What if?

An Open Introduction to Non-Classical Logics

F20

What if?

What if?

An Open Introduction to Non-Classical Logics

Remixed by Audrey Yap & Richard Zach

Fall 2020

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.

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

Typeset in Baskervald X and Nimbus Sans by LaTeX.

This version of *What if?* is revision 8be15d6 (2021-07-11), with content generated from *Open Logic Text* revision 4772e75 (2024-08-29). Free download at:

https://builds.openlogicproject.org/courses/what-if/

Contents

Preface

This is an introductory textbook on non-classical logics. Or rather, it will be when it's done. Right now it's a work in progress. We will use it as the main text in our courses on non-classical logic at the Universities of Victoria and Calgary, respectively, in Fall 2020. 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. The textbook *Sets, Logic, Computation*, is also based on the OLP, provides this background. But the required material is also included here: the basics of classical propositional logic in part I, and the material on set theory, and some introductory material on proofs and induction, in part X.

Introduction

Classical logic is very useful, widely used, has a long history, and is relatively simple. But it has limitations: for instance, it does not (and cannot) deal well with certain locutions of natural language such as tense and subjunctive mood, nor with certain constructions such as "Audrey knows that p." It makes certain assumptions, for instance that every sentence is either true or false and never both. It pronounces some formulas tautologies and some arguments as valid, even though these tautologies and arguments formalize arguments in English which some do not consider true or valid, at least not obviously. Thus it seems there are examples where classical logic is not expressive enough, or even where classical logic gets things wrong.

This book discusses some alternative, *non-classical* logics. These non-classical logics are either more expressive than classical logic or have different tautologies or valid arguments. For instance, temporal logic extends classical logic by operators that express tense; conditional logics have an additional, different conditional ("if—then") that does not suffer from the so-called paradoxes of the material conditional. All of these logics *extend* classical logic by new operators or connectives, and fall into the broad category of intensional logics. Other logics such as many-valued, intuitionistic, and paraconsistent logics have the same basic connectives as classical logic, but different infrences count as valid. In many-valued and intutionistic logic, for instance, the law of excluded middle $A \lor \neg A$ fails to hold; in paraconsistent logic the inference *ex contradictione quodlibet*, $\bot \vDash A$ for arbitrary A.

After we review the basics of classical propositional logic in part I, we begin our discussion of non-classical logics in II. There we will relax one assumption classical logic makes: that everything is either true or false. There are good reasons to think that some sentences of English are neither—they have some intermediate truth value. Examples of this are sentences involving vagueness ("Mary is rich"), sentences the truth of which is not yet determined ("There will be a sea battle tomorrow"), and—an important case for philosopers—sentences that are paradoxical such as "This sentence is false." One of the earliest non-classical logics allow truth values in addition to the classical "true" and "false"; they are called *many-valued*. We cover them in part II.

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

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 interpretation, 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: $M, w \Vdash \Box A$ iff $M, v \Vdash A$ for all worlds v. This corresponds to Leibniz's idea that what's necessarily true is what's true in every possible world.

"Necessarily" is not the only way to interpret the $\Box$ operator, but it is the standard one—"necessarily" and "possibly" are the so-called *alethic* modalities. Other interpretations read $\Box$ as "it is known (by some person A) that," as "some person A believes that," "it ought to be the case that," or "it will always be true that." These are epistemic, doxastic, deontic, and temporal modalities, respectively. Different interpretations of $\Box$ will make different formulas 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. We discuss modal logics in general in parts III and IV and epistemic logics in particular in part V.

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 $M, w \Vdash \Box A$ if $M, v \Vdash A$ for all worlds v which are accessible from w. The resulting semantics is very versatile and powerful, and the basic idea can be used to provide semantic interpretations for logics based on other intensional operators. One such logic is a close relative of modal logic called temporal logic. Instead of having just one modality $\Box$ (plus its dual $\Diamond$), it has *temporal operators* such as "always P," "p will be true", etc. We study these in part VI.

Whereas the material conditional is best read as an English indicative conditional ("If p is true then q is true"), subjunctive conditionals are in the subjunctive mood: "if p were true then q would be true." While a material conditional with a false antecedent is true, a subjunctive conditional need not be, e.g., "if humans had tails, they would be able to fly." In part VII, we discuss logics of counterfactual coditionals.

Intuitionistic logic is 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. As mentioned above, intuitionistic logic is non-classical because it has fewer valid inferences and theorems, e.g., $A \lor \neg A$ and

$\neg\neg A \rightarrow A$ fail in general. Intuitively, this is a consequence of the intuitionist principle that something shouldn't count as true—you should not assert it—unless you have a proof of it. And obviously there are cases where we neither have a proof of A nor a proof of $\neg A$. Intuitionistic logic can be given a relational semantics very much like modal logic. We discuss it in part VIII.

One of the weirdest features of classical logic is the principle of explosion: from a contradiction, anything follows. This principle flows from the way we set up the semantics of classical logic, but it is *very* counterintuitive and goes against what we actually do when we reason. After all, once you discover that some things you believe are contradictory, you don't (usually) go on to conclude arbitrary claims since they follow from your beliefs! This has led logicians to develop systems of logic in which the inference *ex contradictione, quodlibet* is blocked. Some of these are simply further weakenings of classical logic, designed just to get rid of explosion. Some are more philosophically motivated. Part of what makes the principle of explosion weird is that when it pronounces that A and $\neg A$ together entail B, there is no connection at all between the premises and the conclusion. And shouldn't there be such a connection in any valid argument? Shouldn't the premises be *relevant* to the conclusion? This leads to so-valled *relevant* (or *relevance*) logic. Another motivation is the philosophical position called *dialetheism:* the belief that there can be true contradictions. (If you believe that contradictions can be true, then you would not want them to entail anything whatsoever, e.g., something false.) All of these logic fall under the umbrella term *paraconsistent.* We discuss paraconsistent logics in part IX.

PART I

Remind me, how does logic work again?

CHAPTER 1

Syntax and Semantics

1.1 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 $\overline{v}(A)$ for any formula A. A formula is satisfied in a valuation v iff $\overline{v}(A) = \mathbb{T}$—we write this as $v \vDash A$. This relation can also be defined by induction on the structure of A, using the truth functions for the logical connectives to define, say, satisfaction of $A \land B$ in terms of satisfaction (or not) of A and B.

On the basis of the satisfaction relation $v \vDash A$ for sentences we can then define the basic semantic notions of tautology, entailment, and satisfiability. A formula is a tautology, $\vDash A$, if every valuation satisfies it, i.e., $\overline{v}(A) = \mathbb{T}$ for any v. It is entailed by a set of formulas, $\Gamma \vDash A$, if every valuation that satisfies all the formulas in Γ also satisfies A. And a set of formulas is satisfiable if some valuation satisfies all formulas in it at the same time. Because formulas are inductively defined, and satisfaction is in turn defined by induction on the structure of formulas, we can use induction to prove properties of our semantics and to relate the semantic notions defined.

1.2 Propositional Formulas

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

1. A countably infinite set At_0 of propositional variables p_0, p_1, ...

2. The propositional constant for falsity $\bot$.

3. The logical connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional)

4. Punctuation marks: (,), and the comma.

We denote this language of propositional logic by $\mathcal{L}_0$.

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 $\bot$ symbol is variously called "falsity," "falsum," "absurdity," or "bottom." The $\top$ symbol is variously called "truth," "verum," or "top."

Definition 1.1 (Formula). The set $\mathrm{Frm}(\mathcal{L}_0)$ of *formulas* of propositional logic is defined inductively as follows:

1. $\bot$ is an atomic formula.
2. Every propositional variable p_i is an atomic formula.
3. If A is a formula, then $\neg A$ is 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 $\bot$, p_i. "Atomic formula" thus means any formula of this form.

The other cases of the definition give rules for constructing new formulas out of formulas already constructed. At the second stage, we can use them to construct formulas out of atomic formulas. At the third stage, we construct new formulas from the atomic formulas and those obtained in the second stage, and so on. A formula is anything that is eventually constructed at such a stage, and nothing else.

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 1.2. Formulas constructed using the defined operators are to be understood as follows:

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

Definition 1.3 (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., $A \equiv (B \lor C)$ means: the string of symbols A is the same string as the one obtained by concatenating an opening parenthesis, the string B, the $\lor$ 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 $\lor$, etc.

1.3 Preliminaries

Theorem 1.4 (*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 $(A \land B)$ whenever it holds for A and B;*
3. *it holds for $(A \lor B)$ whenever it holds for A and B;*
4. *it holds for $(A \to B)$ whenever it holds for A and B;*

then P holds for all formulas.

Proof. Let S be the collection of all formulas with property P. Clearly $S \subseteq \mathrm{Frm}(\mathcal{L}_0)$. S satisfies all the conditions of Definition 1.1: it contains all atomic formulas and is closed under the logical operators. $\mathrm{Frm}(\mathscr{L}_0)$ is the smallest such class, so $\mathrm{Frm}(\mathscr{L}_0) \subseteq S$. So $\mathrm{Frm}(\mathscr{L}_0) = S$, and every formula has property P. □

Proposition 1.5. *Any formula in* $\mathrm{Frm}(\mathscr{L}_0)$ *is* balanced, *in that it has as many left parentheses as right ones.*

Proposition 1.6. *No proper initial segment of a formula is a formula.*

Proposition 1.7 (Unique Readability). *Any formula A in* $\mathrm{Frm}(\mathscr{L}_0)$ *has exactly one parsing as one of the following*

1. $\bot$.
2. p_n *for some* $p_n \in \mathrm{At}_0$.
3. $\neg B$ *for some formula* B.
4. $(B \wedge C)$ *for some formulas* B *and* C.
5. $(B \vee C)$ *for some formulas* B *and* C.
6. $(B \rightarrow C)$ *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. □

Definition 1.8 (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 $A[B_1/p_1, \dots, B_n/p_n]$.

1.4 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 1.9 (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 1.10.

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

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 1.11. *Every formula A in* $\mathrm{Frm}(\mathscr{L}_0)$ *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 $A \equiv (B \land C)$, then $\langle B_0, \dots, B_n, C_0, \dots, C_m, (B_n \land C_m) \rangle$ is a formation sequence for A.
3. If $A \equiv (B \lor C)$, then $\langle B_0, \dots, B_n, C_0, \dots, C_m, (B_n \lor C_m) \rangle$ is a formation sequence for A.
4. If $A \equiv (B \to C)$, then $\langle B_0, \dots, B_n, C_0, \dots, C_m, (B_n \to C_m) \rangle$ is a formation sequence for A.

By the principle of induction on formulas, every formula has a formation sequence. □

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 1.12. *Suppose that $\langle A_0, \dots, A_n \rangle$ is a formation sequence for A_n, and that $k \le n$. Then $\langle A_0, \dots, A_k \rangle$ is a formation sequence for A_k.*

Proof. Exercise. □

Theorem 1.13. $\mathrm{Frm}(\mathscr{L}_0)$ *is the set of all strings of symbols in the language* $\mathscr{L}_0$ *with a formation sequence.*

Proof. Let F be the set of all strings of symbols in the language $\mathscr{L}_0$ that have a formation sequence. We have seen in Proposition 1.11 that $\mathrm{Frm}(\mathscr{L}_0) \subseteq F$, so now we prove the converse.

Suppose A has a formation sequence $\langle A_0, \dots, A_n \rangle$. We prove that $A \in \mathrm{Frm}(\mathscr{L}_0)$ by strong induction on n. Our induction hypothesis is that every string of symbols with a formation sequence of length $m < n$ is in $\mathrm{Frm}(\mathscr{L}_0)$. 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. $A_n \equiv (A_j \to A_k)$.

Now we reason by cases. If A_n is atomic then $A_n \in \mathrm{Frm}(\mathscr{L}_0)$. Suppose instead that $A \equiv (A_j \wedge A_k)$. By Lemma 1.12, $\langle A_0, \dots, A_j \rangle$ 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 $\mathrm{Frm}(\mathscr{L}_0)$, and so by the definition of a formula, so is $(A_j \wedge A_k)$. The other cases follow by parallel reasoning. □

1.5 Valuations and Satisfaction

Definition 1.14 (Valuations). Let $\{\mathbb{T}, \mathbb{F}\}$ be the set of the two truth values, "true" and "false." A *valuation* for $\mathscr{L}_0$ is a function v assigning either $\mathbb{T}$ or $\mathbb{F}$ to the propositional variables of the language, i.e., $v \colon \mathrm{At}_0 \to \{\mathbb{T}, \mathbb{F}\}$.

Definition 1.15. Given a valuation v, define the evaluation function $\overline{v}\colon \mathrm{Frm}(\mathscr{L}_0) \to \{\mathbb{T}, \mathbb{F}\}$ inductively by:

$$\overline{v}(\bot) = \mathbb{F};$$
$$\overline{v}(p_n) = v(p_n);$$
$$\overline{v}(\neg A) = \begin{cases} \mathbb{T} & \text{if } \overline{v}(A) = \mathbb{F}; \\ \mathbb{F} & \text{otherwise.} \end{cases}$$
$$\overline{v}(A \wedge B) = \begin{cases} \mathbb{T} & \text{if } \overline{v}(A) = \mathbb{T} \text{ and } \overline{v}(B) = \mathbb{T}; \\ \mathbb{F} & \text{if } \overline{v}(A) = \mathbb{F} \text{ or } \overline{v}(B) = \mathbb{F}. \end{cases}$$
$$\overline{v}(A \vee B) = \begin{cases} \mathbb{T} & \text{if } \overline{v}(A) = \mathbb{T} \text{ or } \overline{v}(B) = \mathbb{T}; \\ \mathbb{F} & \text{if } \overline{v}(A) = \mathbb{F} \text{ and } \overline{v}(B) = \mathbb{F}. \end{cases}$$
$$\overline{v}(A \to B) = \begin{cases} \mathbb{T} & \text{if } \overline{v}(A) = \mathbb{F} \text{ or } \overline{v}(B) = \mathbb{T}; \\ \mathbb{F} & \text{if } \overline{v}(A) = \mathbb{T} \text{ and } \overline{v}(B) = \mathbb{F}. \end{cases}$$

The clauses correspond to the following truth tables:

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

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

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

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

Theorem 1.16 (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 1.17 (Satisfaction). We can inductively define the notion of *satisfaction of a formula A by a valuation v*, $v \vDash A$, as follows. (We write $v \nvDash A$ to mean "not $v \vDash A$.")

1. $A \equiv \bot$: $v \nvDash A$.

2. $A \equiv p_i$: $v \vDash A$ iff $v(p_i) = \mathbb{T}$.

3. $A \equiv \neg B$: $v \vDash A$ iff $v \nvDash B$.

4. $A \equiv (B \wedge C)$: $v \vDash A$ iff $v \vDash B$ and $v \vDash C$.

5. $A \equiv (B \vee C)$: $v \vDash A$ iff $v \vDash B$ or $v \vDash C$ (or both).

6. $A \equiv (B \rightarrow C)$: $v \vDash A$ iff $v \nvDash B$ or $v \vDash C$ (or both).

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

Proposition 1.18. *$v \vDash A$ iff $\overline{v}(A) = \mathbb{T}$.*

Proof. By induction on A. □

1.6 Semantic Notions

We define the following semantic notions:

Definition 1.19.

1. A formula A is *satisfiable* if for some v, $v \vDash A$; it is *unsatisfiable* if for no v, $v \vDash A$;

2. A formula A is a *tautology* if $v \vDash A$ for all valuations v;

3. A formula A is *contingent* if it is satisfiable but not a tautology;

4. If Γ is a set of formulas, $\Gamma \vDash A$ ("Γ entails A") if and only if $v \vDash A$ for every valuation v for which $v \vDash \Gamma$.

5. If Γ is a set of formulas, Γ is *satisfiable* if there is a valuation v for which $v \vDash \Gamma$, and Γ is *unsatisfiable* otherwise.

Proposition 1.20.

1. *A is a tautology if and only if $\emptyset \vDash A$;*
2. *If $\Gamma \vDash A$ and $\Gamma \vDash A \rightarrow B$ then $\Gamma \vDash B$;*
3. *If Γ is satisfiable then every finite subset of Γ is also satisfiable;*
4. *Monotonicity: if $\Gamma \subseteq \Delta$ and $\Gamma \vDash A$ then also $\Delta \vDash A$;*
5. *Transitivity: if $\Gamma \vDash A$ and $\Delta \cup \{A\} \vDash B$ then $\Gamma \cup \Delta \vDash B$.*

Proof. Exercise. □

Proposition 1.21. *$\Gamma \vDash A$ if and only if $\Gamma \cup \{\neg A\}$ is unsatisfiable.*

Proof. Exercise. □

Theorem 1.22 (Semantic Deduction Theorem). *$\Gamma \vDash A \rightarrow B$ if and only if $\Gamma \cup \{A\} \vDash B$.*

Proof. Exercise. □

Problems

Problem 1.1. Prove Proposition 1.5

Problem 1.2. Prove Proposition 1.6

Problem 1.3. 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) $(\neg p_0 \land p_1)$; and (iii) $((\neg p_0 \to p_1) \land p_2)$. In each case specify the relevant substitution.

