$  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  $\Gamma \vdash B \rightarrow (A \rightarrow B)$  by [eq. \(D.7\)](#), and [Proposition D.19](#) gives  $\Gamma \vdash A \rightarrow B$ . If  $B \in \{A\}$  then  $\Gamma \vdash A \rightarrow B$  because then last sentence  $A \rightarrow B$  is the same as  $A \rightarrow A$ , 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  $C \rightarrow B$  and  $C$ , for some formula  $C$ , i.e.,  $\Gamma \cup \{A\} \vdash C \rightarrow B$  and  $\Gamma \cup \{A\} \vdash C$ , and

the respective derivations are shorter, so the inductive hypothesis applies to them. We thus have both:

$$\Gamma \vdash A \rightarrow (C \rightarrow B);$$

$$\Gamma \vdash A \rightarrow C.$$

But also

$$\Gamma \vdash (A \rightarrow (C \rightarrow B)) \rightarrow ((A \rightarrow C) \rightarrow (A \rightarrow B)),$$

by eq. (D.8), and two applications of Proposition D.19 give  $\Gamma \vdash A \rightarrow B$ , as required.  $\square$

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.  $\vdash (A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C));$

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  $\Gamma$  is inconsistent.*

*Proof.* If  $\Gamma \cup \{A\}$  is inconsistent, then  $\Gamma \cup \{A\} \vdash \perp$ . 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**,  $\Gamma \vdash \perp$ , i.e.,  $\Gamma$  is inconsistent.  $\square$

**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  $\vdash \neg A \rightarrow (A \rightarrow \perp)$  by **eq. (D.10)**. So by two applications of **Proposition D.19**, we have  $\Gamma \cup \{\neg A\} \vdash \perp$ .

Now assume  $\Gamma \cup \{\neg A\}$  is inconsistent, i.e.,  $\Gamma \cup \{\neg A\} \vdash \perp$ . By the deduction theorem,  $\Gamma \vdash \neg A \rightarrow \perp$ .  $\Gamma \vdash (\neg A \rightarrow \perp) \rightarrow \neg\neg A$  by **eq. (D.13)**, so  $\Gamma \vdash \neg\neg A$  by **Proposition D.19**. Since  $\Gamma \vdash \neg\neg A \rightarrow A$  (**eq. (D.14)**), we have  $\Gamma \vdash A$  by **Proposition D.19** again.  $\square$

**Proposition D.24.** If  $\Gamma \vdash A$  and  $\neg A \in \Gamma$ , then  $\Gamma$  is inconsistent.

*Proof.*  $\Gamma \vdash \neg A \rightarrow (A \rightarrow \perp)$  by **eq. (D.10)**.  $\Gamma \vdash \perp$  by two applications of **Proposition D.19**.  $\square$

**Proposition D.25.** If  $\Gamma \cup \{A\}$  and  $\Gamma \cup \{\neg A\}$  are both inconsistent, then  $\Gamma$  is inconsistent.

*Proof.* Exercise.  $\square$

## D.9 Derivability and the Propositional Connectives

We establish that the derivability relation  $\vdash$  of axiomatic deduction is strong enough to establish some basic facts involving the propositional connectives, such as that  $A \wedge B \vdash A$  and  $A, A \rightarrow B \vdash B$  (modus ponens). These facts are needed for the proof of the completeness theorem.

**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.      $\square$

**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  $\vdash \neg A \rightarrow (A \rightarrow \perp)$  and  $\vdash \neg B \rightarrow (B \rightarrow \perp)$ . So by the deduction theorem, we have  $\{\neg A\} \vdash A \rightarrow \perp$  and  $\{\neg B\} \vdash B \rightarrow \perp$ . From eq. (D.6) we get  $\{\neg A, \neg B\} \vdash (A \vee B) \rightarrow \perp$ . By the deduction theorem,  $\{A \vee B, \neg A, \neg B\} \vdash \perp$ .

2. From eq. (D.4) and eq. (D.5) by modus ponens.      $\square$

**Proposition D.28.**     1.  $A, A \rightarrow B \vdash B$ .

2. Both  $\neg A \vdash A \rightarrow B$  and  $B \vdash A \rightarrow B$ .

*Proof.*     1. We can derive:

- |    |                   |          |
|----|-------------------|----------|
| 1. | $A$               | HYP      |
| 2. | $A \rightarrow B$ | HYP      |
| 3. | $B$               | 1, 2, MP |

2. By eq. (D.10) and eq. (D.7) and the deduction theorem, respectively.      $\square$

## 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  $\Gamma$ , it is also a consequence of them;
3. if a set of formulas  $\Gamma$  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  $v \models A$  for each valuation  $v$ .*

*Proof.* Do truth tables for each axiom to verify that they are tautologies.  $\square$

**Theorem D.30 (Soundness).** *If  $\Gamma \vdash A$  then  $\Gamma \models A$ .*

*Proof.* By induction on the length of the derivation of  $A$  from  $\Gamma$ . If there are no steps justified by inferences, then all formulas in the derivation are either instances of axioms or are in  $\Gamma$ . By the previous proposition, all the axioms are tautologies, and hence if  $A$  is an axiom then  $\Gamma \models A$ . If  $A \in \Gamma$ , then trivially  $\Gamma \models A$ .

If the last step of the derivation of  $A$  is justified by modus ponens, then there are formulas  $B$  and  $B \rightarrow A$  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,  $\Gamma \models B$  and  $\Gamma \models B \rightarrow A$ . Then  $\Gamma \models A$  by [Theorem C.22](#).