1. p_1
2. $(\neg p_0 \land p_0)$
3. $((p_0 \lor p_1) \land p_2)$
4. $\neg((p_0 \to p_1) \land p_2)$
5. $((\neg(p_0 \to p_1) \to (p_0 \lor p_1)) \land \neg(p_0 \land p_1))$

Problem 1.4. Give a mathematically rigorous definition of $A[B/p]$ by induction.

Problem 1.5. Consider adding to $\mathscr{L}_0$ a ternary connective $\diamond$ with evaluation given by

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

Write down the truth table for this connective.

Problem 1.6. Prove Proposition 1.18

Problem 1.7. For each of the following four formulas determine whether it is (a) satisfiable, (b) tautology, and (c) contingent.

1. $(p_0 \to (\neg p_1 \to \neg p_0))$.
2. $((p_0 \land \neg p_1) \to (\neg p_0 \land p_2)) \leftrightarrow ((p_2 \to p_0) \to (p_0 \to p_1))$.
3. $(p_0 \leftrightarrow p_1) \to (p_2 \leftrightarrow \neg p_1)$.
4. $((p_0 \leftrightarrow (\neg p_1 \land p_2)) \lor (p_2 \to (p_0 \leftrightarrow p_1)))$.

Problem 1.8. Prove Proposition 1.20

Problem 1.9. Prove Proposition 1.21

Problem 1.10. Prove Theorem 1.22

CHAPTER 2

Axiomatic Derivations

2.1 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 Γ." However $\vdash$ is defined, we want it to match up with $\vDash$, that is:

1. $\vdash A$ if and only if $\vDash A$

2. $\Gamma \vdash A$ if and only if $\Gamma \vDash A$

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 $\Gamma \vDash A$. Completeness is harder to establish, and some logics have no complete derivation systems. First-order logic does. Kurt Gödel was the first one to prove completeness for a derivation system of first-order logic in his 1929 dissertation.

Another concept that is connected to derivation systems is that of *consistency*. A set of sentences is called inconsistent if anything whatsoever can be derived from it, and consistent otherwise. Inconsistency is the syntactic counterpart to unsatisfiablity: like unsatisfiable sets, inconsistent sets of sentences do not make good theories, they are defective in a fundamental way. Consistent sets of sentences may not be true or useful, but at least they pass that minimal threshold of logical usefulness. For different derivation systems the specific definition of consistency of sets of sentences might differ, but like $\vdash$, we want consistency to coincide with its semantic counterpart, satisfiability. We want it to always be the case that Γ is consistent if and only if it is satisfiable. Here, the "if" direction amounts to completeness (consistency guarantees satisfiability), and the "only if" direction amounts to soundness (satisfiability guarantees consistency). In fact, for classical first-order logic, the two versions of soundness and completeness are equivalent.

2.2 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 Γ 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

$$A \rightarrow (B \rightarrow A) \qquad B \rightarrow (B \lor C) \qquad (B \land C) \rightarrow B$$

are common axioms that govern $\rightarrow$, $\lor$ and $\land$. 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 Γ is taken as the set of sentences in that derivation which are justified by (2) above). A is a theorem if A has a derivation where Γ is empty, i.e., every sentence in the derivation is justified either by (1) or (3). For instance, here is a derivation that shows that $\vdash A \rightarrow (B \rightarrow (B \lor A))$:

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

The sentence on line 1 is of the form of the axiom $A \rightarrow (A \lor B)$ (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 $B \rightarrow (B \lor A)$, i.e., line 1.

A set Γ is inconsistent if $\Gamma \vdash \bot$. A complete axiom system will also prove that $\bot \to A$ for any A, and so if Γ is inconsistent, then $\Gamma \vdash A$ for any A.

Systems of axiomatic derivations for logic were first given by Gottlob Frege in his 1879 *Begriffsschrift*, which for this reason is often considered the first work of modern logic. They were perfected in Alfred North Whitehead and Bertrand Russell's *Principia Mathematica* and by David Hilbert and his students in the 1920s. They are thus often called "Frege systems" or "Hilbert systems." They are very versatile in that it is often easy to find an axiomatic system for a logic. Because derivations have a very simple structure and only one or two inference rules, it is also relatively easy to prove things *about* them. However, they are very hard to use in practice, i.e., it is difficult to find and write proofs.

2.3 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 2.1 (Derivability). If Γ is a set of formulas of $\mathcal{L}$ then a *derivation* from Γ 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 2.2 (Rule of inference). A *rule of inference* gives a sufficient condition for what counts as a correct inference step in a derivation from Γ.

For instance, since any one-element sequence A with $A \in \Gamma$ trivially counts as a derivation, the following might be a very simple rule of inference:

> If $A \in \Gamma$, then A is always a correct inference step in any derivation from Γ.

Similarly, if A is one of the axioms, then A by itself is a derivation, and so this is also a rule of inference:

> If A is an axiom, then A is a correct inference step.

It gets more interesting if the rule of inference appeals to formulas that appear before the step considered. The following rule is called *modus ponens:*

> If $B \to A$ and B occur higher up in the derivation, then A is a correct inference step.

If this is the only rule of inference, then our definition of derivation above amounts to this: $A_1, \ldots, A_n$ is a derivation iff for each $i \leq n$ one of the following holds:

1. $A_i \in \Gamma$; or
2. A_i is an axiom; or
3. for some $j < i$, A_j is $B \to A_i$, and for some $k < i$, A_k is B.

The last clause says that A_i follows from A_j (B) and A_k ($B \to A_i$) by modus ponens. If we can go from 1 to n, and each time we find a formula A_i that is either in Γ, an axiom, or which a rule of inference tells us that it is a correct inference step, then the entire sequence counts as a correct derivation.

Definition 2.3 (Derivability). A formula A is *derivable* from Γ, written $\Gamma \vdash A$, if there is a derivation from Γ ending in A.

Definition 2.4 (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.

2.4 Axiom and Rules for the Propositional Connectives

Definition 2.5 (Axioms). The set of Ax_0 of *axioms* for the propositional connectives comprises all formulas of the following forms:

$$(A \land B) \to A \tag{2.1}$$
$$(A \land B) \to B \tag{2.2}$$
$$A \to (B \to (A \land B)) \tag{2.3}$$
$$A \to (A \lor B) \tag{2.4}$$
$$A \to (B \lor A) \tag{2.5}$$
$$(A \to C) \to ((B \to C) \to ((A \lor B) \to C)) \tag{2.6}$$
$$A \to (B \to A) \tag{2.7}$$
$$(A \to (B \to C)) \to ((A \to B) \to (A \to C)) \tag{2.8}$$
$$(A \to B) \to ((A \to \neg B) \to \neg A) \tag{2.9}$$
$$\neg A \to (A \to B) \tag{2.10}$$
$$\top \tag{2.11}$$

$$\bot \to A \tag{2.12}$$

$$(A \to \bot) \to \neg A \tag{2.13}$$

$$\neg\neg A \to A \tag{2.14}$$

Definition 2.6 (Modus ponens). If B and $B \to A$ already occur in a derivation, then A is a correct inference step.

We'll abbreviate the rule modus ponens as "MP."

2.5 Examples of Derivations

Example 2.7. Suppose we want to prove $(\neg D \lor E) \to (D \to E)$. Clearly, this is not an instance of any of our axioms, so we have to use the MP rule to derive it. Our only rule is MP, which given A and $A \to B$ allows us to justify B. One strategy would be to use eq. (2.6) with A being $\neg D$, B being E, and C being $D \to E$, i.e., the instance

$$(\neg D \to (D \to E)) \to ((E \to (D \to E)) \to ((\neg D \lor E) \to (D \to E))).$$

Why? Two applications of MP yield the last part, which is what we want. And we easily see that $\neg D \to (D \to E)$ is an instance of eq. (2.10), and $E \to (D \to E)$ is an instance of eq. (2.7). So our derivation is:

1.	$\neg D \to (D \to E)$	eq. (2.10)
2.	$(\neg D \to (D \to E)) \to$ $((E \to (D \to E)) \to ((\neg D \lor E) \to (D \to E)))$	eq. (2.6)
3.	$((E \to (D \to E)) \to ((\neg D \lor E) \to (D \to E))$	1, 2, MP
4.	$E \to (D \to E)$	eq. (2.7)
5.	$(\neg D \lor E) \to (D \to E)$	3, 4, MP

Example 2.8. Let's try to find a derivation of $D \to D$. It is not an instance of an axiom, so we have to use MP to derive it. eq. (2.7) is an axiom of the form $A \to B$ to which we could apply MP. To

be useful, of course, the B which MP would justify as a correct step in this case would have to be $D \to D$, since this is what we want to derive. That means A would also have to be D, i.e., we might look at this instance of eq. (2.7):

$$D \to (D \to D)$$

In order to apply MP, we would also need to justify the corresponding second premise, namely A. But in our case, that would be D, and we won't be able to derive D by itself. So we need a different strategy.

The other axiom involving just $\to$ is eq. (2.8), i.e.,

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

We could get to the last nested conditional by applying MP twice. Again, that would mean that we want an instance of eq. (2.8) where $A \to C$ is $D \to D$, the formula we are aiming for. Then of course, A and C are both D. How should we pick B so that both $A \to (B \to C)$ and $A \to B$, i.e., in our case $D \to (B \to D)$ and $D \to B$, are also derivable? Well, the first of these is already an instance of eq. (2.7), whatever we decide B to be. And $D \to B$ would be another instance of eq. (2.7) if B were $(D \to D)$. So, our derivation is:

1.	$D \to ((D \to D) \to D)$	eq. (2.7)
2.	$(D \to ((D \to D) \to D)) \to$ $((D \to (D \to D)) \to (D \to D))$	eq. (2.8)
3.	$(D \to (D \to D)) \to (D \to D)$	1, 2, MP
4.	$D \to (D \to D)$	eq. (2.7)
5.	$D \to D$	3, 4, MP

Example 2.9. Sometimes we want to show that there is a derivation of some formula from some other formulas Γ. For instance, let's show that we can derive $A \to C$ from $\Gamma = \{A \to B, B \to C\}$.

1.	$A \to B$	Hyp
2.	$B \to C$	Hyp
3.	$(B \to C) \to (A \to (B \to C))$	eq. (2.7)
4.	$A \to (B \to C)$	2, 3, MP
5.	$(A \to (B \to C)) \to$ $((A \to B) \to (A \to C))$	eq. (2.8)
6.	$((A \to B) \to (A \to C))$	4, 5, MP
7.	$A \to C$	1, 6, MP

The lines labelled "Hyp" (for "hypothesis") indicate that the formula on that line is an element of Γ.

Proposition 2.10. *If $\Gamma \vdash A \to B$ and $\Gamma \vdash B \to C$, then $\Gamma \vdash A \to C$*

Proof. Suppose $\Gamma \vdash A \to B$ and $\Gamma \vdash B \to C$. Then there is a derivation of $A \to B$ from Γ; and a derivation of $B \to C$ from Γ as well. Combine these into a single derivation by concatenating them. Now add lines 3–7 of the derivation in the preceding example. This is a derivation of $A \to C$—which is the last line of the new derivation—from Γ. Note that the justifications of lines 4 and 7 remain valid if the reference to line number 2 is replaced by reference to the last line of the derivation of $A \to B$, and reference to line number 1 by reference to the last line of the derivation of $B \to C$. □

2.6 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 2.11 (Derivability). A formula A is *derivable* from Γ, written $\Gamma \vdash A$, if there is a derivation from Γ ending in A.

Definition 2.12 (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 2.13 (Consistency). A set Γ of formulas is *consistent* if and only if $\Gamma \nvdash \bot$; it is *inconsistent* otherwise.

Proposition 2.14 (Reflexivity). *If $A \in \Gamma$, then $\Gamma \vdash A$.*

Proof. The formula A by itself is a derivation of A from Γ. □

Proposition 2.15 (Monotonicity). *If $\Gamma \subseteq \Delta$ and $\Gamma \vdash A$, then $\Delta \vdash A$.*

Proof. Any derivation of A from Γ is also a derivation of A from Δ. □

Proposition 2.16 (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 Γ, i.e., a derivation $A_1, \dots, A_k = A$ where every A_i is an axiom, an element of Γ, or correct by a rule of inference. Now consider the sequence

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

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 $A_k = A$. □

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

Proposition 2.17. *Γ is inconsistent iff $\Gamma \vdash A$ for every A.*

Proof. Exercise. □

Proposition 2.18 (Compactness).

1. *If $\Gamma \vdash A$ then there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \vdash A$.*
2. *If every finite subset of Γ is consistent, then Γ is consistent.*

Proof.

1. If $\Gamma \vdash A$, then there is a finite sequence of formulas $A_1, \ldots, A_n$ so that $A \equiv A_n$ and each A_i is either a logical axiom, an element of Γ or follows from previous formulas by modus ponens. Take Γ_0 to be those A_i which are in Γ. Then the derivation is likewise a derivation from Γ_0, and so $\Gamma_0 \vdash A$.
2. This is the contrapositive of (1) for the special case $A \equiv \bot$. □

2.7 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 2.14 says we can always assert that $\Gamma \vdash A$ when we know that $A \in \Gamma$. Proposition 2.15 says that if $\Gamma \vdash A$ then also $\Gamma \cup \{B\} \vdash A$. And Proposition 2.16 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 2.19. *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 2.16, $\Gamma \vdash B$. □

The most important result we'll use in this context is the deduction theorem:

Theorem 2.20 (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 2.15. Also, $\Gamma \cup \{A\} \vdash A$ by Proposition 2.14. So, by Proposition 2.19, $\Gamma \cup \{A\} \vdash B$.

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. (2.7), and Proposition 2.19 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 2.8.

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:

$$\Gamma \vdash A \to (C \to B);$$
$$\Gamma \vdash A \to C.$$

But also

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

by eq. (2.8), and two applications of Proposition 2.19 give $\Gamma \vdash A \to B$, as required. □

Notice how eq. (2.7) and eq. (2.8) were chosen precisely so that the Deduction Theorem would hold.

The following are some useful facts about derivability, which we leave as exercises.

Proposition 2.21. *1.* $\vdash (A \to B) \to ((B \to C) \to (A \to C)$*;*

2. If $\Gamma \cup \{\neg A\} \vdash \neg B$ *then* $\Gamma \cup \{B\} \vdash A$ *(Contraposition);*

3. $\{A, \neg A\} \vdash B$ *(Ex Falso Quodlibet, Explosion);*

4. $\{\neg\neg A\} \vdash A$ *(Double Negation Elimination);*

5. If $\Gamma \vdash \neg\neg A$ *then* $\Gamma \vdash A$*;*

Problems

Problem 2.1. Show that the following hold by exhibiting derivations from the axioms:

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

Problem 2.2. Prove Proposition 2.17.

Problem 2.3. Prove Proposition 2.21

CHAPTER 3

Sequent calculus

3.1 The Sequent Calculus

While many derivation systems operate with arrangements of sentences, the sequent calculus operates with *sequents.* A sequent is an expression of the form

$$A_1, \ldots, A_m \Rightarrow B_1, \ldots, B_m,$$

that is a pair of sequences of sentences, separated by the sequent symbol $\Rightarrow$. Either sequence may be empty. A derivation in the sequent calculus is a tree of sequents, where the topmost sequents are of a special form (they are called "initial sequents" or "axioms") and every other sequent follows from the sequents immediately above it by one of the rules of inference. The rules of inference either manipulate the sentences in the sequents (adding, removing, or rearranging them on either the left or the right), or they introduce a complex formula in the conclusion of the rule. For instance, the $\land$L rule allows the inference from $A, \Gamma \Rightarrow \Delta$ to $A \land B, \Gamma \Rightarrow \Delta$, and the $\to$R allows the inference from $A, \Gamma \Rightarrow \Delta, B$ to $\Gamma \Rightarrow \Delta, A \to B$, for any Γ, Δ, A, and B. (In particular, Γ and Δ may be empty.)

The $\vdash$ relation based on the sequent calculus is defined as follows: $\Gamma \vdash A$ iff there is some sequence Γ_0 such that every A in Γ_0 is in Γ and there is a derivation with the sequent $\Gamma_0 \Rightarrow A$ at its root. A is a theorem in the sequent calculus if the sequent $\Rightarrow A$ has a derivation. For instance, here is a derivation that shows that $\vdash (A \wedge B) \to A$:

$$\dfrac{\dfrac{A \Rightarrow A}{A \wedge B \Rightarrow A}\;\wedge\mathrm{L}}{\Rightarrow (A \wedge B) \to A}\;\to\mathrm{R}$$

A set Γ is inconsistent in the sequent calculus if there is a derivation of $\Gamma_0 \Rightarrow$ (where every $A \in \Gamma_0$ is in Γ and the right side of the sequent is empty). Using the rule WR, any sentence can be derived from an inconsistent set.

The sequent calculus was invented in the 1930s by Gerhard Gentzen. Because of its systematic and symmetric design, it is a very useful formalism for developing a theory of derivations. It is relatively easy to find derivations in the sequent calculus, but these derivations are often hard to read and their connection to proofs are sometimes not easy to see. It has proved to be a very elegant approach to derivation systems, however, and many logics have sequent calculus systems.

3.2 Rules and Derivations

For the following, let $\Gamma, \Delta, \Pi, \Lambda$ represent finite sequences of sentences.

Definition 3.1 (Sequent). A *sequent* is an expression of the form

$$\Gamma \Rightarrow \Delta$$

where Γ and Δ are finite (possibly empty) sequences of sentences of the language $\mathcal{L}$. Γ is called the *antecedent*, while Δ is the *succedent*.

The intuitive idea behind a sequent is: if all of the sentences in the antecedent hold, then at least one of the sentences in the succedent holds. That is, if $\Gamma = \langle A_1, \ldots, A_m \rangle$ and $\Delta = \langle B_1, \ldots, B_n \rangle$, then $\Gamma \Rightarrow \Delta$ holds iff

$$(A_1 \wedge \cdots \wedge A_m) \rightarrow (B_1 \vee \cdots \vee B_n)$$

holds. There are two special cases: where Γ is empty and when Δ is empty. When Γ is empty, i.e., $m = 0$, $\Rightarrow \Delta$ holds iff $B_1 \vee \cdots \vee B_n$ holds. When Δ is empty, i.e., $n = 0$, $\Gamma \Rightarrow$ holds iff $\neg(A_1 \wedge \cdots \wedge A_m)$ does. We say a sequent is valid iff the corresponding sentence is valid.

If Γ is a sequence of sentences, we write Γ, A for the result of appending A to the right end of Γ (and A, Γ for the result of appending A to the left end of Γ). If Δ is a sequence of sentences also, then Γ, Δ is the concatenation of the two sequences.

Definition 3.2 (Initial Sequent). An *initial sequent* is a sequent of one of the following forms:

1. $A \Rightarrow A$
2. $\bot \Rightarrow$

for any sentence A in the language.

Derivations in the sequent calculus are certain trees of sequents, where the topmost sequents are initial sequents, and if a sequent stands below one or two other sequents, it must follow correctly by a rule of inference. The rules for ***LK*** are divided into two main types: *logical* rules and *structural* rules. The logical rules are named for the main operator of the sentence containing A and/or B in the lower sequent. Each one comes in two versions, one for inferring a sequent with the sentence containing the logical operator on the left, and one with the sentence on the right.

3.3 Propositional Rules

Rules for ¬

$$\frac{\Gamma \Rightarrow \Delta, A}{\neg A, \Gamma \Rightarrow \Delta}\ \neg\mathrm{L} \qquad \frac{A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg A}\ \neg\mathrm{R}$$

Rules for ∧

$$\frac{A, \Gamma \Rightarrow \Delta}{A \wedge B, \Gamma \Rightarrow \Delta}\ \wedge\mathrm{L}$$

$$\frac{B, \Gamma \Rightarrow \Delta}{A \wedge B, \Gamma \Rightarrow \Delta}\ \wedge\mathrm{L}$$

$$\frac{\Gamma \Rightarrow \Delta, A \qquad \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \wedge B}\ \wedge\mathrm{R}$$

Rules for ∨

$$\frac{A, \Gamma \Rightarrow \Delta \qquad B, \Gamma \Rightarrow \Delta}{A \vee B, \Gamma \Rightarrow \Delta}\ \vee\mathrm{L}$$

$$\frac{\Gamma \Rightarrow \Delta, A}{\Gamma \Rightarrow \Delta, A \vee B}\ \vee\mathrm{R}$$

$$\frac{\Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \vee B}\ \vee\mathrm{R}$$

Rules for →

$$\frac{\Gamma \Rightarrow \Delta, A \qquad B, \Pi \Rightarrow \Lambda}{A \rightarrow B, \Gamma, \Pi \Rightarrow \Delta, \Lambda}\ \rightarrow\mathrm{L} \qquad \frac{A, \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \rightarrow B}\ \rightarrow\mathrm{R}$$

3.4 Structural Rules

We also need a few rules that allow us to rearrange sentences in the left and right side of a sequent. Since the logical rules require that the sentences in the premise which the rule acts upon stand either to the far left or to the far right, we need an “exchange” rule that allows us to move sentences to the right position. It’s also important sometimes to be able to combine two identical sentences into one, and to add a sentence on either side.

Weakening

$$\frac{\Gamma \Rightarrow \Delta}{A, \Gamma \Rightarrow \Delta}\ \text{WL} \qquad \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, A}\ \text{WR}$$

Contraction

$$\frac{A, A, \Gamma \Rightarrow \Delta}{A, \Gamma \Rightarrow \Delta}\ \text{CL} \qquad \frac{\Gamma \Rightarrow \Delta, A, A}{\Gamma \Rightarrow \Delta, A}\ \text{CR}$$

Exchange

$$\frac{\Gamma, A, B, \Pi \Rightarrow \Delta}{\Gamma, B, A, \Pi \Rightarrow \Delta}\ \text{XL} \qquad \frac{\Gamma \Rightarrow \Delta, A, B, \Lambda}{\Gamma \Rightarrow \Delta, B, A, \Lambda}\ \text{XR}$$

A series of weakening, contraction, and exchange inferences will often be indicated by double inference lines.

The following rule, called "cut," is not strictly speaking necessary, but makes it a lot easier to reuse and combine derivations.

$$\frac{\Gamma \Rightarrow \Delta, A \qquad A, \Pi \Rightarrow \Lambda}{\Gamma, \Pi \Rightarrow \Delta, \Lambda}\ \text{Cut}$$

3.5 Derivations

We've said what an initial sequent looks like, and we've given the rules of inference. Derivations in the sequent calculus are inductively generated from these: each derivation either is an initial sequent on its own, or consists of one or two derivations followed by an inference.

Definition 3.3 (*LK* derivation). An *LK-derivation* of a sequent S is a finite tree of sequents satisfying the following conditions:

1. The topmost sequents of the tree are initial sequents.

2. The bottommost sequent of the tree is S.

3. Every sequent in the tree except S is a premise of a correct application of an inference rule whose conclusion stands directly below that sequent in the tree.

We then say that S is the *end-sequent* of the derivation and that S is *derivable in* ***LK*** (or ***LK***-derivable).

Example 3.4. Every initial sequent, e.g., $C \Rightarrow C$ is a derivation. We can obtain a new derivation from this by applying, say, the WL rule,

$$\frac{\Gamma \Rightarrow \Delta}{A, \Gamma \Rightarrow \Delta}\ \text{WL}$$

The rule, however, is meant to be general: we can replace the A in the rule with any sentence, e.g., also with D. If the premise matches our initial sequent $C \Rightarrow C$, that means that both Γ and Δ are just C, and the conclusion would then be $D, C \Rightarrow C$. So, the following is a derivation:

$$\frac{C \Rightarrow C}{D, C \Rightarrow C}\ \text{WL}$$

We can now apply another rule, say XL, which allows us to switch two sentences on the left. So, the following is also a correct derivation:

$$\frac{\dfrac{C \Rightarrow C}{D, C \Rightarrow C}\ \text{WL}}{C, D \Rightarrow C}\ \text{XL}$$

In this application of the rule, which was given as

$$\frac{\Gamma, A, B, \Pi \Rightarrow \Delta}{\Gamma, B, A, \Pi \Rightarrow \Delta,}\ \text{XL}$$

both Γ and Π were empty, Δ is C, and the roles of A and B are played by D and C, respectively. In much the same way, we also see that

$$\dfrac{D \Rightarrow D}{C, D \Rightarrow D}\ \text{WL}$$

is a derivation. Now we can take these two derivations, and combine them using $\wedge$R. That rule was

$$\dfrac{\Gamma \Rightarrow \Delta, A \qquad \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \wedge B}\ \wedge\text{R}$$

In our case, the premises must match the last sequents of the derivations ending in the premises. That means that Γ is C, D, Δ is empty, A is C and B is D. So the conclusion, if the inference should be correct, is $C, D \Rightarrow C \wedge D$.

$$\dfrac{\dfrac{\dfrac{C \Rightarrow C}{D, C \Rightarrow C}\ \text{WL}}{C, D \Rightarrow C}\ \text{XL} \qquad \dfrac{D \Rightarrow D}{C, D \Rightarrow D}\ \text{WL}}{C, D \Rightarrow C \wedge D}\ \wedge\text{R}$$

Of course, we can also reverse the premises, then A would be D and B would be C.

$$\dfrac{\dfrac{D \Rightarrow D}{C, D \Rightarrow D}\ \text{WL} \qquad \dfrac{\dfrac{C \Rightarrow C}{D, C \Rightarrow C}\ \text{WL}}{C, D \Rightarrow C}\ \text{XL}}{C, D \Rightarrow D \wedge C}\ \wedge\text{R}$$

3.6 Examples of Derivations

Example 3.5. Give an *LK*-derivation for the sequent $A \wedge B \Rightarrow A$.

We begin by writing the desired end-sequent at the bottom of the derivation.

$$\dfrac{}{A \wedge B \Rightarrow A}$$

Next, we need to figure out what kind of inference could have a lower sequent of this form. This could be a structural rule, but it is a good idea to start by looking for a logical rule. The only logical connective occurring in the lower sequent is $\wedge$, so we're looking for an $\wedge$ rule, and since the $\wedge$ symbol occurs in the antecedent, we're looking at the $\wedge$L rule.

$$\dfrac{}{A \wedge B \Rightarrow A}\ \wedge\text{L}$$

There are two options for what could have been the upper sequent of the $\wedge$L inference: we could have an upper sequent of $A \Rightarrow A$, or of $B \Rightarrow A$. Clearly, $A \Rightarrow A$ is an initial sequent (which is a good thing), while $B \Rightarrow A$ is not derivable in general. We fill in the upper sequent:

$$\dfrac{A \Rightarrow A}{A \wedge B \Rightarrow A}\ \wedge\text{L}$$

We now have a correct ***LK***-derivation of the sequent $A \wedge B \Rightarrow A$.

Example 3.6. Give an ***LK***-derivation for the sequent $\neg A \vee B \Rightarrow A \rightarrow B$.

Begin by writing the desired end-sequent at the bottom of the derivation.

$$\dfrac{}{\neg A \vee B \Rightarrow A \rightarrow B}$$

To find a logical rule that could give us this end-sequent, we look at the logical connectives in the end-sequent: $\neg$, $\vee$, and $\rightarrow$. We only care at the moment about $\vee$ and $\rightarrow$ because they are main operators of sentences in the end-sequent, while $\neg$ is inside the scope of another connective, so we will take care of it later. Our options for logical rules for the final inference are therefore the $\vee$L rule and the $\rightarrow$R rule. We could pick either rule, really, but let's pick the $\rightarrow$R rule (if for no reason other than it allows us to put off splitting into two branches). According to the form of $\rightarrow$R inferences which can yield the lower sequent, this must look like:

$$
\begin{prooftree}
\AxiomC{}
\UnaryInfC{$A, \neg A \lor B \ \Rightarrow \ B$}
\RightLabel{$\to$R}
\UnaryInfC{$\neg A \lor B \ \Rightarrow \ A \to B$}
\end{prooftree}
$$

If we move $\neg A \lor B$ to the outside of the antecedent, we can apply the $\lor$L rule. According to the schema, this must split into two upper sequents as follows:

$$
\begin{prooftree}
\AxiomC{}
\UnaryInfC{$\neg A, A \ \Rightarrow \ B$}
\AxiomC{}
\UnaryInfC{$B, A \ \Rightarrow \ B$}
\RightLabel{$\lor$L}
\BinaryInfC{$\neg A \lor B, A \ \Rightarrow \ B$}
\RightLabel{XR}
\UnaryInfC{$A, \neg A \lor B \ \Rightarrow \ B$}
\RightLabel{$\to$R}
\UnaryInfC{$\neg A \lor B \ \Rightarrow \ A \to B$}
\end{prooftree}
$$

Remember that we are trying to wind our way up to initial sequents; we seem to be pretty close! The right branch is just one weakening and one exchange away from an initial sequent and then it is done:

$$
\begin{prooftree}
\AxiomC{}
\UnaryInfC{$\neg A, A \ \Rightarrow \ B$}
\AxiomC{$B \ \Rightarrow \ B$}
\RightLabel{WL}
\UnaryInfC{$A, B \ \Rightarrow \ B$}
\RightLabel{XL}
\UnaryInfC{$B, A \ \Rightarrow \ B$}
\RightLabel{$\lor$L}
\BinaryInfC{$\neg A \lor B, A \ \Rightarrow \ B$}
\RightLabel{XR}
\UnaryInfC{$A, \neg A \lor B \ \Rightarrow \ B$}
\RightLabel{$\to$R}
\UnaryInfC{$\neg A \lor B \ \Rightarrow \ A \to B$}
\end{prooftree}
$$

Now looking at the left branch, the only logical connective in any sentence is the $\neg$ symbol in the antecedent sentences, so we're looking at an instance of the $\neg$L rule.

$$
\begin{prooftree}
\AxiomC{}
\UnaryInfC{$A \ \Rightarrow \ B, A$}
\RightLabel{$\neg$L}
\UnaryInfC{$\neg A, A \ \Rightarrow \ B$}
\AxiomC{$B \ \Rightarrow \ B$}
\RightLabel{WL}
\UnaryInfC{$A, B \ \Rightarrow \ B$}
\RightLabel{XL}
\UnaryInfC{$B, A \ \Rightarrow \ B$}
\RightLabel{$\lor$L}
\BinaryInfC{$\neg A \lor B, A \ \Rightarrow \ B$}
\RightLabel{XR}
\UnaryInfC{$A, \neg A \lor B \ \Rightarrow \ B$}
\RightLabel{$\to$R}
\UnaryInfC{$\neg A \lor B \ \Rightarrow \ A \to B$}
\end{prooftree}
$$

Similarly to how we finished off the right branch, we are just one weakening and one exchange away from finishing off this left branch as well.

$$
\begin{prooftree}
\AxiomC{$A \Rightarrow A$}
\RightLabel{WR}
\UnaryInfC{$A \Rightarrow A, B$}
\RightLabel{XR}
\UnaryInfC{$A \Rightarrow B, A$}
\RightLabel{$\neg$L}
\UnaryInfC{$\neg A, A \Rightarrow B$}
\AxiomC{$B \Rightarrow B$}
\RightLabel{WL}
\UnaryInfC{$A, B \Rightarrow B$}
\RightLabel{XL}
\UnaryInfC{$B, A \Rightarrow B$}
\RightLabel{$\vee$L}
\BinaryInfC{$\neg A \vee B, A \Rightarrow B$}
\RightLabel{XR}
\UnaryInfC{$A, \neg A \vee B \Rightarrow B$}
\RightLabel{$\rightarrow$R}
\UnaryInfC{$\neg A \vee B \Rightarrow A \rightarrow B$}
\end{prooftree}
$$

Example 3.7. Give an *LK*-derivation of the sequent $\neg A \vee \neg B \Rightarrow \neg(A \wedge B)$

Using the techniques from above, we start by writing the desired end-sequent at the bottom.

$$
\begin{prooftree}
\AxiomC{}
\UnaryInfC{$\neg A \vee \neg B \Rightarrow \neg(A \wedge B)$}
\end{prooftree}
$$

The available main connectives of sentences in the end-sequent are the $\vee$ symbol and the $\neg$ symbol. It would work to apply either the $\vee$L or the $\neg$R rule here, but we start with the $\neg$R rule because it avoids splitting up into two branches for a moment:

$$
\begin{prooftree}
\AxiomC{}
\UnaryInfC{$A \wedge B, \neg A \vee \neg B \Rightarrow$}
\RightLabel{$\neg$R}
\UnaryInfC{$\neg A \vee \neg B \Rightarrow \neg(A \wedge B)$}
\end{prooftree}
$$

Now we have a choice of whether to look at the $\wedge$L or the $\vee$L rule. Let's see what happens when we apply the $\wedge$L rule: we have a choice to start with either the sequent $A, \neg A \vee B \Rightarrow$ or the sequent $B, \neg A \vee B \Rightarrow$. Since the derivation is symmetric with regards to A and B, let's go with the former:

$$
\begin{prooftree}
\AxiomC{}
\UnaryInfC{$A, \neg A \vee \neg B \Rightarrow$}
\RightLabel{$\wedge$L}
\UnaryInfC{$A \wedge B, \neg A \vee \neg B \Rightarrow$}
\RightLabel{$\neg$R}
\UnaryInfC{$\neg A \vee \neg B \Rightarrow \neg(A \wedge B)$}
\end{prooftree}
$$

Continuing to fill in the derivation, we see that we run into a problem:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \Rightarrow A}{\neg A, A \Rightarrow}\neg\mathrm{L} \qquad \dfrac{\dfrac{}{A \Rightarrow B}\,?}{\neg B, A \Rightarrow}\neg\mathrm{L}}{\neg A \lor \neg B, A \Rightarrow}\lor\mathrm{L}}{A, \neg A \lor \neg B \Rightarrow}\mathrm{XL}}{A \land B, \neg A \lor \neg B \Rightarrow}\land\mathrm{L}}{\neg A \lor \neg B \Rightarrow \neg(A \land B)}\neg\mathrm{R}$$

The top of the right branch cannot be reduced any further, and it cannot be brought by way of structural inferences to an initial sequent, so this is not the right path to take. So clearly, it was a mistake to apply the $\land$L rule above. Going back to what we had before and carrying out the $\lor$L rule instead, we get

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\neg A, A \land B \Rightarrow} \qquad \dfrac{}{\neg B, A \land B \Rightarrow}}{\neg A \lor \neg B, A \land B \Rightarrow}\lor\mathrm{L}}{A \land B, \neg A \lor \neg B \Rightarrow}\mathrm{XL}}{\neg A \lor \neg B \Rightarrow \neg(A \land B)}\neg\mathrm{R}$$

Completing each branch as we've done before, we get

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \Rightarrow A}{A \land B \Rightarrow A}\land\mathrm{L}}{\neg A, A \land B \Rightarrow}\neg\mathrm{L} \qquad \dfrac{\dfrac{B \Rightarrow B}{A \land B \Rightarrow B}\land\mathrm{L}}{\neg B, A \land B \Rightarrow}\neg\mathrm{L}}{\neg A \lor \neg B, A \land B \Rightarrow}\lor\mathrm{L}}{A \land B, \neg A \lor \neg B \Rightarrow}\mathrm{XL}}{\neg A \lor \neg B \Rightarrow \neg(A \land B)}\neg\mathrm{R}$$

(We could have carried out the $\land$ rules lower than the $\neg$ rules in these steps and still obtained a correct derivation).