**Corollary D.31.** *If  $\vdash A$ , then  $A$  is a tautology.*

**Corollary D.32.** *If  $\Gamma$  is satisfiable, then it is consistent.*

*Proof.* We prove the contrapositive. Suppose that  $\Gamma$  is not consistent. Then  $\Gamma \vdash \perp$ , i.e., there is a derivation of  $\perp$  from  $\Gamma$ . By [Theorem D.30](#), any valuation  $v$  that satisfies  $\Gamma$  must satisfy  $\perp$ . Since  $v \not\models \perp$  for every valuation  $v$ , no  $v$  can satisfy  $\Gamma$ , i.e.,  $\Gamma$  is not satisfiable.  $\square$

## Problems

**Problem D.1.** Show that the following hold by exhibiting derivations from the axioms:

1.  $(A \wedge B) \rightarrow (B \wedge A)$
2.  $((A \wedge B) \rightarrow C) \rightarrow (A \rightarrow (B \rightarrow C))$
3.  $\neg(A \vee B) \rightarrow \neg A$

**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  $\Gamma_0 = \{B_1, \dots, B_n\} \subseteq \Gamma$  such that there is a closed tableau for the assumptions

$$\{\mathbb{F} A, \mathbb{T} B_1, \dots, \mathbb{T} B_n\}$$

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  $\Gamma$  is inconsistent in the tableau calculus if there is a closed tableau for assumptions

$$\{\mathbb{T} B_1, \dots, \mathbb{T} B_n\}$$

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 building 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 where 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 $\neg$

$$\frac{\mathsf{T} \neg A}{\mathsf{F} A} \neg\mathsf{T}$$

$$\frac{\mathsf{F} \neg A}{\mathsf{T} A} \neg\mathsf{F}$$

### Rules for $\wedge$

$$\frac{\mathsf{T} A \wedge B}{\mathsf{T} A \quad \mathsf{T} B} \wedge\mathsf{T}$$

$$\frac{\mathsf{F} A \wedge B}{\mathsf{F} A \quad | \quad \mathsf{F} B} \wedge\mathsf{F}$$

### Rules for $\vee$

$$\frac{\mathsf{T} A \vee B}{\mathsf{T} A \quad | \quad \mathsf{T} B} \vee\mathsf{T}$$

$$\frac{\mathsf{F} A \vee B}{\mathsf{F} A \quad \mathsf{F} B} \vee\mathsf{F}$$

### Rules for $\rightarrow$

$$\frac{\mathsf{T} A \rightarrow B}{\mathsf{F} A \quad | \quad \mathsf{T} B} \rightarrow\mathsf{T}$$

$$\frac{\mathsf{F} A \rightarrow B}{\mathsf{T} A \quad \mathsf{F} B} \rightarrow\mathsf{F}$$

### The Cut Rule

$$\frac{}{\mathsf{T} A \quad | \quad \mathsf{F} A} \text{Cut}$$

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  $\mathsf{T} A$ , the other  $\mathsf{F} A$ . 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  $S_1A_1, \dots, S_nA_n$  (where each  $S_i$  is either  $\mathbb{T}$  or  $\mathbb{F}$ ) is a finite 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  $(A \wedge B) \rightarrow A$ .

We begin by writing the corresponding assumption at the top of the tableau.

- |    |   |            |
|----|---|------------|
| 1. | $\mathbb{F} (A \wedge B) \rightarrow A$ | 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  $\rightarrow\mathbb{F}$ .

- |    |  |                            |
|----|--|----------------------------|
| 1. | $\mathbb{F} (A \wedge B) \rightarrow A \checkmark$ | Assumption                 |
| 2. | $\mathbb{T} A \wedge B$                            | $\rightarrow \mathbb{F} 1$ |
| 3. | $\mathbb{F} A$                                     | $\rightarrow \mathbb{F} 1$ |

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 2. Applying the  $\wedge \mathbb{T}$  rule results in:

- |    |  |                            |
|----|--|----------------------------|
| 1. | $\mathbb{F} (A \wedge B) \rightarrow A \checkmark$ | Assumption                 |
| 2. | $\mathbb{T} A \wedge B \checkmark$                 | $\rightarrow \mathbb{F} 1$ |
| 3. | $\mathbb{F} A$                                     | $\rightarrow \mathbb{F} 1$ |
| 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  $(A \wedge B) \rightarrow A$ .

**Example E.5.** Now let's find a closed tableau for  $(\neg A \vee B) \rightarrow (A \rightarrow B)$ .

We begin with the corresponding assumption:

- |    |  |            |
|----|--|------------|
| 1. | $\mathbb{F} (\neg A \vee B) \rightarrow (A \rightarrow B)$ | Assumption |
|----|--|------------|

The one signed formula in this tableau has main operator  $\rightarrow$  and sign  $\mathbb{F}$ , so we apply the  $\rightarrow \mathbb{F}$  rule to it to obtain:

1.	$\mathbb{F} (\neg A \vee B) \rightarrow (A \rightarrow B) \checkmark$	Assumption
2.	$\mathbb{T} \neg A \vee B$	$\rightarrow \mathbb{F} 1$
3.	$\mathbb{F} (A \rightarrow B)$	$\rightarrow \mathbb{F} 1$