Example 3.8. So far we haven't used the contraction rule, but it is sometimes required. Here's an example where that happens. Suppose we want to prove $\Rightarrow A \lor \neg A$. Applying $\lor$R backwards would give us one of these two derivations:

$$\dfrac{\dfrac{}{\Rightarrow A}}{\Rightarrow A \lor \neg A}\lor\mathrm{R} \qquad\qquad \dfrac{\dfrac{\dfrac{}{A \Rightarrow}}{\Rightarrow \neg A}\neg\mathrm{R}}{\Rightarrow A \lor \neg A}\lor\mathrm{R}$$

Neither of these of course ends in an initial sequent. The trick is to realize that the contraction rule allows us to combine two copies of a sentence into one—and when we're searching for a proof, i.e., going from bottom to top, we can keep a copy of $A \lor \neg A$ in the premise, e.g.,

$$\cfrac{\cfrac{\cfrac{}{\Rightarrow A \lor \neg A, A}}{\Rightarrow A \lor \neg A, A \lor \neg A}\ \lor\mathrm{R}}{\Rightarrow A \lor \neg A}\ \mathrm{CR}$$

Now we can apply $\lor$R a second time, and also get $\neg A$, which leads to a complete derivation.

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{A \Rightarrow A}{\Rightarrow A, \neg A}\ \neg\mathrm{R}}{\Rightarrow A, A \lor \neg A}\ \lor\mathrm{R}}{\Rightarrow A \lor \neg A, A}\ \mathrm{XR}}{\Rightarrow A \lor \neg A, A \lor \neg A}\ \lor\mathrm{R}}{\Rightarrow A \lor \neg A}\ \mathrm{CR}$$

3.7 Proof-Theoretic Notions

Just as we've defined a number of important semantic notions (validity, entailment, satisfiabilty), we now define corresponding *proof-theoretic notions.* These are not defined by appeal to satisfaction of sentences in structures, but by appeal to the derivability or non-derivability of certain sequents. It was an important discovery that these notions coincide. That they do is the content of the *soundness* and *completeness theorem.*

Definition 3.9 (Theorems). A sentence A is a *theorem* if there is a derivation in *LK* of the sequent $\Rightarrow A$. We write $\vdash A$ if A is a theorem and $\nvdash A$ if it is not.

Definition 3.10 (Derivability). A sentence A is *derivable from* a set of sentences Γ, $\Gamma \vdash A$, iff there is a finite subset $\Gamma_0 \subseteq \Gamma$ and a sequence Γ_0' of the sentences in Γ_0 such that **LK** derives $\Gamma_0' \Rightarrow A$. If A is not derivable from Γ we write $\Gamma \nvdash A$.

Because of the contraction, weakening, and exchange rules, the order and number of sentences in Γ_0' does not matter: if a sequent $\Gamma_0' \Rightarrow A$ is derivable, then so is $\Gamma_0'' \Rightarrow A$ for any Γ_0'' that contains the same sentences as Γ_0'. For instance, if $\Gamma_0 = \{B, C\}$ then both $\Gamma_0' = \langle B, B, C\rangle$ and $\Gamma_0'' = \langle C, C, B\rangle$ are sequences containing just the sentences in Γ_0. If a sequent containing one is derivable, so is the other, e.g.:

$$\dfrac{\dfrac{\dfrac{\vdots \atop B, B, C \Rightarrow A}{B, C \Rightarrow A}\,\text{CL}}{C, B \Rightarrow A}\,\text{XL}}{C, C, B \Rightarrow A}\,\text{WL}$$

From now on we'll say that if Γ_0 is a finite set of sentences then $\Gamma_0 \Rightarrow A$ is any sequent where the antecedent is a sequence of sentences in Γ_0 and tacitly include contractions, exchanges, and weakenings if necessary.

Definition 3.11 (Consistency). A set of sentences Γ is *inconsistent* iff there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that **LK** derives $\Gamma_0 \Rightarrow$. If Γ is not inconsistent, i.e., if for every finite $\Gamma_0 \subseteq \Gamma$, **LK** does not derive $\Gamma_0 \Rightarrow$, we say it is *consistent*.

Proposition 3.12 (Reflexivity). *If $A \in \Gamma$, then $\Gamma \vdash A$.*

Proof. The initial sequent $A \Rightarrow A$ is derivable, and $\{A\} \subseteq \Gamma$. □

Proposition 3.13 (Monotonicity). *If $\Gamma \subseteq \Delta$ and $\Gamma \vdash A$, then $\Delta \vdash A$.*

Proof. Suppose $\Gamma \vdash A$, i.e., there is a finite $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \Rightarrow A$ is derivable. Since $\Gamma \subseteq \Delta$, then Γ_0 is also a finite subset of Δ. The derivation of $\Gamma_0 \Rightarrow A$ thus also shows $\Delta \vdash A$. □

Proposition 3.14 (Transitivity). *If $\Gamma \vdash A$ and $\{A\} \cup \Delta \vdash B$, then $\Gamma \cup \Delta \vdash B$.*

Proof. If $\Gamma \vdash A$, there is a finite $\Gamma_0 \subseteq \Gamma$ and a derivation π_0 of $\Gamma_0 \Rightarrow A$. If $\{A\} \cup \Delta \vdash B$, then for some finite subset $\Delta_0 \subseteq \Delta$, there is a derivation π_1 of $A, \Delta_0 \Rightarrow B$. Consider the following derivation:

$$\dfrac{\begin{array}{c}\vdots\,\pi_0\\ \Gamma_0 \Rightarrow A\end{array} \qquad \begin{array}{c}\vdots\,\pi_1\\ A, \Delta_0 \Rightarrow B\end{array}}{\Gamma_0, \Delta_0 \Rightarrow B}\ \text{Cut}$$

Since $\Gamma_0 \cup \Delta_0 \subseteq \Gamma \cup \Delta$, this shows $\Gamma \cup \Delta \vdash B$. □

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

Proposition 3.15. *Γ is inconsistent iff $\Gamma \vdash A$ for every sentence A.*

Proof. Exercise. □

Proposition 3.16 (Compactness).

1. *If $\Gamma \vdash A$ then there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \vdash A$.*
2. *If every finite subset of Γ is consistent, then Γ is consistent.*

Proof.

1. If $\Gamma \vdash A$, then there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that the sequent $\Gamma_0 \Rightarrow A$ has a derivation. Consequently, $\Gamma_0 \vdash A$.

2. If Γ is inconsistent, there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that **LK** derives $\Gamma_0 \Rightarrow$. But then Γ_0 is a finite subset of Γ that is inconsistent. □

Problems

Problem 3.1. Give derivations of the following sequents:

1. $A \land (B \land C) \Rightarrow (A \land B) \land C$.
2. $A \lor (B \lor C) \Rightarrow (A \lor B) \lor C$.
3. $A \to (B \to C) \Rightarrow B \to (A \to C)$.
4. $A \Rightarrow \neg\neg A$.

Problem 3.2. Give derivations of the following sequents:

1. $(A \lor B) \to C \Rightarrow A \to C$.
2. $(A \to C) \land (B \to C) \Rightarrow (A \lor B) \to C$.
3. $\Rightarrow \neg(A \land \neg A)$.
4. $B \to A \Rightarrow \neg A \to \neg B$.
5. $\Rightarrow (A \to \neg A) \to \neg A$.
6. $\Rightarrow \neg(A \to B) \to \neg B$.
7. $A \to C \Rightarrow \neg(A \land \neg C)$.
8. $A \land \neg C \Rightarrow \neg(A \to C)$.
9. $A \lor B, \neg B \Rightarrow A$.
10. $\neg A \lor \neg B \Rightarrow \neg(A \land B)$.
11. $\Rightarrow (\neg A \land \neg B) \to \neg(A \lor B)$.

12. $\Rightarrow \neg(A \lor B) \to (\neg A \land \neg B)$.

Problem 3.3. Give derivations of the following sequents:

1. $\neg(A \to B) \Rightarrow A$.
2. $\neg(A \land B) \Rightarrow \neg A \lor \neg B$.
3. $A \to B \Rightarrow \neg A \lor B$.
4. $\Rightarrow \neg\neg A \to A$.
5. $A \to B, \neg A \to B \Rightarrow B$.
6. $(A \land B) \to C \Rightarrow (A \to C) \lor (B \to C)$.
7. $(A \to B) \to A \Rightarrow A$.
8. $\Rightarrow (A \to B) \lor (B \to C)$.

(These all require the CR rule.)

Problem 3.4. Prove ??

PART II

Does everything have to be true or false?

CHAPTER 4

Syntax and Semantics

4.1 Introduction

In classical logic, we deal with formulas that are built from propositional variables using the propositional connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$. When we define a semantics for classical logic, we do so using the two truth values $\mathbb{T}$ and $\mathbb{F}$. We interpret propositional variables in a valuation $\mathfrak{v}$, which assigns these truth values $\mathbb{T}$, $\mathbb{F}$ to the propositional variables. Any valuation then determines a truth value $\overline{\mathfrak{v}}(A)$ for any formula A, and A formula is satisfied in a valuation $\mathfrak{v}$, $\mathfrak{v} \vDash A$, iff $\overline{\mathfrak{v}}(A) = \mathbb{T}$.

Many-valued logics are generalizations of classical two-valued logic by allowing more truth values than just $\mathbb{T}$ and $\mathbb{F}$. So in many-valued logic, a valuation $\mathfrak{v}$ is a function assigning to every propositional variable p one of a range of possible truth values. We'll generally call the set of allowed truth values V. Classical logic is a many-valued logic where $V = \{\mathbb{T},\mathbb{F}\}$, and the truth value $\overline{\mathfrak{v}}(A)$ is computed using the familiar characteristic truth tables for the connectives.

Once we add additional truth values, we have more than one natural option for how to compute $\overline{\mathfrak{v}}(A)$ for the connectives we read as "and," "or," "not," and "if—then." So a many-valued logic is determined not just by the set of truth values, but also by the *truth functions* we decide to use for each connective. Once these are selected for a many-valued logic $\boldsymbol{L}$, however, the truth value $\overline{v}_L(A)$ is uniquely determined by the valuation, just like in classical logic. Many-valued logics, like classical logic, are *truth functional.*

With this semantic building blocks in hand, we can go on to define the analogs of the semantic concepts of tautology, entailment, and satisfiability. In classical logic, a formula is a tautology if its truth value $\overline{v}(A) = \mathbb{T}$ for any v. In many-valued logic, we have to generalize this a bit as well. First of all, there is no requirement that the set of truth values V contains $\mathbb{T}$. For instance, some many-valued logics use numbers, such as all rational numbers between 0 and 1 as their set of truth values. In such a case, 1 usually plays the rule of $\mathbb{T}$. In other logics, not just one but several truth values do. So, we require that every many-valued logic have a set V^+ of *designated values.* We can then say that a formula is satisfied in a valuation v, $v \vDash_L A$, iff $\overline{v}_L(A) \in V^+$. A formula A is a tautology of the logic, $\vDash_L A$, iff $\overline{v}(A) \in V^+$ for any v. And, finally, we say that A is entailed by a set of formulas, $\Gamma \vDash_L A$, if every valuation that satisfies all the formulas in Γ also satisfies A.

4.2 Languages and Connectives

Classical propositional logic, and many other logics, use a set supply of *propositional constants* and *connectives.* For instance, we use the following as primitives:

1. The propositional constant for falsity $\bot$.

2. The logical connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional)

In addition to the primitive connectives above, we also use symbols defined as abbreviations, such as $\leftrightarrow$ (biconditional), $\top$ (truth)

The same connectives are used in many-valued logics as well. However, it is often useful to include different versions of, say, conjunction, in the same logic, and that would require different symbols to keep the versions separate. Some many-valued logics also include connectives that have no equivalent in classical logic. So, we'll be a bit more general than usual.

Definition 4.1. A *propositional language* consists of a set $\mathcal{L}$ of *connectives*. Each connective $\star$ has an *arity*; a connective of arity n is said to be *n-place*. Connectives of arity 0 are also called *constants*; connectives of arity 1 are called *unary*, and connectives of arity 2, *binary*.

Example 4.2. The standard language of propositional logic $\mathcal{L}_0$ consists of the following connectives (with associated arities): $\bot$ (0) $\lnot$ (1), $\land$ (2), $\lor$ (2), $\rightarrow$ (2). Most logics we consider will use this language. Some logics by tradition an convention use different symbols for some connectives. For instance, in product logic, the conjunction symbol is often $\odot$ instead of $\land$. Sometimes it is convenient to add a new operator, e.g., the determinateness operator $\vartriangle$ (1-place).

4.3 Formulas

Definition 4.3 (Formula). The set $\mathrm{Frm}(\mathcal{L})$ of *formulas* of a propositional language $\mathcal{L}$ is defined inductively as follows:

1. Every propositional variable p_i is an atomic formula.

2. Every 0-place connective (propositional constant) of $\mathcal{L}$ is an atomic formula.

3. If $\star$ is an n-place connective of $\mathcal{L}$, and A_1, …, A_n are formulas, then $\star(A_1, \dots, A_n)$ is a formula.

4. Nothing else is a formula.

If $\star$ is 1-place, then $\star(A_1)$ will often be written simply as $\star A_1$. If $\star$ is 2-place $\star(A_1, A_2)$ will often be written as $(A_1 \star A_2)$.

As usual, we will often silently leave out the outermost parentheses.

Example 4.4. In the standard language $\mathcal{L}_0$, $p_1 \to (p_1 \land \neg p_2)$ is a formula. In the language of product logic, it would be written instead as $p_1 \to (p_1 \odot \neg p_2)$. If we add the 1-place $\triangle$ to the language, we would also have formulas such as $\triangle(p_1 \land p_2) \to (\triangle p_1 \land \triangle p_2)$.

4.4 Matrices

A many-valued logic is defined by its language, its set of truth values V, a subset of designated truth values, and truth functions for its connective. Together, these elements are called a *matrix*.

Definition 4.5 (Matrix). A *matrix* for the logic $\boldsymbol{L}$ consists of:

1. a set of connectives making up a language $\mathcal{L}$;
2. a set $V \neq \emptyset$ of truth values;
3. a set $V^+ \subseteq V$ of designated truth values;
4. for each n-place connective $\star$ in $\mathcal{L}$, a truth function $\widetilde{\star} : V^n \to V$. If $n = 0$, then $\widetilde{\star}$ is just an element of V.

Example 4.6. The matrix for classical logic $\boldsymbol{C}$ consists of:

1. The standard propositional language $\mathcal{L}_0$ with $\bot$, $\neg$, $\land$, $\lor$, $\to$.
2. The set of truth values $V = \{\mathbb{T}, \mathbb{F}\}$.
3. $\mathbb{T}$ is the only designated value, i.e., $V^+ = \{\mathbb{T}\}$.

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

$\widetilde{\wedge}$	$\mathbb{T}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$

$\widetilde{\vee}$	$\mathbb{T}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$

$\widetilde{\rightarrow}$	$\mathbb{T}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$

Figure 4.1: Truth functions for classical logic ***C***.

4. For $\bot$, we have $\widetilde{\bot} = \mathbb{F}$. The other truth functions are given by the usual truth tables (see Figure 4.1).

4.5 Valuations and Satisfaction

Definition 4.7 (Valuations). Let V be a set of truth values. A *valuation* for $\mathcal{L}$ into V is a function v assigning an element of V to the propositional variables of the language, i.e., $v\colon \mathrm{At}_0 \to V$.