We now have a choice as to whether to apply  $\vee \mathbb{T}$  to line 2 or  $\rightarrow \mathbb{F}$  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.	$\mathbb{F} (\neg A \vee B) \rightarrow (A \rightarrow B) \checkmark$	Assumption
2.	$\mathbb{T} \neg A \vee B \checkmark$	$\rightarrow \mathbb{F} 1$
3.	$\mathbb{F} (A \rightarrow B)$	$\rightarrow \mathbb{F} 1$
$\swarrow \quad \searrow$		
4.	$\mathbb{T} \neg A \quad \mathbb{T} B$	$\vee \mathbb{T} 2$

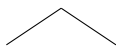
We have not applied the  $\rightarrow \mathbb{F}$  rule to line 3 yet: let's do that now. To save time, we apply it to both branches. Recall that we write a checkmark next to a signed formula only if we have applied the corresponding rule in every open branch. So it's a good idea to apply a rule at the 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.	$\mathbb{F} (\neg A \vee B) \rightarrow (A \rightarrow B) \checkmark$	Assumption
2.	$\mathbb{T} \neg A \vee B \checkmark$	$\rightarrow \mathbb{F} 1$
3.	$\mathbb{F} (A \rightarrow B) \checkmark$	$\rightarrow \mathbb{F} 1$
$\swarrow \quad \searrow$		
4.	$\mathbb{T} \neg A \quad \mathbb{T} B$	$\vee \mathbb{T} 2$
5.	$\mathbb{T} A \quad \mathbb{T} A$	$\rightarrow \mathbb{F} 3$
6.	$\mathbb{F} B \quad \mathbb{F} B$	$\rightarrow \mathbb{F} 3$
$\otimes$		

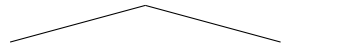
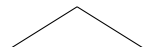
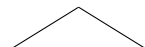
The right branch is now closed. On the left branch, we can still apply the  $\neg\mathsf{T}$  rule to line 4. This results in  $\mathsf{F} A$  and closes the left branch:

1.	$\mathsf{F} (\neg A \vee B) \rightarrow (A \rightarrow B) \checkmark$	Assumption
2.	$\mathsf{T} \neg A \vee B \checkmark$	$\rightarrow\mathsf{F} 1$
3.	$\mathsf{F} (A \rightarrow B) \checkmark$	$\rightarrow\mathsf{F} 1$
		
4.	$\mathsf{T} \neg A$	$\vee\mathsf{T} 2$
5.	$\mathsf{T} A$	$\rightarrow\mathsf{F} 3$
6.	$\mathsf{F} B$	$\rightarrow\mathsf{F} 3$
7.	$\mathsf{F} A$	$\neg\mathsf{T} 4$
	$\otimes$	

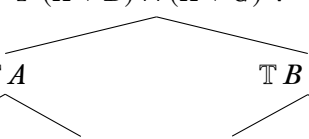
**Example E.6.** We can give tableaux for any number of signed formulas as assumptions. Often it is also necessary to apply more than one rule that allows branching; and in general a tableau can have any number of branches. For instance, consider a tableau for  $\{\mathsf{T} A \vee (B \wedge C), \mathsf{F} (A \vee B) \wedge (A \vee C)\}$ . We start by applying the  $\vee\mathsf{T}$  to the first assumption:

1.	$\mathsf{T} A \vee (B \wedge C) \checkmark$	Assumption
2.	$\mathsf{F} (A \vee B) \wedge (A \vee C)$	Assumption
		
3.	$\mathsf{T} A$	$\vee\mathsf{T} 1$

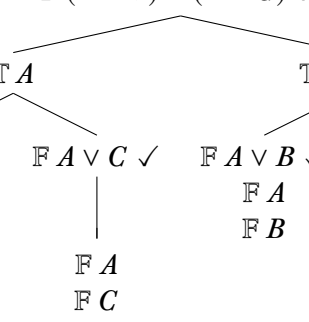
Now we can apply the  $\wedge\mathsf{F}$  rule to line 2. We do this on both branches simultaneously, and can therefore check off line 2:

1.	$\mathsf{T} A \vee (B \wedge C) \checkmark$	Assumption
2.	$\mathsf{F} (A \vee B) \wedge (A \vee C) \checkmark$	Assumption
		
3.	$\mathsf{T} A$	$\vee\mathsf{T} 1$
		
4.	$\mathsf{F} A \vee B$	$\wedge\mathsf{F} 2$
		
	$\mathsf{F} A \vee C$	

Now we can apply  $\vee\mathbb{F}$  to all the branches containing  $A \vee B$ :

1.	$\mathbb{T} A \vee (B \wedge C) \checkmark$	Assumption
2.	$\mathbb{F} (A \vee B) \wedge (A \vee C) \checkmark$	Assumption
		
3.	$\mathbb{T} A$ $\mathbb{T} B \wedge C$	$\vee\mathbb{T} 1$
4.	$\mathbb{F} A \vee B \checkmark$ $\mathbb{F} A \vee C$ $\mathbb{F} A \vee B \checkmark$ $\mathbb{F} A \vee C$	$\wedge\mathbb{F} 2$
5.	$\mathbb{F} A$ $\mathbb{F} A$	$\vee\mathbb{F} 4$
6.	$\mathbb{F} B$ $\mathbb{F} B$	$\vee\mathbb{F} 4$
	$\otimes$	

The leftmost branch is now closed. Let's now apply  $\vee\mathbb{F}$  to  $A \vee C$ :

1.	$\mathbb{T} A \vee (B \wedge C) \checkmark$	Assump
2.	$\mathbb{F} (A \vee B) \wedge (A \vee C) \checkmark$	Assump
		