Definition 4.8. Given a valuation v into the set of truth values V of a many-valued logic $\boldsymbol{L}$, define the evaluation function $\overline{v}\colon \mathrm{Frm}(\mathcal{L}) \to V$ inductively by:

1. $\overline{v}(p_n) = v(p_n)$;

2. If $\star$ is a 0-place connective, then $\overline{v}(\star) = \widetilde{\star}_{\boldsymbol{L}}$;

3. If $\star$ is an n-place connective, then

$$\overline{v}(\star(A_1, \dots, A_n)) = \widetilde{\star}_{\boldsymbol{L}}(\overline{v}(A_1), \dots, \overline{v}(A_n)).$$

Definition 4.9 (Satisfaction). The formula A is *satisfied* by a valuation v, $v \vDash_{\boldsymbol{L}} A$, iff $\overline{v}_{\boldsymbol{L}}(A) \in V^+$, where V^+ is the set of designated truth values of $\boldsymbol{L}$.

We write $v \nvDash_L A$ to mean "not $v \vDash_L A$." If Γ is a set of formulas, $v \vDash_L \Gamma$ iff $v \vDash_L A$ for every $A \in \Gamma$.

4.6 Semantic Notions

Suppose a many-valued logic $\boldsymbol{L}$ is given by a matrix. Then we can define the usual semantic notions for $\boldsymbol{L}$.

Definition 4.10.

1. A formula A is *satisfiable* if for some v, $v \vDash A$; it is *unsatisfiable* if for no v, $v \vDash A$;
2. A formula A is a *tautology* if $v \vDash A$ for all valuations v;
3. If Γ is a set of formulas, $\Gamma \vDash A$ ("Γ entails A") if and only if $v \vDash A$ for every valuation v for which $v \vDash \Gamma$.
4. If Γ is a set of formulas, Γ is *satisfiable* if there is a valuation v for which $v \vDash \Gamma$, and Γ is *unsatisfiable* otherwise.

We have some of the same facts for these notions as we do for the case of classical logic:

Proposition 4.11.

1. *A is a tautology if and only if $\emptyset \vDash A$;*
2. *If Γ is satisfiable then every finite subset of Γ is also satisfiable;*
3. *Monotonicity: if $\Gamma \subseteq \Delta$ and $\Gamma \vDash A$ then also $\Delta \vDash A$;*
4. *Transitivity: if $\Gamma \vDash A$ and $\Delta \cup \{A\} \vDash B$ then $\Gamma \cup \Delta \vDash B$;*

Proof. Exercise. □

In classical logic we can connect entailment and the conditional. For instance, we have the validity of *modus ponens*: If $\Gamma \vDash A$ and $\Gamma \vDash A \rightarrow B$ then $\Gamma \vDash B$. Another important relationship between $\vDash$ and $\rightarrow$ in classical logic is the semantic deduction theorem: $\Gamma \vDash A \rightarrow B$ if and only if $\Gamma \cup \{A\} \vDash B$. These results *do not* always hold in many-valued logics. Whether they do depends on the truth function $\widetilde{\rightarrow}$.

4.7 Many-valued logics as sublogics of C

The usual many-valued logics are all defined using matrices in which the value of a truth-function for arguments in $\{\mathbb{T},\mathbb{F}\}$ agrees with the classical truth functions. Specifically, in these logics, if $x \in \{\mathbb{T},\mathbb{F}\}$, then $\widetilde{\neg}_L(x) = \widetilde{\neg}_C(x)$, and for $\star$ any one of $\wedge$, $\vee$, $\to$, if $x, y \in \{\mathbb{T},\mathbb{F}\}$, then $\widetilde{\star}_L(x,y) = \widetilde{\star}_C(x,y)$. In other words, the truth functions for $\neg$, $\wedge$, $\vee$, $\to$ restricted to $\{\mathbb{T},\mathbb{F}\}$ are exactly the classical truth functions.

Proposition 4.12. *Suppose that a many-valued logic* $\mathbf{L}$ *contains the connectives* $\neg$, $\wedge$, $\vee$, $\to$ *in its language,* $\mathbb{T},\mathbb{F} \in V$, *and its truth functions satisfy:*

1. $\widetilde{\neg}_L(x) = \widetilde{\neg}_C(x)$ *if* $x = \mathbb{T}$ *or* $x = \mathbb{F}$*;*
2. $\widetilde{\wedge}_L(x,y) = \widetilde{\wedge}_C(x,y)$,
3. $\widetilde{\vee}_L(x,y) = \widetilde{\vee}_C(x,y)$,
4. $\widetilde{\to}_L(x,y) = \widetilde{\to}_C(x,y)$, *if* $x, y \in \{\mathbb{T},\mathbb{F}\}$.

Then, for any valuation v *into* V *such that* $v(p) \in \{\mathbb{T},\mathbb{F}\}$, $\overline{v}_L(A) = \overline{v}_C(A)$.

Proof. By induction on A.

1. If $A \equiv p$ is atomic, we have $\overline{v}_L(A) = v(p) = \overline{v}_C(A)$.
2. If $A \equiv \neg B$, we have

$$\begin{aligned}
\overline{v}_L(A) &= \widetilde{\neg}_L(\overline{v}_L(B)) && \text{by Definition 4.8} \\
&= \widetilde{\neg}_L(\overline{v}_C(B)) && \text{by inductive hypothesis} \\
&= \widetilde{\neg}_C(\overline{v}_C(B)) && \text{by assumption (1),} \\
& && \text{since } \overline{v}_C(B) \in \{\mathbb{T},\mathbb{F}\}, \\
&= \overline{v}_C(A) && \text{by Definition 4.8.}
\end{aligned}$$

3. If $A \equiv (B \wedge C)$, we have

$$\overline{v}_L(A) = \widetilde{\wedge}_L(\overline{v}_L(B), \overline{v}_L(C)) \quad \text{by Definition 4.8}$$

$$\begin{aligned}
&= \widetilde{\land}_L(\overline{v}_C(B), \overline{v}_C(C)) && \text{by inductive hypothesis} \\
&= \widetilde{\land}_C(\overline{v}_C(B), \overline{v}_C(C)) && \text{by assumption (2),} \\
& && \text{since } \overline{v}_C(B), \overline{v}_C(C) \in \{\mathbb{T}, \mathbb{F}\}, \\
&= \overline{v}_C(A) && \text{by Definition 4.8.}
\end{aligned}$$

The cases where $A \equiv (B \lor C)$ and $A \equiv (B \to C)$ are similar. □

Corollary 4.13. *If a many-valued logic satisfies the conditions of Proposition 4.12, $\mathbb{T} \in V^+$ and $\mathbb{F} \notin V^+$, then $\vDash_L \subseteq \vDash_C$, i.e., if $\Gamma \vDash_L B$ then $\Gamma \vDash_C B$. In particular, every tautology of* **L** *is also a classical tautology.*

Proof. We prove the contrapositive. Suppose $\Gamma \nvDash_C B$. Then there is some valuation $v\colon \mathrm{At}_0 \to \{\mathbb{T}, \mathbb{F}\}$ such that $\overline{v}_C(A) = \mathbb{T}$ for all $A \in \Gamma$ and $\overline{v}_C(B) = \mathbb{F}$. Since $\mathbb{T}, \mathbb{F} \in V$, the valuation v is also a valuation for $\boldsymbol{L}$. By Proposition 4.12, $\overline{v}_L(A) = \mathbb{T}$ for all $A \in \Gamma$ and $\overline{v}_L(B) = \mathbb{F}$. Since $\mathbb{T} \in V^+$ and $\mathbb{F} \notin V^+$ that means $v \vDash_L \Gamma$ and $v \nvDash_L B$, i.e., $\Gamma \nvDash_L B$. □

Problems

Problem 4.1. Prove Proposition 4.11

CHAPTER 5

Three-valued Logics

5.1 Introduction

If we just add one more value $\mathbb{U}$ to $\mathbb{T}$ and $\mathbb{F}$, we get a three-valued logic. Even though there is only one more truth value, the possibilities for defining the truth-functions for $\neg$, $\wedge$, $\vee$, and $\rightarrow$ are quite numerous. Then a logic might use any combination of these truth functions, and you also have a choice of making only $\mathbb{T}$ designated, or both $\mathbb{T}$ and $\mathbb{U}$.

We present here a selection of the most well-known three-valued logics, their motivations, and some of their properties.

5.2 Łukasiewicz logic

One of the first published, worked out proposals for a many-valued logic is due to the Polish philosopher Jan Łukasiewicz in 1921. Łukasiewicz was motivated by Aristotle's sea battle problem: It seems that, *today*, the sentence "There will be a sea battle tomorrow" is neither true nor false: its truth value is not yet settled. Łukasiewicz proposed to introduce a third truth value, to such "future contingent" sentences.

> I can assume without contradiction that my presence in Warsaw at a certain moment of next year, e.g., at noon on 21 December, is at the present time determined neither positively nor negatively. Hence it is possible, but not necessary, that I shall be present in Warsaw at the given time. On this assumption the proposition "I shall be in Warsaw at noon on 21 December of next year," can at the present time be neither true nor false. For if it were true now, my future presence in Warsaw would have to be necessary, which is contradictory to the assumption. If it were false now, on the other hand, my future presence in Warsaw would have to be impossible, which is also contradictory to the assumption. Therefore the proposition considered is at the moment neither true nor false and must possess a third value, different from "0" or falsity and "1" or truth. This value we can designate by "$\frac{1}{2}$." It represents "the possible," and joins "the true" and "the false" as a third value.

We will use $\mathbb{U}$ for Łukasiewicz's third truth value.[1]

The truth functions for the connectives $\neg$, $\wedge$, and $\vee$ are easy to determine on this interpretation: the negation of a future contingent sentence is also a future contingent sentence, so $\widetilde{\neg}(\mathbb{U}) = \mathbb{U}$. If one conjunct of a conjunction is undetermined and the other is true, the conjunction is also undetermined—after all, depending on how the future contingent conjunct turns out, the conjunction might turn out to be true, and it might turn out to be false. So

$$\widetilde{\wedge}(\mathbb{T}, \mathbb{U}) = \widetilde{\wedge}(\mathbb{U}, \mathbb{T}) = \mathbb{U}.$$

If the other conjunct is false, however, it cannot turn out true, so

$$\widetilde{\wedge}(\mathbb{F}, \mathbb{U}) = \widetilde{\wedge}(\mathbb{F}, \mathbb{U}) = \mathbb{F}.$$

[1] Łukasiewicz here uses "possible" in a way that is uncommon today, namely to mean possible but not necessary.

The other values (if the arguments are settled truth values, $\mathbb{T}$ or $\mathbb{F}$, are like in classical logic.

For the conditional, the situation is a little trickier. Suppose q is a future contingent statement. If p is false, then $p \to q$ will be true, regardless of how q turns out, so we should set $\widetilde{\to}(\mathbb{F}, \mathbb{U}) = \mathbb{T}$. And if p is true, then $q \to p$ will be true, regardless of what q turns out to be, so $\widetilde{\to}(\mathbb{U}, \mathbb{T}) = \mathbb{T}$. If p is true, then $p \to q$ might turn out to be true or false, so $\widetilde{\to}(\mathbb{T}, \mathbb{U}) = \mathbb{U}$. Similarly, if p is false, then $q \to p$ might turn out to be true or false, so $\widetilde{\to}(\mathbb{U}, \mathbb{F}) = \mathbb{U}$. This leaves the case where p and q are both future contingents. On the basis of the motivation, we should really assign $\mathbb{U}$ in this case. However, this would make $A \to A$ *not* a tautology. Łukasiewicz had not trouble giving up $A \lor \lnot A$ and $\lnot(A \land \lnot A)$, but balked at giving up $A \to A$. So he stipulated $\widetilde{\to}(\mathbb{U}, \mathbb{U}) = \mathbb{T}$.

Definition 5.1. Three-valued Łukasiewicz logic is defined using the matrix:

1. The standard propositional language $\mathcal{L}_0$ with $\lnot$, $\land$, $\lor$, $\to$.

2. The set of truth values $V = \{\mathbb{T}, \mathbb{U}, \mathbb{F}\}$.

3. $\mathbb{T}$ is the only designated value, i.e., $V^+ = \{\mathbb{T}\}$.

4. Truth functions are given by the following tables:

$\widetilde{\lnot}$	
$\mathbb{T}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{T}$

$\widetilde{\land}_{Ł_3}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$

$\widetilde{\lor}_{Ł_3}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{U}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$

$\widetilde{\to}_{Ł_3}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$

As can easily be seen, any formula A containing only $\lnot$, $\land$,

and $\lor$ will take the truth value $\mathbb{U}$ if all its propositional variables are assigned $\mathbb{U}$. So for instance, the classical tautologies $p \lor \neg p$ and $\neg(p \land \neg p)$ are not tautologies in $Ł_3$, since $\overline{v}(A) = \mathbb{U}$ whenever $v(p) = \mathbb{U}$.

On valuations where $v(p) = \mathbb{T}$ or $\mathbb{F}$, $\overline{v}(A)$ will coincide with its classical truth value.

Proposition 5.2. *If $v(p) \in \{\mathbb{T}, \mathbb{F}\}$ for all p in A, then $\overline{v}_{Ł_3}(A) = \overline{v}_C(A)$.*

Many classical tautologies *are* also tautologies in $Ł_3$, e.g, $\neg p \to (p \to q)$. Just like in classical logic, we can use truth tables to verify this:

p	q	$\neg$	p	$\to$	$(p$	$\to$	$q)$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{U}$
$\mathbb{T}$	$\mathbb{F}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{T}$	$\mathbb{U}$
$\mathbb{U}$	$\mathbb{F}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{U}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$

One might therefore perhaps think that although not all classical tautologies are tautologies in $Ł_3$, they should at least take either the value $\mathbb{T}$ or the value $\mathbb{U}$ on every valuation. This is not the case. A counterexample is given by

$$\neg(p \to \neg p) \lor \neg(\neg p \to p)$$

which is $\mathbb{F}$ if p is $\mathbb{U}$.

Łukasiewicz hoped to build a logic of possibility on the basis of his three-valued system, by introducing a one-place connective $\Diamond A$ (for "A is possible") and a corresponding $\Box A$ (for "A is necessary"):

$\widetilde{\Diamond}$	
$\mathbb{T}$	$\mathbb{T}$
$\mathbb{U}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{F}$

$\widetilde{\Box}$	
$\mathbb{T}$	$\mathbb{T}$
$\mathbb{U}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{F}$

In other words, p is possible iff it is not already settled as false; and p is necessary iff it is already settled as true.

However, the shortcomings of this proposed modal logic soon became evident: However things turn out, $p \wedge \neg p$ can never turn out to be true. So even if it is not now settled (and therefore undetermined), it should count as impossible, i.e., $\neg\Diamond(p \wedge \neg p)$ should be a tautology. However, if $v(p) = \mathbb{U}$, then $\overline{v}(\neg\Diamond(p \wedge \neg p)) = \mathbb{U}$. Although Łukasiewicz was correct that two truth values will not be enough to accommodate modal distinctions such as possiblity and necessity, introducing a third truth value is also not enough.

5.3 Kleene logics

Stephen Kleene introduced two three-valued logics motivated by a logic in which truth values are thought of the outcomes of computational procedures: a procedure may yield $\mathbb{T}$ or $\mathbb{F}$, but it may also fail to terminate. In that case the corresponding truth value is undefined, represented by the truth value $\mathbb{U}$.

To compute the negation of a proposition A, you would first compute the value of A, and then return the opposite of the result. If the computation of A does not terminate, then the entire procedure does not either: so the negation of $\mathbb{U}$ is $\mathbb{U}$.

To compute a conjunction $A \wedge B$, there are two options: one can first compute A, then B, and then the result would be $\mathbb{T}$ if the outcome of both is $\mathbb{T}$, and $\mathbb{F}$ otherwise. If either computation fails to halt, the entire procedure does as well. So in this case, the if one conjunct is undefined, the conjunction is as well. The same goes for disjunction.

However, if we can evaluate A and B in parallel, we can do better. Then, if one of the two procedures halts and returns $\mathbb{F}$, we can stop, as the answer must be false. So in that case a conjunction with one false conjunct is false, even if the other conjunct is undefined. Similarly, when computing a disjunction in parallel, we can stop once the procedure for one of the two disjuncts has returned true: then the disjunction must be true. So in this case we can know what the outcome of a compound claim is, even if one of the components is undefined. On this interpretation, we might read $\mathbb{U}$ as "unknown" rather than "undefined."

The two interpretations give rise to Kleene's strong and weak logic. The conditional is defined as equivalent to $\neg A \lor B$.

Definition 5.3. *Strong Kleene logic* **Ks** is defined using the matrix:

1. The standard propositional language $\mathcal{L}_0$ with $\neg$, $\land$, $\lor$, $\rightarrow$.
2. The set of truth values $V = \{\mathbb{T}, \mathbb{U}, \mathbb{F}\}$.
3. $\mathbb{T}$ is the only designated value, i.e., $V^+ = \{\mathbb{T}\}$.
4. Truth functions are given by the following tables:

$\widetilde{\neg}$	
$\mathbb{T}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{T}$

$\widetilde{\land}_{Ks}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$

$\widetilde{\lor}_{Ks}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{U}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$

$\widetilde{\rightarrow}_{Ks}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$

Definition 5.4. *Weak Kleene logic* **Kw** is defined using the matrix:

1. The standard propositional language $\mathcal{L}_0$ with $\neg$, $\land$, $\lor$, $\rightarrow$.
2. The set of truth values $V = \{\mathbb{T}, \mathbb{U}, \mathbb{F}\}$.