3.	$\mathbb{T} A$ $\mathbb{T} B \wedge C$	$\vee\mathbb{T} 1$
4.	$\mathbb{F} A \vee B \checkmark$ $\mathbb{F} A \vee C \checkmark$ $\mathbb{F} A \vee B \checkmark$ $\mathbb{F} A \vee C \checkmark$	$\wedge\mathbb{F} 2$
5.	$\mathbb{F} A$ $\mathbb{F} A$	$\vee\mathbb{F} 4$
6.	$\mathbb{F} B$ $\mathbb{F} B$	$\vee\mathbb{F} 4$
7.	$\otimes$ $\mathbb{F} A$	$\vee\mathbb{F} 4$
8.	$\mathbb{F} C$ $\mathbb{F} C$	$\vee\mathbb{F} 4$
	$\otimes$	

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:



1.	$\text{T } A \vee (B \wedge C) \checkmark$				Assump
2.	$\text{F } (A \vee B) \wedge (A \vee C) \checkmark$				Assump
3.	$\text{T } A$		$\text{T } B \wedge C \checkmark$		$\vee\text{T } 1$
4.	$\text{F } A \vee B \checkmark$	$\text{F } A \vee C \checkmark$	$\text{F } A \vee B \checkmark$	$\text{F } A \vee C \checkmark$	$\wedge\text{F } 2$
5.	$\text{F } A$	$\text{F } A$	$\text{F } A$	$\text{F } A$	$\vee\text{F } 4$
6.	$\text{F } B$	$\text{F } C$	$\text{F } B$	$\text{F } C$	$\vee\text{F } 4$
7.	$\otimes$	$\otimes$	$\text{T } B$	$\text{T } B$	$\wedge\text{T } 3$
8.			$\text{T } C$	$\text{T } C$	$\wedge\text{T } 3$
			$\otimes$	$\otimes$	

For comparison, here's a closed tableau for the same set of assumptions in which the rules are applied in a different order:

1.	$\text{T } A \vee (B \wedge C) \checkmark$		Assumption
2.	$\text{F } (A \vee B) \wedge (A \vee C) \checkmark$		Assumption
3.	$\text{F } A \vee B \checkmark$	$\text{F } A \vee C \checkmark$	$\wedge\text{F } 2$
4.	$\text{F } A$	$\text{F } A$	$\vee\text{F } 3$
5.	$\text{F } B$	$\text{F } C$	$\vee\text{F } 3$
6.	$\text{T } A$	$\text{T } B \wedge C \checkmark$	$\vee\text{T } 1$
7.	$\otimes$	$\text{T } B$	$\wedge\text{T } 6$
8.		$\text{T } C$	$\wedge\text{T } 6$
		$\otimes$	

## E.6 Proof-Theoretic Notions

Just as we've defined a number of important semantic notions (validity, entailment, satisfiability), 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  $\Gamma$ ,  $\Gamma \vdash A$  iff there is a finite set  $\{B_1, \dots, B_n\} \subseteq \Gamma$  and a closed tableau for the set

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

If  $A$  is not derivable from  $\Gamma$  we write  $\Gamma \nvdash A$ .

**Definition E.9 (Consistency).** A set of sentences  $\Gamma$  is *inconsistent* iff there is a finite set  $\{B_1, \dots, B_n\} \subseteq \Gamma$  and a closed tableau for the set

$$\{\mathbb{T} B_1, \dots, \mathbb{T} B_n\}.$$

If  $\Gamma$  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  $\Gamma$  and the tableau

- |    |                |            |
|----|----------------|------------|
| 1. | $\mathbb{F} A$ | Assumption |
| 2. | $\mathbb{T} A$ | Assumption |
|    | $\otimes$      |            |

is closed. □

**Proposition E.11 (Monotonicity).** *If  $\Gamma \subseteq \Delta$  and  $\Gamma \vdash A$ , then  $\Delta \vdash A$ .*

*Proof.* Any finite subset of  $\Gamma$  is also a finite subset of  $\Delta$ .  $\square$

**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  $\Delta_0 = \{C_1, \dots, C_n\} \subseteq \Delta$  such that

$$\{\mathbb{F} B, \mathbb{T} A, \mathbb{T} C_1, \dots, \mathbb{T} C_n\}$$

has a closed tableau. If  $\Gamma \vdash A$  then there are  $D_1, \dots, D_m$  such that

$$\{\mathbb{F} A, \mathbb{T} D_1, \dots, \mathbb{T} D_m\}$$

has a closed tableau.

Now consider the tableau with assumptions

$$\mathbb{F} B, \mathbb{T} C_1, \dots, \mathbb{T} C_n, \mathbb{T} D_1, \dots, \mathbb{T} D_m.$$

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

$$\{\mathbb{F} B, \mathbb{T} A, \mathbb{T} C_1, \dots, \mathbb{T} C_n\}$$

are available. Since there is a closed tableau for these assumptions, we can attach it to that branch; every branch through  $\mathbb{T} A$  closes. On the other branch, all of

$$\{\mathbb{F} A, \mathbb{T} D_1, \dots, \mathbb{T} D_m\}$$

are available, so we can also complete the other side to obtain a closed tableau. This shows  $\Gamma \cup \Delta \vdash B$ .  $\square$

Note that this means that in particular if  $\Gamma \vdash A$  and  $A \vdash B$ , then  $\Gamma \vdash B$ . It follows also that if  $A_1, \dots, A_n \vdash B$  and  $\Gamma \vdash A_i$  for each  $i$ , then  $\Gamma \vdash B$ .