3. $\mathbb{T}$ is the only designated value, i.e., $V^+ = \{\mathbb{T}\}$.

4. Truth functions are given by the following tables:

$\widetilde{\neg}$	
$\mathbb{T}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{T}$

$\widetilde{\wedge}_{Kw}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{U}$	$\mathbb{F}$

$\widetilde{\vee}_{Kw}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{T}$
$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$

$\widetilde{\rightarrow}_{Kw}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{T}$

Proposition 5.5. *$\boldsymbol{Ks}$ and $\boldsymbol{Kw}$ have no tautologies.*

Proof. If $v(p) = \mathbb{U}$ for all propositional variables p, then any formula A will have truth value $\overline{v}(A) = \mathbb{U}$, since

$$\widetilde{\neg}(\mathbb{U}) = \widetilde{\vee}(\mathbb{U},\mathbb{U}) = \widetilde{\wedge}(\mathbb{U},\mathbb{U}) = \widetilde{\rightarrow}(\mathbb{U},\mathbb{U}) = \mathbb{U}$$

in both logics. As $\mathbb{U} \notin V^+$ for either $\boldsymbol{Ks}$ or $\boldsymbol{Kw}$, on this valuation, A will not be designated. □

Although both weak and strong Kleene logic have no tautologies, they have non-trivial consequence relations.

Dmitry Bochvar interpreted $\mathbb{U}$ as “meaningless” and attempted to use it to solve paradoxes such as the Liar paradox by stipulating that paradoxical sentences take the value $\mathbb{U}$. He introduced a logic which is essentially weak Kleene logic extended by additional connectives, two of which are “external negation” and the “is undefined” operator:

$\widetilde{\sim}$	
$\mathbb{T}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{T}$

$\widetilde{+}$	
$\mathbb{T}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{F}$

5.4 Gödel logics

Kurt Gödel introduced a sequence of n-valued logics that each contain all formulas valid in intuitionistic logic, and are contained in classical logic. Here is the first interesting one:

Definition 5.6. 3*-valued Gödel logic* **G** is defined using the matrix:

1. The standard propositional language $\mathcal{L}_0$ with $\bot$, $\neg$, $\wedge$, $\vee$, $\rightarrow$.
2. The set of truth values $V = \{\mathbb{T}, \mathbb{U}, \mathbb{F}\}$.
3. $\mathbb{T}$ is the only designated value, i.e., $V^+ = \{\mathbb{T}\}$.
4. For $\bot$, we have $\widetilde{\bot} = \mathbb{F}$. Truth functions for the remaining connectives are given by the following tables:

$\widetilde{\neg}_G$	
$\mathbb{T}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$

$\widetilde{\wedge}_G$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{U}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$

$\widetilde{\vee}_G$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{U}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{U}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$

$\widetilde{\rightarrow}_G$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{U}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$

You'll notice that the truth tables for $\wedge$ and $\vee$ are the same as in Łukasiewicz and strong Kleene logic, but the truth tables for $\neg$ and $\rightarrow$ differ for each. In Gödel logic, $\widetilde{\neg}(\mathbb{U}) = \mathbb{F}$. In contrast to Łukasiewicz logic and Kleene logic, $\widetilde{\rightarrow}(\mathbb{U}, \mathbb{F}) = \mathbb{F}$; in contrast to Kleene logic (but as in Łukasiewicz logic), $\widetilde{\rightarrow}(\mathbb{U}, \mathbb{U}) = \mathbb{T}$.

As the connection to intuitionistic logic alluded to above suggests, G_3 is close to intuitionistic logic. All intuitionistic truths are tautologies in G_3, and many classical tautologies that are not valid intuitionistically also fail to be tautologies in G_3. For instance, the following are not tautologies:

$$
\begin{array}{ll}
p \lor \neg p & (p \to q) \to (\neg p \lor q) \\
\neg\neg p \to p & \neg(\neg p \land \neg q) \to (p \lor q) \\
((p \to q) \to p) \to p & \neg(p \to q) \to (p \land \neg q)
\end{array}
$$

However, not every tautology of $G3$ is also intuitionistically valid, e.g., $\neg\neg p \lor \neg p$ or $(p \to q) \lor (q \to p)$.

5.5 Designating not just $\mathbb{T}$

So far the logics we've seen all had the set of designated truth values $V^+ = \{\mathbb{T}\}$, i.e., something counts as true iff its truth value is $\mathbb{T}$. But one might also count something as true if it's just not $\mathbb{F}$. Then one would get a logic by stipulating in the matrix, e.g., that $V^+ = \{\mathbb{T}, \mathbb{U}\}$.

Definition 5.7. The *logic of paradox* LP is defined using the matrix:

1. The standard propositional language $\mathcal{L}_0$ with $\neg$, $\land$, $\lor$, $\to$.
2. The set of truth values $V = \{\mathbb{T}, \mathbb{U}, \mathbb{F}\}$.
3. $\mathbb{T}$ and $\mathbb{U}$ are designated, i.e., $V^+ = \{\mathbb{T}, \mathbb{U}\}$.
4. Truth functions are the same as in strong Kleene logic.

Definition 5.8. Halldén's *logic of nonsense* Hal is defined using the matrix:

1. The standard propositional language $\mathcal{L}_0$ with $\neg$, $\land$, $\lor$, $\to$ and a 1-place connective $+$.
2. The set of truth values $V = \{\mathbb{T}, \mathbb{U}, \mathbb{F}\}$.

3. $\mathbb{T}$ and $\mathbb{U}$ are designated, i.e., $V^+ = \{\mathbb{T}, \mathbb{U}\}$.

4. Truth functions are the same as weak Kleene logic, plus the "is meaningless" operator:

$\widetilde{+}$	
$\mathbb{T}$	$\mathbb{F}$
$\mathbb{U}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{F}$

By contrast to the Kleene logics with which they share truth tables, these *do* have tautologies.

Proposition 5.9. *The tautologies of* ***LP*** *are the same as the tautologies of classical propositional logic.*

Proof. By Proposition 4.12, if $\vDash_{\boldsymbol{LP}} A$ then $\vDash_{\boldsymbol{C}} A$. To show the reverse, we show that if there is a valuation $v\colon \mathrm{At}_0 \to \{\mathbb{F}, \mathbb{T}, \mathbb{U}\}$ such that $\overline{v}_{\boldsymbol{Ks}}(A) = \mathbb{F}$ then there is a valuation $v'\colon \mathrm{At}_0 \to \{\mathbb{F}, \mathbb{T}\}$ such that $\overline{v'}_{\boldsymbol{C}}(A) = \mathbb{F}$. This establishes the result for ***LP***, since ***Ks*** and ***LP*** have the same characteristic truth functions, and $\mathbb{F}$ is the only truth value of ***LP*** that is not designated (that is the only difference between ***LP*** and ***Ks***). Thus, if $\nvDash_{\boldsymbol{LP}} A$, for some valuation v, $\overline{v}_{\boldsymbol{LP}}(A) = \overline{v}_{\boldsymbol{Ks}}(A) = \mathbb{F}$. By the claim we're proving, $\overline{v'}_{\boldsymbol{C}}(A) = \mathbb{F}$, i.e., $\nvDash_{\boldsymbol{C}} A$.

To establish the claim, we first define v' as

$$v'(p) = \begin{cases} \mathbb{T} & \text{if } v(p) \in \{\mathbb{T}, \mathbb{U}\} \\ \mathbb{F} & \text{otherwise} \end{cases}$$

We now show by induction on A that (a) if $\overline{v}_{\boldsymbol{Ks}}(A) = \mathbb{F}$ then $\overline{v'}_{\boldsymbol{C}}(A) = \mathbb{F}$, and (b) if $\overline{v}_{\boldsymbol{Ks}}(A) = \mathbb{T}$ then $\overline{v'}_{\boldsymbol{C}}(A) = \mathbb{T}$

1. Induction basis: $A \equiv p$. By Definition 4.8, $\overline{v}_{\boldsymbol{Ks}}(A) = v(p) = \overline{v'}_{\boldsymbol{C}}(A)$, which implies both (a) and (b).

   For the induction step, consider the cases:

2. $A \equiv \neg B$.

   a) Suppose $\overline{v}_{Ks}(\neg B) = \mathbb{F}$. By the definition of $\widetilde{\neg}_{Ks}$, $\overline{v}_{Ks}(B) = \mathbb{T}$. By inductive hypothesis, case (b), we get $\overline{v'}_C(B) = \mathbb{T}$, so $\overline{v'}_C(\neg B) = \mathbb{F}$.

   b) Suppose $\overline{v}_{Ks}(\neg B) = \mathbb{T}$. By the definition of $\widetilde{\neg}_{Ks}$, $\overline{v}_{Ks}(B) = \mathbb{F}$. By inductive hypothesis, case (a), we get $\overline{v'}_C(B) = \mathbb{F}$, so $\overline{v'}_C(\neg B) = \mathbb{T}$.

3. $A \equiv (B \wedge C)$.

   a) Suppose $\overline{v}_{Ks}(B \wedge C) = \mathbb{F}$. By the definition of $\widetilde{\wedge}_{Ks}$, $\overline{v}_{Ks}(B) = \mathbb{F}$ or $\overline{v}_{Ks}(B) = \mathbb{F}$. By inductive hypothesis, case (a), we get $\overline{v'}_C(B) = \mathbb{F}$ or $\overline{v'}_C(C) = \mathbb{F}$, so $\overline{v'}_C(B \wedge C) = \mathbb{F}$.

   b) Suppose $\overline{v}_{Ks}(B \wedge C) = \mathbb{T}$. By the definition of $\widetilde{\wedge}_{Ks}$, $\overline{v}_{Ks}(B) = \mathbb{T}$ and $\overline{v}_{Ks}(B) = \mathbb{T}$. By inductive hypothesis, case (b), we get $\overline{v'}_C(B) = \mathbb{T}$ and $\overline{v'}_C(C) = \mathbb{T}$, so $\overline{v'}_C(B \wedge C) = \mathbb{T}$.

The other two cases are similar, and left as exercises. Alternatively, the proof above establishes the result for all formulas only containing $\neg$ and $\wedge$. One may now appeal to the facts that in both ***Ks*** and ***C***, for any v, $\overline{v}(B \vee C) = \overline{v}(\neg(\neg B \wedge \neg C))$ and $\overline{v}(B \rightarrow C) = \overline{v}(\neg(B \wedge \neg C))$. □

Although they have the same tautologies as classical logic, their consequence relations are different. ***LP***, for instance, is *paraconsistent* in that $\neg p, p \nvDash q$, and so the principle of explosion $\neg A, A \vDash B$ does not hold in general. (It holds for some cases of A and B, e.g., if B is a tautology.)

What if you make $\mathbb{U}$ designated in ***Ł***$_3$?

Definition 5.10. The logic *3-valued R-Mingle* ***RM***$_3$ is defined using the matrix:

1. The standard propositional language $\mathcal{L}_0$ with $\bot$, $\neg$, $\wedge$, $\vee$,

$\to$.

2. The set of truth values $V = \{\mathbb{T}, \mathbb{U}, \mathbb{F}\}$.
3. $\mathbb{T}$ and $\mathbb{U}$ are designated, i.e., $V^+ = \{\mathbb{T}, \mathbb{U}\}$.
4. Truth functions are the same as Łukasiewicz logic $\L_3$.

Different truth tables can sometimes generate the same logic (entailment relation) just by changing the designated values. E.g., this happens if in Gödel logic we take $V^+ = \{\mathbb{T}, \mathbb{U}\}$ instead of $\{\mathbb{T}\}$.

Proposition 5.11. *The matrix with $V = \{\mathbb{F}, \mathbb{U}, \mathbb{T}\}$, $V^+ = \{\mathbb{T}, \mathbb{U}\}$, and the truth functions of 3-valued Gödel logic defines classical logic.*

Proof. Exercise. □

Problems

Problem 5.1. Suppose we define $\overline{v}(A \leftrightarrow B) = \overline{v}((A \to B) \land (B \to A))$ in $\L_3$. What truth table would $\leftrightarrow$ have?

Problem 5.2. Show that the following are tautologies in $\L_3$:

1. $p \to (q \to p)$
2. $\neg(p \land q) \leftrightarrow (\neg p \lor \neg q)$
3. $\neg(p \lor q) \leftrightarrow (\neg p \land \neg q)$

(In (2) and (3), take $A \leftrightarrow B$ as an abbreviation for $(A \to B) \land (B \to A)$, or refer to your solution to problem 5.1.)

Problem 5.3. Show that the following classical tautologies are not tautologies in $\L_3$:

1. $(\neg p \land p) \to q)$
2. $((p \to q) \to p) \to p$

3. $(p \to (p \to q)) \to (p \to q)$

Problem 5.4. Which of the following relations hold in Łukasiewicz logic? Give a truth table for each.

1. $p, p \to q \vDash q$
2. $\neg\neg p \vDash p$
3. $p \land q \vDash p$
4. $p \vDash p \land p$
5. $p \vDash p \lor q$

Problem 5.5. Show that $\Box p \leftrightarrow \neg\Diamond\neg p$ and $\Diamond p \leftrightarrow \neg\Box\neg p$ are tautologies in $Ł_3$, extended with the truth tables for $\Box$ and $\Diamond$.

Problem 5.6. Which of the following relations hold in (a) strong and (b) weak Kleene logic? Give a truth table for each.

1. $p, p \to q \vDash q$
2. $p \lor q, \neg p \vDash q$
3. $p \land q \vDash p$
4. $p \vDash p \land p$
5. $p \vDash p \lor q$

Problem 5.7. Can you define $\sim$ in Bochvar's logic in terms of $\neg$ and $+$, i.e., find a formula with only the propositional variable p and not involving $\sim$ which always takes the same truth value as $\sim p$? Give a truth table to show you're right.

Problem 5.8. Give truth tables to show that the following are tautologies of G_3:

$$\neg\neg p \lor \neg p$$

$$(p \to q) \lor (q \to p)$$
$$\neg(p \land q) \to (\neg p \lor \neg q)$$
$$(p \to q) \lor (q \to r) \lor (r \to s)$$

Problem 5.9. Give truth tables that show that the following are not tautologies of G_3

$$(p \to q) \to (\neg p \lor q)$$
$$\neg(\neg p \land \neg q) \to (p \lor q)$$
$$((p \to q) \to p) \to p$$
$$\neg(p \to q) \to (p \land \neg q)$$

Problem 5.10. Which of the following relations hold in Gödel logic? Give a truth table for each.

1. $p, p \to q \vDash q$
2. $p \lor q, \neg p \vDash q$
3. $p \land q \vDash p$
4. $p \vDash p \land p$
5. $p \vDash p \lor q$

Problem 5.11. Complete the proof Proposition 5.9, i.e., establish (a) and (b) for the cases where $A \equiv (B \lor C)$ and $A \equiv (B \to C)$.

Problem 5.12. Prove that every classical tautology is a tautology in ***Hal***.

Problem 5.13. Which of the following relations hold in (a) ***LP*** and in (b) ***Hal***? Give a truth table for each.

1. $p, p \to q \vDash q$
2. $\neg q, p \to q \vDash \neg p$
3. $p \lor q, \neg p \vDash q$

4. $\neg p, p \vDash q$

5. $p \vDash p \lor q$

6. $p \to q, q \to r \vDash p \to r$

Problem 5.14. Which of the following relations hold in $\mathbf{RM}_3$?

1. $p, p \to q \vDash q$

2. $p \lor q, \neg p \vDash q$

3. $\neg p, p \vDash q$

4. $p \vDash p \lor q$

Problem 5.15. Prove Proposition 5.11 by showing that for the logic $\mathbf{L}$ defined just like Gödel logic but with $V^+ = \{\mathbb{T}, \mathbb{U}\}$, if $\Gamma \nvDash_{\mathbf{L}} B$ then $\Gamma \nvDash_{\mathbf{C}} B$. Use the ideas of Proposition 5.9, except instead of proving properties (a) and (b), show that $\overline{v}_{\mathbf{G}}(A) = \mathbb{F}$ iff $\overline{v'}_{\mathbf{C}}(A) = \mathbb{F}$ (and hence that $\overline{v}_{\mathbf{G}}(A) \in \{\mathbb{T}, \mathbb{U}\}$ iff $\overline{v'}_{\mathbf{C}}(A) = \mathbb{T}$). Explain why this establishes the proposition.

CHAPTER 6

Sequent Calculus

6.1 Introduction

The sequent calculus for classical logic is an efficient and simple derivation system. If a many-valued logic is defined by a matrix with finitely many truth values, i.e., V is finite, it is possible to provide a sequent calculus for it. The idea for how to do this comes from considering the meanings of sequents and the form of inference rules in the classical case.