**Proposition E.13.**  *$\Gamma$  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  $\Gamma$  is consistent, then  $\Gamma$  is consistent.*

*Proof.* 1. If  $\Gamma \vdash A$ , then there is a finite subset  $\Gamma_0 = \{B_1, \dots, B_n\}$  and a closed tableau for

$$\{\mathbb{F} A, \mathbb{T} B_1, \dots, \mathbb{T} B_n\}$$

This tableau also shows  $\Gamma_0 \vdash A$ .

2. If  $\Gamma$  is inconsistent, then for some finite subset  $\Gamma_0 = \{B_1, \dots, B_n\}$  there is a closed tableau for

$$\{\mathbb{T} B_1, \dots, \mathbb{T} B_n\}$$

This closed tableau shows that  $\Gamma_0$  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  $\Gamma$  is inconsistent.*

*Proof.* There are finite  $\Gamma_0 = \{B_1, \dots, B_n\}$  and  $\Gamma_1 = \{C_1, \dots, C_m\} \subseteq \Gamma$  such that

$$\{\mathbb{F} A, \mathbb{T} B_1, \dots, \mathbb{T} B_n\}$$

$$\{\mathbb{T} A, \mathbb{T} C_1, \dots, \mathbb{T} C_m\}$$

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  $\Gamma$  is inconsistent.  $\square$

**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

$$\{\mathbb{F} A, \mathbb{T} B_1, \dots, \mathbb{T} B_n\}$$

Using the  $\neg\mathbb{T}$  rule, this can be turned into a closed tableau for

$$\{\mathbb{T} \neg A, \mathbb{T} B_1, \dots, \mathbb{T} B_n\}.$$

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.  $\square$

**Proposition E.17.** *If  $\Gamma \vdash A$  and  $\neg A \in \Gamma$ , then  $\Gamma$  is inconsistent.*

*Proof.* Suppose  $\Gamma \vdash A$  and  $\neg A \in \Gamma$ . Then there are  $B_1, \dots, B_n \in \Gamma$  such that

$$\{\mathbb{F} A, \mathbb{T} B_1, \dots, \mathbb{T} B_n\}$$

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  $\Gamma$  is inconsistent, since all assumptions are in  $\Gamma$ .  $\square$

**Proposition E.18.** *If  $\Gamma \cup \{A\}$  and  $\Gamma \cup \{\neg A\}$  are both inconsistent, then  $\Gamma$  is inconsistent.*

*Proof.* If there are  $B_1, \dots, B_n \in \Gamma$  and  $C_1, \dots, C_m \in \Gamma$  such that

$$\{\mathbb{T} A, \mathbb{T} B_1, \dots, \mathbb{T} B_n\} \text{ and } \{\mathbb{T} \neg A, \mathbb{T} C_1, \dots, \mathbb{T} C_m\}$$

both have closed tableaux, we can construct a single, combined tableau that shows that  $\Gamma$  is inconsistent by using as assumptions  $\mathbb{T} B_1, \dots, \mathbb{T} B_n$  together with  $\mathbb{T} C_1, \dots, \mathbb{T} C_m$ , followed by an application of the Cut rule. This yields two branches, one starting with  $\mathbb{T} A$ , the other with  $\mathbb{F} A$ .

On the left left side, add the part of the first tableau below its assumptions. Here, every rule application is still correct, since each of the assumptions of the first tableau, including  $\mathbb{T} A$ , is available. Thus, every branch below  $\mathbb{T} A$  closes.

On the right side, add the part of the second tableau below its assumption, with the results of any applications of  $\neg\mathbb{T}$  to  $\mathbb{T} \neg A$  removed. The conclusion of  $\neg\mathbb{T}$  to  $\mathbb{T} \neg A$  is  $\mathbb{F} A$ , which is nevertheless available, as it is the conclusion of the Cut rule on the right side of the combined tableau.

If a branch in the second tableau was closed because it contained the assumption  $\mathbb{T} \neg A$  (which no longer appears as an assumption in the combined tableau) as well as  $\mathbb{F} \neg A$ , we can applying  $\neg\mathbb{F}$  to  $\mathbb{F} \neg A$  to obtain  $\mathbb{T} A$ . Now the corresponding branch in the combined tableau also closes, because it contains the right-hand conclusion of the Cut rule,  $\mathbb{F} A$ . If a branch in the second tableau closed for any other reason, the corresponding branch in the combined tableau also closes, since any signed formulas other than  $\mathbb{T} \neg A$  occurring on the branch in the old, second tableau also occur on the corresponding branch in the combined tableau.  $\square$

## E.8 Derivability and the Propositional Connectives

We establish that the derivability relation  $\vdash$  of tableaux is strong enough to establish some basic facts involving the propositional connectives, such as that  $A \wedge B \vdash A$  and  $A, A \rightarrow B \vdash B$  (modus ponens). These facts are needed for the proof of the completeness theorem.

**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
	$\swarrow \quad \searrow$	
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.	$\mathbb{F} B$	$\neg \mathbb{T} 3$
	$\swarrow \quad \searrow$	
6.	$\mathbb{T} A \quad \mathbb{T} B$	$\vee \mathbb{T} 1$
	$\otimes \quad \otimes$	

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.  $A, A \rightarrow B \vdash B$ .