Now recall that a sequent

$$A_1, \dots, A_n \Rightarrow B_1, \dots, B_n$$

can be interpreted as the formula

$$(A_1 \land \cdots \land A_m) \to (B_1 \lor \cdots \lor B_n)$$

In other words, A valuation $\mathfrak{v}$ *satisfies* a sequent $\Gamma \Rightarrow \Delta$ iff either $\overline{\mathfrak{v}}(A) = \mathbb{F}$ for some $A \in \Gamma$ or $\overline{\mathfrak{v}}(A) = \mathbb{T}$ for some $A \in \Delta$. On this interpretation, initial sequents $A \Rightarrow A$ are always satisfied, because either $\overline{\mathfrak{v}}(A) = \mathbb{T}$ or $\overline{(}A) = \mathbb{F}$.

Here are the inference rules for the conditional in ***LK***, with side formulas Γ, Δ left out:

$$\frac{\Rightarrow A \qquad B \Rightarrow}{A \to B \Rightarrow} \to\text{L} \qquad\qquad \frac{A \Rightarrow B}{\Rightarrow A \to B} \to\text{R}$$

If we apply the above semantic interpretation of a sequent, we can read the $\to$L rule as saying that if $\overline{v}(A) = \mathbb{T}$ and $\overline{v}(B) = \mathbb{F}$, then $\overline{v}(A \to B) = \mathbb{F}$. Similarly, the $\to$R rule says that if either $\overline{v}(A) = \mathbb{F}$ or $\overline{v}(B) = \mathbb{T}$, then $\overline{v}(A \to B) = \mathbb{T}$. And in fact, these conditionals are actually biconditionals. In the case of the $\wedge$L and $\vee$R rules in their standard formulation, the corresponding conditionals would not be biconditionals. But there are alternative versions of these rules where they are:

$$\frac{A, B, \Gamma \Rightarrow \Delta}{A \wedge B, \Gamma \Rightarrow \Delta} \wedge\text{L} \qquad\qquad \frac{\Gamma \Rightarrow \Delta, A, B}{\Gamma \Rightarrow \Delta, A \vee B} \vee\text{R}$$

This basic idea, applied to an n-valued logic, then results in a sequent calculus with n instead of two places, one for each truth value. For a three-valued logic with $V = \{\mathbb{F}, \mathbb{U}, \mathbb{T}\}$, a sequent is an expression $\Gamma \mid \Pi \mid \Delta$. It is satisfied in a valuation v iff either $\overline{v}(A) = \mathbb{F}$ for some $A \in \Gamma$ or $\overline{v}(A) = \mathbb{T}$ for some $A \in \Delta$ or $\overline{v}(A) = \mathbb{U}$ for some $A \in \Pi$. Consequently, initial sequents $A \mid A \mid A$ are always satisfied.

6.2 Rules and Derivations

For the following, let $\Gamma, \Delta, \Pi, \Lambda$ represent finite sequences of sentences.

Definition 6.1 (Sequent). An *n-sided sequent* is an expression of the form

$$\Gamma_1 \mid \ldots \mid \Gamma_n$$

where each Γ_1 is a finite (possibly empty) sequences of sentences of the language $\mathcal{L}$.

Definition 6.2 (Initial Sequent). An *n-sided initial sequent* is an n-sided sequent of the form $A \mid \ldots \mid A$ for any sentence A in the language.

If the language contains a 0-place connective $\star$, i.e., a propositional constant, then we also take the sequent $\ldots \mid \star \mid \ldots$ where $\star$ appears in the space for the truth value associated with $\widetilde{\star} \in V$, and is empty otherwise.

For each connective of an n-valued logic $\boldsymbol{L}$, there is a logical rule for each truth value that this connective can take in $\boldsymbol{L}$. Derivations in an n-sided sequent calculus for $\boldsymbol{L}$ are trees of sequents, where the topmost sequents are initial sequents, and if a sequent stands below one or more other sequents, it must follow correctly by a rule of inference for the connectives of $\boldsymbol{L}$.

Definition 6.3 (Theorems). A sentence A is a *theorem* of an n-valued logic $\boldsymbol{L}$ if there is a derivation of the n-sequent containing A in each position corresponding to a designated truth value of $\boldsymbol{L}$. We write $\vdash_{\boldsymbol{L}} A$ if A is a theorem and $\nvdash_{\boldsymbol{L}} A$ if it is not.

Definition 6.4 (Derivability). A sentence A is *derivable from* a set of sentences Γ in an n-valued logic $\boldsymbol{L}$, $\Gamma \vdash_{\boldsymbol{L}} A$, iff there is a finite subset $\Gamma_0 \subseteq \Gamma$ and a sequence Γ_0' of the sentences in Γ_0 such that the following sequent has a derivation:

$$\Lambda_1 \mid \ldots \mid \Lambda_n$$

where Λ_i is A if position i corresponds to a designated truth value, and Γ_0'otherwise. If A is not derivable from Γ we write $\Gamma \nvdash A$.

For instance, 3-valued Łukasiewicz logic has a 3-sided sequent calculus. In a 3-sided sequent $\Gamma \mid \Pi \mid \Delta$, Γ corresponds to $\mathbb{F}$, Δ to $\mathbb{T}$, and Π to $\mathbb{U}$. Axioms are $A \mid A \mid A$. Since only $\mathbb{T}$ is designated, $\Gamma \vdash_{\boldsymbol{Ł}_3} A$ iff the sequent $\Gamma \mid \Gamma \mid A$ has a derivation. (If $\mathbb{U}$ were also designated, we would need a derivation of $\Gamma \mid A \mid A$.)

6.3 Structural Rules

The structural rules for n-sided sequent calculus operate as in the classical case, except for each position i.

$$\frac{\Gamma_1 \mid \dots \mid \quad \Gamma_i \mid \dots \mid \Gamma_n}{\Gamma_1 \mid \dots \mid A, \Gamma_i \mid \dots \mid \Gamma_n}\,\mathrm{W}i$$

$$\frac{\Gamma_1 \mid \dots \mid A, A, \Gamma_i \mid \dots \mid \Gamma_n}{\Gamma_1 \mid \dots \mid \quad A, \Gamma_i \mid \dots \mid \Gamma_n}\,\mathrm{C}i$$

$$\frac{\Gamma_1 \mid \dots \mid \Gamma_i, A, B, \Gamma_i' \mid \dots \mid \Gamma_n}{\Gamma_1 \mid \dots \mid \Gamma_i, B, A, \Gamma_i' \mid \dots \mid \Gamma_n}\,\mathrm{X}i$$

A series of weakening, contraction, and exchange inferences will often be indicated by double inference lines.

The Cut rule comes in several forms, one for every combination of distinct positions in the sequent $i \neq j$:

$$\frac{\Gamma_1 \mid \dots \mid A, \Gamma_i \mid \dots \mid \Gamma_n \qquad \Delta_1 \mid \dots \mid A, \Delta_j \mid \dots \mid \Delta_n}{\Gamma_1, \Delta_1 \mid \dots \mid \Gamma_n, \Delta_n}\,\mathrm{Cut}i, j$$

6.4 Propositional Rules for Selected Logics

The inference rules for a connective in an n-sided sequent calculus only depend on the characteristic truth function for the connective. Thus, if some connective is defined by the same truth function in different logics, these n-sided sequent rules for the connective are the same in those logics.

Rules for ¬

The following rules for ¬ apply to Łukasiewicz and Kleene logics, and their variants.

$$\frac{\Gamma \mid \Pi \mid \Delta, A}{\neg A, \Gamma \mid \Pi \mid \Delta} \neg\mathbb{F}$$

$$\frac{\Gamma \mid A, \Pi \mid \Delta}{\Gamma \mid \neg A, \Pi \mid \Delta} \neg\mathbb{U}$$

$$\frac{A, \Gamma \mid \Pi \mid \Delta}{\Gamma \mid \Pi \mid \Delta, \neg A} \neg\mathbb{T}$$

The following rules for $\neg$ apply to Gödel logic.

$$\frac{\Gamma \mid A, \Pi \mid \Delta, A}{\neg A, \Gamma \mid \Pi \mid \Delta} \neg_G\mathbb{F} \qquad \frac{A, \Gamma \mid \Pi \mid \Delta}{\Gamma \mid \Pi \mid \Delta, \neg A} \neg_G\mathbb{T}$$

(In Gödel logic, $\neg A$ can never take the value $\mathbb{U}$, so there is no rule for the middle position.)

Rules for $\wedge$

These are the rules for $\wedge$ in Łukasiewicz, strong Kleene, and Gödel logic.

$$\frac{A, B, \Gamma \mid \Pi \mid \Delta}{A \wedge B, \Gamma \mid \Pi \mid \Delta} \wedge\mathbb{F}$$

$$\frac{\Gamma \mid A, \Pi \mid A, \Delta \qquad \Gamma \mid B, \Pi \mid B, \Delta \qquad \Gamma \mid A, B, \Pi \mid \Delta}{\Gamma \mid A \wedge B, \Pi \mid \Delta} \wedge\mathbb{U}$$

$$\frac{\Gamma \mid \Pi \mid \Delta, A \qquad \Gamma \mid \Pi \mid \Delta, B}{\Gamma \mid \Pi \mid \Delta, A \wedge B} \wedge\mathbb{T}$$

Rules for $\vee$

These are the rules for $\vee$ in Łukasiewicz, strong Kleene, and Gödel logic.

$$\frac{A, \Gamma \mid \Pi \mid \Delta \qquad B, \Gamma \mid \Pi \mid \Delta}{A \vee B, \Gamma \mid \Pi \mid \Delta} \vee\mathbb{F}$$

$$\frac{A, \Gamma \mid A, \Pi \mid \Delta \qquad B, \Gamma \mid B, \Pi \mid \Delta \qquad \Gamma \mid A, B, \Pi \mid \Delta}{\Gamma \mid A \vee B, \Pi \mid \Delta} \vee\mathbb{U}$$

$$\frac{\Gamma \mid \Pi \mid \Delta, A, B}{\Gamma \mid \Pi \mid \Delta, A \vee B} \vee\mathbb{T}$$

Rules for $\to$

These are the rules for $\to$ in Łukasiewicz logic.

$$\frac{\Gamma \mid \Pi \mid \Delta, A \qquad B, \Gamma \mid \Pi \mid \Delta}{A \to B, \Gamma \mid \Pi \mid \Delta} \to_{Ł_3}\mathbb{F}$$

$$\frac{\Gamma \mid A, B, \Pi \mid \Delta \qquad B, \Gamma \mid \Pi \mid \Delta, A}{\Gamma \mid A \to B, \Pi \mid \Delta} \to_{Ł_3}\mathbb{U}$$

$$\frac{A, \Gamma \mid B, \Pi \mid \Delta, B \qquad A, \Gamma \mid A, \Pi \mid \Delta, B}{\Gamma \mid \Pi \mid \Delta, A \to B} \to_{Ł_3}\mathbb{T}$$

These are the rules for $\to$ in strong Kleene logic.

$$\frac{\Gamma \mid \Pi \mid \Delta, A \qquad B, \Gamma \mid \Pi \mid \Delta}{A \to B, \Gamma \mid \Pi \mid \Delta} \to_{Ks}\mathbb{F}$$

$$\frac{B, \Gamma \mid B, \Pi \mid \Delta \qquad \Gamma \mid A, B, \Pi \mid \Delta \qquad \Gamma \mid A, \Pi \mid \Delta, A}{\Gamma \mid A \to B, \Pi \mid \Delta} \to_{Ks}\mathbb{U}$$

$$\frac{A, \Gamma \mid \Pi \mid \Delta, B}{\Gamma \mid \Pi \mid \Delta, A \to B} \to_{Ks}\mathbb{T}$$

These are the rules for $\to$ in Gödel logic.

$$\frac{\Gamma \mid A, \Pi \mid \Delta, A \qquad B, \Gamma \mid \Pi \mid \Delta}{A \to B, \Gamma \mid \Pi \mid \Delta} \to_{G_3}\mathbb{F}$$

$$\frac{\Gamma \mid B, \Pi \mid \Delta \qquad \Gamma \mid \Pi \mid \Delta, A}{\Gamma \mid A \to B, \Pi \mid \Delta} \to_{G_3} \mathbb{U}$$

$$\frac{A, \Gamma \mid B, \Pi \mid \Delta, B \qquad A, \Gamma \mid A, \Pi \mid \Delta, B}{\Gamma \mid \Pi \mid \Delta, A \to B} \to_{G_3} \mathbb{T}$$

$$
\dfrac{
\dfrac{
\dfrac{\dfrac{\dfrac{A \mid A \mid A}{A \mid A \mid B,A}\,\mathrm{W}\mathbb{T}}{A \mid B,A \mid B,A}\,\mathrm{W}\mathbb{U}}{A \mid A,B,A \mid B,A}\,\mathrm{W}\mathbb{U}
\qquad
\dfrac{\dfrac{\dfrac{A \mid A \mid A}{A \mid A \mid A,A}\,\mathrm{W}\mathbb{T}}{A \mid A \mid B,A,A}\,\mathrm{W}\mathbb{T}}{B,A \mid A \mid B,A,A}\,\mathrm{W}\mathbb{F}
}{A \mid A \to B,A \mid B,A}\,{\to}\mathbb{U}
\qquad
\dfrac{
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{B \mid B \mid B}{B \mid A,B \mid B}\,\mathrm{W}\mathbb{U}}{B \mid B,A \mid B}\,\mathrm{X}\mathbb{U}}{B \mid A,B,A \mid B}\,\mathrm{W}\mathbb{U}}{A,B \mid A,B,A \mid B}\,\mathrm{W}\mathbb{F}}{B,A \mid A,B,A \mid B}\,\mathrm{X}\mathbb{F}
\qquad
\dfrac{\dfrac{\dfrac{A \mid A \mid A}{A \mid A \mid B,A}\,\mathrm{W}\mathbb{T}}{B,A \mid A \mid B,A}\,\mathrm{W}\mathbb{F}}{B,B,A \mid A \mid B,A}\,\mathrm{W}\mathbb{F}
}{B,A \mid A \to B,A \mid B}\,{\to}\mathbb{U}
}{A \to B,A \mid A \to B,A \mid B}\,{\to}\mathbb{F}
$$

Figure 6.1: Example derivation in $Ł_3$

CHAPTER 7

Infinite-valued Logics

7.1 Introduction

The number of truth values of a matrix need not be finite. An obvious choice for a set of infinitely many truth values is the set of rational numbers between 0 and 1, $V_\infty = [0,1] \cap \mathbb{Q}$, i.e.,

$$V_\infty = \{\frac{n}{m} : n, m \in \mathbb{N} \text{ and } n \le m\}.$$

When considering this infinite truth value set, it is often useful to also consider the subsets

$$V_m = \{\frac{n}{m-1} : n \in \mathbb{N} \text{ and } n \le m\}$$

For instance, V_5 is the set with 5 evenly spaced truth values,

$$V_5 = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}.$$

In logics based on these truth value sets, usually only 1 is designated, i.e., $V^+ = \{1\}$. In other words, we let 1 play the role of (absolute) truth, 0 as absolute falsity, but formulas may take any intermediate value in V.

One can also consider the set $V_{[0,1]} = [0,1]$ of all *real* numbers between 0 and 1, or other infinite subsets of $[0,1]$, however. Logics with this truth value set are often called *fuzzy*.

7.2 Łukasiewicz logic

Definition 7.1. Infinite-valued Łukasiewicz logic Ł_∞ is defined using the matrix:

1. The standard propositional language $\mathcal{L}_0$ with $\neg$, $\wedge$, $\vee$, $\to$.
2. The set of truth values V_∞.
3. 1 is the only designated value, i.e., $V^+ = \{1\}$.
4. Truth functions are given by the following functions:

$$\begin{aligned}
\widetilde{\neg}_{\text{Ł}}(x) &= 1 - x \\
\widetilde{\wedge}_{\text{Ł}}(x, y) &= \min(x, y) \\
\widetilde{\vee}_{\text{Ł}}(x, y) &= \max(x, y) \\
\widetilde{\to}_{\text{Ł}}(x, y) &= \min(1, 1 - (x - y)) = \begin{cases} 1 & \text{if } x \leq y \\ 1 - (x - y) & \text{otherwise.} \end{cases}
\end{aligned}$$

m-valued Łukasiewicz logic is defined the same, except $V = V_m$.

Proposition 7.2. *The logic Ł_3 defined by Definition 5.1 is the same as Ł_3 defined by Definition 7.1.*

Proof. This can be seen by comparing the truth tables for the connectives given in Definition 5.1 with the truth tables determined by the equations in Definition 7.1:

$\widetilde{\neg}$	
1	0
1/2	1/2
0	1

$\widetilde{\wedge}_{Ł_3}$	1	1/2	0
1	1	1/2	0
1/2	1/2	1/2	0
0	0	0	0

$\widetilde{\vee}_{Ł_3}$	1	1/2	0
1	1	1	1
1/2	1	1/2	1/2
0	1	1/2	0

$\widetilde{\rightarrow}_{Ł_3}$	1	1/2	0
1	1	1/2	0
1/2	1	1	1/2
0	1	1	1

□

Proposition 7.3. *If $\Gamma \vDash_{Ł_\infty} B$ then $\Gamma \vDash_{Ł_m} B$ for all $m \geq 2$.*

Proof. Exercise. □

In fact, the converse holds as well.