2. Both  $\neg A \vdash A \rightarrow B$  and  $B \vdash A \rightarrow B$ .

*Proof.*     1.  $\{\mathbb{F} B, \mathbb{T} A \rightarrow B, \mathbb{T} A\}$  has a closed tableau:

- |    |   |                          |
|----|---|--------------------------|
| 1. | $\mathbb{F} B$  | Assumption               |
| 2. | $\mathbb{T} A \rightarrow B$  | Assumption               |
| 3. | $\mathbb{T} A$  | Assumption               |
|    |  |                          |
| 4. | $\mathbb{F} A \quad \mathbb{T} B$   | $\rightarrow\mathbb{T}2$ |
|    | $\otimes \quad \otimes$   |                          |

2. Both  $\{\mathbb{F} A \rightarrow B, \mathbb{T} \neg A\}$  and  $\{\mathbb{F} A \rightarrow B, \mathbb{T} B\}$  have closed tableaux:

- |    |                              |                          |
|----|------------------------------|--------------------------|
| 1. | $\mathbb{F} A \rightarrow B$ | Assumption               |
| 2. | $\mathbb{T} \neg A$          | Assumption               |
| 3. | $\mathbb{T} A$               | $\rightarrow\mathbb{F}1$ |
| 4. | $\mathbb{F} B$               | $\rightarrow\mathbb{F}1$ |
| 5. | $\mathbb{F} A$               | $\neg\mathbb{T}2$        |
|    | $\otimes$                    |                          |

- |    |                              |                          |
|----|------------------------------|--------------------------|
| 1. | $\mathbb{F} A \rightarrow B$ | Assumption               |
| 2. | $\mathbb{T} B$               | Assumption               |
| 3. | $\mathbb{T} A$               | $\rightarrow\mathbb{F}1$ |
| 4. | $\mathbb{F} B$               | $\rightarrow\mathbb{F}1$ |
|    | $\otimes$                    |                          |

## 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  $v \models A$ , and it *satisfies*  $\mathbb{F} A$  iff  $v \not\models A$ .  $v$  *satisfies* a set of signed formulas  $\Gamma$  iff it satisfies every  $S A \in \Gamma$ .  $\Gamma$  is *satisfiable* if there is a valuation that satisfies it, and *unsatisfiable* otherwise.

**Theorem E.23 (Soundness).** *If  $\Gamma$  has a closed tableau,  $\Gamma$  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  $\Gamma$ . So if  $\Gamma$  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  $\Gamma$  be the set of signed formulas on that branch, and let  $SA \in \Gamma$  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.,  $\Gamma$  together with the conclusions of the rule, is still satisfiable. If the rule results in a split branch, we have to show that at least one of the two resulting branches is satisfiable.

First, we consider the possible inferences that do not result in a split branch.

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  $v \models \Gamma$ . In particular,  $v \models \neg B$ . Thus,  $v \not\models B$ , 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  $v \models \Gamma$ , in particular  $v \models B \wedge C$ . Then  $v \models B$  and  $v \models C$ . This means that  $v$  satisfies both  $\mathbb{T} B$  and  $\mathbb{T} C$ .
4. The branch is expanded by applying  $\vee\mathbb{F}$  to  $\mathbb{F} B \vee C \in \Gamma$ : Exercise.
5. The branch is expanded by applying  $\rightarrow\mathbb{F}$  to  $\mathbb{F} B \rightarrow C \in \Gamma$ : This results in two new signed formulas on the branch:  $\mathbb{T} B$  and  $\mathbb{F} C$ . Suppose  $v \models \Gamma$ , in particular  $v \models B \rightarrow C$ . Then  $v \models B$  and  $v \models C$ . This means that  $v$  satisfies both  $\mathbb{T} B$  and  $\mathbb{F} C$ .

Now let's consider the possible inferences that result in a split branch.

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  $v \models \Gamma$ , in particular  $v \models B \wedge C$ . Then  $v \models B$  or  $v \models C$ . 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  $v \models \Gamma$  and either  $v \models B$  or  $v \models \neg B$ ,  $v$  satisfies either the left or the right branch.  $\square$

**Corollary E.24.** *If  $\Gamma \vdash A$  then  $A$  is a tautology.*

**Corollary E.25.** *If  $\Gamma \vdash A$  then  $\Gamma \models A$ .*

*Proof.* If  $\Gamma \vdash A$  then for some  $B_1, \dots, B_n \in \Gamma$ ,  $\{\mathbb{F} A, \mathbb{T} B_1, \dots, \mathbb{T} B_n\}$  has a closed tableau. By [Theorem E.23](#), every valuation  $v$  either makes some  $B_i$  false or makes  $A$  true. Hence, if  $v \models \Gamma$  then also  $v \models A$ .  $\square$

**Corollary E.26.** *If  $\Gamma$  is satisfiable, then it is consistent.*

*Proof.* We prove the contrapositive. Suppose that  $\Gamma$  is not consistent. Then there are  $B_1, \dots, B_n \in \Gamma$  and a closed tableau for  $\{\mathbb{T} B_1, \dots, \mathbb{T} B_n\}$ . By [Theorem E.23](#), there is no  $v$  such that  $v \models B_i$  for all  $i = 1, \dots, n$ . But then  $\Gamma$  is not satisfiable.  $\square$

## Problems

**Problem E.1.** Give closed tableaux of the following:

1.  $\mathbb{T} A \wedge (B \wedge C), \mathbb{F} (A \wedge B) \wedge C$ .
2.  $\mathbb{T} A \vee (B \vee C), \mathbb{F} (A \vee B) \vee C$ .
3.  $\mathbb{T} A \rightarrow (B \rightarrow C), \mathbb{F} B \rightarrow (A \rightarrow C)$ .
4.  $\mathbb{T} A, \mathbb{F} \neg \neg A$ .

**Problem E.2.** Give closed tableaux of the following:

1.  $\mathbb{T} (A \vee B) \rightarrow C, \mathbb{F} A \rightarrow C$ .
2.  $\mathbb{T} (A \rightarrow C) \wedge (B \rightarrow C), \mathbb{F} (A \vee B) \rightarrow C$ .
3.  $\mathbb{F} \neg(A \wedge \neg A)$ .
4.  $\mathbb{T} B \rightarrow A, \mathbb{F} \neg A \rightarrow \neg B$ .

5.  $\mathbb{F} (A \rightarrow \neg A) \rightarrow \neg A.$
6.  $\mathbb{F} \neg(A \rightarrow B) \rightarrow \neg B.$
7.  $\mathbb{T} A \rightarrow C, \mathbb{F} \neg(A \wedge \neg C).$
8.  $\mathbb{T} A \wedge \neg C, \mathbb{F} \neg(A \rightarrow C).$
9.  $\mathbb{T} A \vee B, \neg B, \mathbb{F} A.$
10.  $\mathbb{T} \neg A \vee \neg B, \mathbb{F} \neg(A \wedge B).$
11.  $\mathbb{F} (\neg A \wedge \neg B) \rightarrow \neg(A \vee B).$
12.  $\mathbb{F} \neg(A \vee B) \rightarrow (\neg A \wedge \neg B).$

**Problem E.3.** Give closed tableaux of the following:

1.  $\mathbb{T} \neg(A \rightarrow B), \mathbb{F} A.$
2.  $\mathbb{T} \neg(A \wedge B), \mathbb{F} \neg A \vee \neg B.$
3.  $\mathbb{T} A \rightarrow B, \mathbb{F} \neg A \vee B.$
4.  $\mathbb{F} \neg \neg A \rightarrow A.$
5.  $\mathbb{T} A \rightarrow B, \mathbb{T} \neg A \rightarrow B, \mathbb{F} B.$
6.  $\mathbb{T} (A \wedge B) \rightarrow C, \mathbb{F} (A \rightarrow C) \vee (B \rightarrow C).$
7.  $\mathbb{T} (A \rightarrow B) \rightarrow A, \mathbb{F} A.$
8.  $\mathbb{F} (A \rightarrow B) \vee (B \rightarrow C).$

**Problem E.4.** Prove [Proposition E.13](#)

**Problem E.5.** Prove that  $\Gamma \vdash \neg A$  iff  $\Gamma \cup \{A\}$  is inconsistent.

**Problem E.6.** 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 derivation system: if a sentence  $A$  follows from some sentences  $\Gamma$ , then there is also a derivation that establishes  $\Gamma \vdash A$ . Thus, the derivation 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 derivation system is unable to produce certain derivations. But who's to say that just because there are no derivations of a certain sort from  $\Gamma$ , it's guaranteed that there is valuation  $v$  with  $v \models \Gamma$ ? Before the completeness theorem was first proved—in fact before we had the

derivation 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  $\Gamma$ , it follows by the completeness theorem that if  $A$  is a consequence of  $\Gamma$ , it is already a consequence of a finite subset of  $\Gamma$ . This is called *compactness*. Equivalently, if every finite subset of  $\Gamma$  is consistent, then  $\Gamma$  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  $\Gamma$  is consistent, it is satisfiable.” Perhaps we can use the information in  $\Gamma$  together with the hypothesis that it is consistent to construct a valuation that satisfies every formula in  $\Gamma$ . After all, we know what kind of valuation we are looking for: one that is as  $\Gamma$  describes it!

If  $\Gamma$  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  $\Gamma$  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  $\Gamma$  comes in: if  $\neg B \in \Gamma$ , then  $B \notin \Gamma$ , or else  $\Gamma$  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  $\Gamma$  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  $\Gamma$ . 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  $\Gamma$ .

This suggests the following idea: we add additional formulas to  $\Gamma$  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  $\Gamma^*$ . Then our construction above would provide us with a valuation  $v$  for which we could prove, by induction, that it satisfies all sentences in  $\Gamma^*$ , and hence also all sentence in  $\Gamma$  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  $\Gamma$  to a consistent and complete set  $\Gamma^*$ .

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  $\Gamma$  and show that it can be extended to a consistent and complete set  $\Gamma^*$  (Lemma F.3). This set  $\Gamma^*$  is what we'll use to define our valuation  $v(\Gamma^*)$ . The valuation is determined by the propositional variables in  $\Gamma^*$  (Definition F.4). We'll use the properties of complete consistent sets to show that indeed  $v(\Gamma^*) \models A$  iff  $A \in \Gamma^*$  (Lemma F.5), and thus in particular,  $v(\Gamma^*) \models \Gamma$ .

### F.3 Complete Consistent Sets of Sentences

**Definition F.1 (Complete set).** A set  $\Gamma$  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  $A$ ,  $\Gamma$  “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  $\Gamma$  is contained in a complete consistent set  $\Gamma^*$ . 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  $\Gamma^*$ . This valuation can then be shown to make all sentences in  $\Gamma^*$  (and hence also all those in  $\Gamma$ ) 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  $\Gamma$  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.  *$A \rightarrow B \in \Gamma$  iff either  $A \notin \Gamma$  or  $B \in \Gamma$ .*

*Proof.* Let us suppose for all of the following that  $\Gamma$  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  $\Gamma$  is complete,  $\neg A \in \Gamma$ . By [Propositions D.24](#) and [E.17](#),  $\Gamma$  is inconsistent. This contradicts the assumption that  $\Gamma$  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  $\Gamma$  is complete,  $\neg A \in \Gamma$  and  $\neg B \in \Gamma$ . By [Propositions D.27](#) and [E.20](#), item (1),  $\Gamma$  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 D.27** and **E.20**, 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 make sure at each stage not to introduce an inconsistency.

**Lemma F.3 (Lindenbaum's Lemma).** *Every consistent set  $\Gamma$  in a language  $\mathcal{L}$  can be extended to a complete and consistent set  $\Gamma^*$ .*

*Proof.* Let  $\Gamma$  be consistent. Let  $A_0, A_1, \dots$  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  $\Gamma_n$  is consistent:  $\Gamma_0$  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  $\Gamma_n$  would be inconsistent by **Propositions D.25** and **E.18**, 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 show that if  $i < n + 1$  then  $\Gamma_i \subseteq \Gamma_{n+1}$ . 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  $i < n + 1$ , then  $\Gamma_i \subseteq \Gamma_n$  by inductive hypothesis (if  $i < n$ ) or the trivial fact that  $\Gamma_n \subseteq \Gamma_n$  (if  $i = n$ ). We get that  $\Gamma_i \subseteq \Gamma_{n+1}$  by transitivity of  $\subseteq$ .

From this it follows that  $\Gamma^*$  is consistent. Here's why: Let  $\Gamma' \subseteq \Gamma^*$  be finite. Each  $B \in \Gamma'$  is also in  $\Gamma_i$  for some  $i$ . Let  $n$  be the largest of these. Since  $\Gamma_i \subseteq \Gamma_n$  if  $i \leq n$ , every  $B \in \Gamma'$  is also  $\in \Gamma_n$ , i.e.,  $\Gamma' \subseteq \Gamma_n$ , and  $\Gamma_n$  is consistent. So, every finite subset  $\Gamma' \subseteq \Gamma^*$  is consistent. By [Propositions D.18](#) and [E.14](#),  $\Gamma^*$  is consistent.

Every sentence of  $\text{Frm}(\mathcal{L})$  appears on the list used to define  $\Gamma^*$ . If  $A_n \notin \Gamma^*$ , then that is because  $\Gamma_n \cup \{A_n\}$  was inconsistent. But then  $\neg A_n \in \Gamma^*$ , so  $\Gamma^*$  is complete.  $\square$

## 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  $\Gamma^*$  determine  $v(\Gamma^*)$ .

**Definition F.4.** Suppose  $\Gamma^*$  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).**  $v(\Gamma^*) \models A$  iff  $A \in \Gamma^*$ .

*Proof.* We prove both directions simultaneously, and by induction on  $A$ .

1.  $A \equiv \perp$ :  $v(\Gamma^*) \not\models \perp$  by definition of satisfaction. On the other hand,  $\perp \notin \Gamma^*$  since  $\Gamma^*$  is consistent.
2.  $A \equiv p$ :  $v(\Gamma^*) \models p$  iff  $v(\Gamma^*)(p) = \mathbb{T}$  (by the definition of satisfaction) iff  $p \in \Gamma^*$  (by the construction of  $v(\Gamma^*)$ ).
3.  $A \equiv \neg B$ :  $v(\Gamma^*) \models A$  iff  $v(\Gamma^*) \not\models B$  (by definition of satisfaction). By induction hypothesis,  $v(\Gamma^*) \not\models B$  iff  $B \notin \Gamma^*$ . Since  $\Gamma^*$  is consistent and complete,  $B \notin \Gamma^*$  iff  $\neg B \in \Gamma^*$ .
4.  $A \equiv B \wedge C$ : exercise.
5.  $A \equiv B \vee C$ :  $v(\Gamma^*) \models A$  iff  $v(\Gamma^*) \models B$  or  $v(\Gamma^*) \models C$  (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.  $A \equiv B \rightarrow C$ : exercise.

## F.6 The Completeness Theorem

Let's combine our results: we arrive at the completeness theorem.

**Theorem F.6 (Completeness Theorem).** *Let  $\Gamma$  be a set of sentences. If  $\Gamma$  is consistent, it is satisfiable.*

*Proof.* Suppose  $\Gamma$  is consistent. By [Lemma F.3](#), there is a  $\Gamma^* \supseteq \Gamma$  which is consistent and complete. By [Lemma F.5](#),  $v(\Gamma^*) \models A$  iff  $A \in \Gamma^*$ . From this it follows in particular that for all  $A \in \Gamma$ ,  $v(\Gamma^*) \models A$ , so  $\Gamma$  is satisfiable.  $\square$

**Corollary F.7 (Completeness Theorem, Second Version).**

*For all  $\Gamma$  and sentences  $A$ : if  $\Gamma \models A$  then  $\Gamma \vdash A$ .*

*Proof.* Note that the  $\Gamma$ '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  $\Delta$ , if  $\Delta$  is consistent, it is satisfiable. By contraposition, if  $\Delta$  is not satisfiable, then  $\Delta$  is inconsistent. We will use this to prove the corollary.

Suppose that  $\Gamma \models A$ . Then  $\Gamma \cup \{\neg A\}$  is unsatisfiable by [Proposition C.21](#). Taking  $\Gamma \cup \{\neg A\}$  as our  $\Delta$ , the previous version of [Theorem F.6](#) gives us that  $\Gamma \cup \{\neg A\}$  is inconsistent. By [Propositions D.23](#) and [E.16](#),  $\Gamma \vdash A$ .  $\square$

## 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](http://openlogicproject.org) for additional information.