Infinite-valued Łukasiewicz logic is the most popular fuzzy logic. In the fuzzy logic literature, the conditional is often defined as $\neg A \vee B$. The result would be an infinite-valued strong Kleene logic.

7.3 Gödel logics

Definition 7.4. Infinite-valued Gödel logic G_∞ is defined using the matrix:

1. The standard propositional language $\mathcal{L}_0$ with $\bot$, $\neg$, $\wedge$, $\vee$, $\rightarrow$.
2. The set of truth values V_∞.
3. 1 is the only designated value, i.e., $V^+ = \{1\}$.
4. Truth functions are given by the following functions:

$$\widetilde{\bot} = 0$$

$$\widetilde{\neg}_G(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\widetilde{\land}_G(x, y) = \min(x, y)$$

$$\widetilde{\lor}_G(x, y) = \max(x, y)$$

$$\widetilde{\to}_G(x, y) = \begin{cases} 1 & \text{if } x \leq y \\ y & \text{otherwise.} \end{cases}$$

m-valued Gödel logic is defined the same, except $V = V_m$.

Proposition 7.5. *The logic* G_3 *defined by Definition 5.6 is the same as* G_3 *defined by Definition 7.4.*

Proof. This can be seen by comparing the truth tables for the connectives given in Definition 5.6 with the truth tables determined by the equations in Definition 7.4:

$\widetilde{\neg}_{G_3}$	
1	0
1/2	0
0	1

$\widetilde{\land}_G$	1	1/2	0
1	1	1/2	0
1/2	1/2	1/2	0
0	0	0	0

$\widetilde{\lor}_G$	1	1/2	0
1	1	1	1
1/2	1	1/2	1/2
0	1	1/2	0

$\widetilde{\to}_G$	1	1/2	0
1	1	1/2	0
1/2	1	1	0
0	1	1	1

□

Proposition 7.6. *If* $\Gamma \vDash_{G_\infty} B$ *then* $\Gamma \vDash_{G_m} B$ *for all* $m \geq 2$.

Proof. Exercise. □

In fact, the converse holds as well.

Like $\mathbf{G}_3$, $\mathbf{G}_\infty$ has all intuitionistically valid formulas as tautologies, and the same examples of non-tautologies are non-tautologies of $\mathbf{G}_\infty$:

$$
\begin{array}{ll}
p \lor \neg p & (p \to q) \to (\neg p \lor q) \\
\neg\neg p \to p & \neg(\neg p \land \neg q) \to (p \lor q) \\
((p \to q) \to p) \to p & \neg(p \to q) \to (p \land \neg q)
\end{array}
$$

The example of an intuitionistically invalid formula that is nevertheless a tautology of $\mathbf{G}_3$, $(p \to q) \lor (q \to p)$, is also a tautology in $\mathbf{G}_\infty$. In fact, $\mathbf{G}_\infty$ can be characterized as intuitionistic logic to which the schema $(A \to B) \lor (B \to A)$ is added. This was shown by Michael Dummett, and so $\mathbf{G}_\infty$ is often referred to as Gödel–Dummett logic $\mathbf{LC}$.

Problems

Problem 7.1. Prove Proposition 7.3.

Problem 7.2. Show that $(p \to q) \lor (q \to p)$ is a tautology of $\mathbf{Ł}_\infty$.

Problem 7.3. Prove Proposition 7.6.

Problem 7.4. Show that $(p \to q) \lor (q \to p)$ is a tautology of $\mathbf{G}_\infty$.

Problem 7.5. Show that $(p \to q) \lor (q \to r) \lor (r \to s)$, which is a tautology of $\mathbf{G}_3$, is not a tautology of $\mathbf{G}_\infty$.

PART III

But isn't truth relative (to a world)?

CHAPTER 8

Syntax and Semantics

8.1 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 $A \strictif B$. 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.

8.2 The Language of Basic Modal Logic

Definition 8.1. The basic language of modal logic contains

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

Definition 8.2. *Formulas* of the basic modal language are inductively defined as follows:

1. $\bot$ is an atomic formula.
2. Every propositional variable p_i is an (atomic) formula.
3. If A is a formula, then $\neg A$ is a formula.
4. If A and B are formulas, then $(A \land B)$ is a formula.
5. If A and B are formulas, then $(A \lor B)$ 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 8.3. Formulas constructed using the defined operators are to be understood as follows:

1. $\top$ abbreviates $\neg\bot$.
2. $A \leftrightarrow B$ abbreviates $(A \rightarrow B) \land (B \rightarrow A)$.

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

8.3 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 8.4. Where A is a modal formula all of whose propositional variables are among p_1, …, p_n, and D_1, …, D_n are also modal formulas, we define $A[D_1/p_1,\ldots,D_n/p_n]$ as the result of simultaneously substituting each D_i for p_i in A. Formally, this is a definition by induction on A:

1. $A \equiv \bot$: $A[D_1/p_1,\ldots,D_n/p_n]$ is $\bot$.
2. $A \equiv q$: $A[D_1/p_1,\ldots,D_n/p_n]$ is q, provided $q \not\equiv p_i$ for $i = 1$, …, n.
3. $A \equiv p_i$: $A[D_1/p_1,\ldots,D_n/p_n]$ is D_i.
4. $A \equiv \neg B$: $A[D_1/p_1,\ldots,D_n/p_n]$ is $\neg B[D_1/p_1,\ldots,D_n/p_n]$.
5. $A \equiv (B \wedge C)$: $A[D_1/p_1,\ldots,D_n/p_n]$ is
$$(B[D_1/p_1,\ldots,D_n/p_n] \wedge C[D_1/p_1,\ldots,D_n/p_n]).$$
6. $A \equiv (B \vee C)$: $A[D_1/p_1,\ldots,D_n/p_n]$ is
$$(B[D_1/p_1,\ldots,D_n/p_n] \vee C[D_1/p_1,\ldots,D_n/p_n]).$$
7. $A \equiv (B \to C)$: $A[D_1/p_1,\ldots,D_n/p_n]$ is
$$(B[D_1/p_1,\ldots,D_n/p_n] \to C[D_1/p_1,\ldots,D_n/p_n]).$$
8. $A \equiv (B \leftrightarrow C)$: $A[D_1/p_1,\ldots,D_n/p_n]$ is
$$(B[D_1/p_1,\ldots,D_n/p_n] \leftrightarrow C[D_1/p_1,\ldots,D_n/p_n]).$$
9. $A \equiv \Box B$: $A[D_1/p_1,\ldots,D_n/p_n]$ is $\Box B[D_1/p_1,\ldots,D_n/p_n]$.
10. $A \equiv \Diamond B$: $A[D_1/p_1,\ldots,D_n/p_n]$ is $\Diamond B[D_1/p_1,\ldots,D_n/p_n]$.

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

Example 8.5. Suppose A is $p_1 \to \Box(p_1 \land p_2)$, D_1 is $\Diamond(p_2 \to p_3)$ and D_2 is $\neg\Box p_1$. Then $A[D_1/p_1, D_2/p_2]$ is

$$\Diamond(p_2 \to p_3) \to \Box(\Diamond(p_2 \to p_3) \land \neg\Box p_1)$$

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

$$\neg\Box p_1 \to \Box(\neg\Box p_1 \land \Diamond(p_2 \to p_3))$$

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:

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

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

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

8.4 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 8.6. 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.

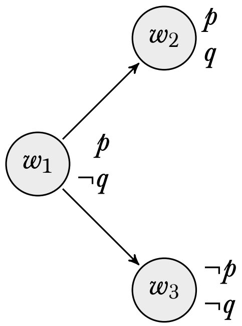


Figure 8.1: A simple model.

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 8.1. 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 8.1 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 $V(q) = \{w_2\}$.

8.5 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 8.7. *Truth of a formula A at w* in a M, in symbols: $M, w \Vdash A$, is defined inductively as follows:

1. $A \equiv \bot$: Never $M, w \Vdash \bot$.
2. $M, w \Vdash p$ iff $w \in V(p)$.
3. $A \equiv \neg B$: $M, w \Vdash A$ iff $M, w \nVdash B$.
4. $A \equiv (B \wedge C)$: $M, w \Vdash A$ iff $M, w \Vdash B$ and $M, w \Vdash C$.
5. $A \equiv (B \vee C)$: $M, w \Vdash A$ iff $M, w \Vdash B$ or $M, w \Vdash C$ (or both).
6. $A \equiv (B \to C)$: $M, w \Vdash A$ iff $M, w \nVdash B$ or $M, w \Vdash C$.
7. $A \equiv \Box B$: $M, w \Vdash A$ iff $M, w' \Vdash B$ for all $w' \in W$ with Rww'.
8. $A \equiv \Diamond B$: $M, w \Vdash A$ iff $M, w' \Vdash B$ for at least one $w' \in W$ with Rww'.

Note that by clause (7), a formula $\Box B$ is true at w whenever there are no w' with 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 $M, w \nVdash \Diamond A$, for any A.

Proposition 8.8.

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

Proof.

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

8.6 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 8.9. A formula A is *true in a model* $M = \langle W, R, V\rangle$, written $M \Vdash A$, if and only if $M, w \Vdash A$ for every $w \in W$.

Proposition 8.10.

1. *If $M \Vdash A$ then $M \nVdash \neg A$, but* not *vice-versa.*
2. *If $M \Vdash A \rightarrow B$ then $M \Vdash A$ only if $M \Vdash B$, but* not *vice-versa.*

Proof.

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

   On the other hand, if $M \nVdash \neg A$ then A is true at some world $w \in W$. It does not follow that $M, w \Vdash A$ for *every* $w \in W$. For instance, in the model of Figure 8.1, $M \nVdash \neg p$, and also $M \nVdash p$.

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

   To show that the converse fails, we need to find a model M such that $M \Vdash A$ only if $M \Vdash B$, but $M \nVdash A \rightarrow B$. Consider again the model of Figure 8.1: $M \nVdash p$ and hence (vacuously) $M \Vdash p$ only if $M \Vdash q$. However, $M \nVdash p \rightarrow q$, as p is true but q false at w_1. □

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

Definition 8.11. A formula A is *valid* in a class $\mathscr{C}$ of models if it is true in every model in $\mathscr{C}$ (i.e., true at every world in every model in $\mathscr{C}$). If A is valid in $\mathscr{C}$, we write $\mathscr{C} \vDash A$, and we write $\vDash A$ if A is valid in the class of *all* models.

Proposition 8.12. *If A is valid in $\mathscr{C}$ it is also valid in each class $\mathscr{C}' \subseteq \mathscr{C}$.*

Proposition 8.13. *If A is valid, then so is $\Box A$.*

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

8.8 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 $\Box p \lor \neg\Box p$—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 8.14. 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 8.15. *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 $v \vDash A$ if and only if $M, w \Vdash A[D_1/p_1, \dots, D_n/p_n]$.*

Proof. By induction on A.

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

2. $A \equiv p_i$:

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

3. $A \equiv \neg B$:

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

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

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

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

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

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

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

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

by definition of $M, w \Vdash$.

□

Proposition 8.16. *All tautological instances are valid.*

Proof. Contrapositively, suppose A is such that $M, w \nVdash A[D_1/p_1, \ldots, D_n/p_n]$, for some model M and world w. Define an assignment v such that $v(p_i) = \mathbb{T}$ if and only if $M, w \Vdash D_i$ (and v assigns arbitrary values to $q \notin \{p_1, \ldots, p_n\}$). Then by Lemma 8.15, $v \nvDash A$, so A is not a tautology. □

8.9 Schemas and Validity

Definition 8.17. A *schema* is a set of formulas comprising all and only the substitution instances of some modal formula C, i.e.,

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

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

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

Definition 8.18. 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 8.19. *The following schema* K *is valid*

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

Proof. We need to show that all instances of the schema are true at every world in every model. So let $M = \langle W, R, V \rangle$ 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 $M, w \Vdash \Box(A \to B)$ 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 $M, w' \Vdash A \to B$ and by the second assumption $M, w' \Vdash A$. It follows that $M, w' \Vdash B$. Since w' was arbitrary, $M, w \Vdash \Box B$. □

Proposition 8.20. *The following schema* DUAL *is valid*

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

Proof. Exercise. □

Proposition 8.21. *If A and $A \to B$ are true at a world in a model then so is B. Hence, the valid formulas are closed under modus ponens.*

Proposition 8.22. *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.)

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

Table 8.1: Valid and (or?) invalid schemas.

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

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

8.10 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 8.23. If Γ is a set of formulas and A a formula, then Γ *entails* A, in symbols: $\Gamma \vDash A$, if and only if for every model

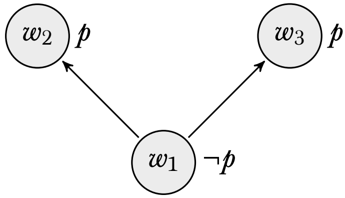


Figure 8.2: Counterexample to $p \to \Diamond p \vDash \Box p \to p$.

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

Example 8.24. 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 $p \to \Diamond p \vDash \Box\neg p \to \neg p$, we might argue as follows: Consider a model $M = \langle W, R, V\rangle$ and $w \in W$, and suppose $M, w \Vdash p \to \Diamond p$. We have to show that $M, w \Vdash \Box\neg p \to \neg p$. Suppose not. Then $M, w \Vdash \Box\neg p$ and $M, w \nVdash \neg p$. Since $M, w \nVdash \neg p$, $M, w \Vdash p$. By assumption, $M, w \Vdash p \to \Diamond p$, hence $M, w \Vdash \Diamond p$. By definition of $M, w \Vdash \Diamond p$, there is some w' with Rww' such that $M, w' \Vdash p$. Since also $M, w \Vdash \Box\neg p$, $M, w' \Vdash \neg p$, a contradiction.

To show that a formula B does not entail another A, we have to give a counterexample, i.e., a model $M = \langle W, R, V\rangle$ where we show that at some world $w \in W$, $M, w \Vdash B$ but $M, w \nVdash A$. Let's show that $p \to \Diamond p \nvDash \Box p \to p$. Consider the model in Figure 8.2. We have $M, w_1 \Vdash \Diamond p$ and hence $M, w_1 \Vdash p \to \Diamond p$. However, since $M, w_1 \Vdash \Box p$ but $M, w_1 \nVdash p$, we have $M, w_1 \nVdash \Box p \to p$.

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

Problems

Problem 8.1. Consider the model of Figure 8.1. Which of the following hold?

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

Problem 8.2. Complete the proof of Proposition 8.8.

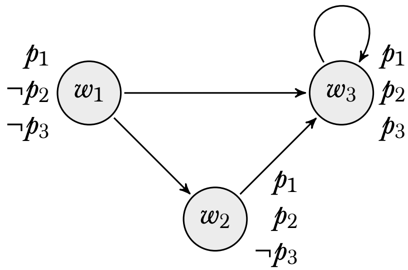
Problem 8.3. Let $M = \langle W, R, V \rangle$ be a model, and suppose $w_1, w_2 \in W$ are such that:

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 8.4. 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 8.5. Consider the following model M for the language comprising p_1, p_2, p_3 as the only propositional variables:



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

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

Problem 8.6. Show that the following are valid:

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

Problem 8.7. Show that $A \to \Box A$ is valid in the class $\mathscr{C}$ of models $M = \langle W, R, V \rangle$ where $W = \{w\}$. Similarly, show that $B \to \Box A$ and $\Diamond A \to B$ are valid in the class of models $M = \langle W, R, V \rangle$ where $R = \emptyset$.

Problem 8.8. Prove Proposition 8.20.

Problem 8.9. Prove the claim in the "only if" part of the proof of Proposition 8.22. (Hint: use induction on A.)

Problem 8.10. Show that none of the following formulas are valid:

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

T: $\Box p \to p$;

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

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

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

Problem 8.11. Prove that the schemas in the first column of Table 8.1 are valid and those in the second column are not valid.

Problem 8.12. Decide whether the following schemas are valid or invalid:

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

Problem 8.13. For each of the following schemas find a model M such that every instance of the formula is true in M:

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

Problem 8.14. Show that $\Box(A \land B) \vDash \Box A$.

Problem 8.15. Show that $\Box(p \to q) \nvDash p \to \Box q$ and $p \to \Box q \nvDash \Box(p \to q)$.

CHAPTER 9

Axiomatic Derivations

9.1 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 $\mathbf{K}$.

Definition 9.1. The rule of *modus ponens* is the inference schema

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

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

Definition 9.2. The rule of *necessitation* is the inference schema

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

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

Definition 9.3. A *derivation* from a set of axioms Σ is a sequence of formulas $B_1, B_2, \ldots, B_n$, where each B_i is either

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

ponens, or

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

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

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

9.2 Proofs in K

In order to practice proofs in the smallest modal system, we show the valid formulas on the left-hand side of Table 8.1 can all be given K-proofs.

Proposition 9.4. $K \vdash \Box A \to \Box(B \to A)$

Proof.

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

□

Proposition 9.5. $K \vdash \Box(A \land B) \to (\Box A \land \Box B)$

Proof.

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

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

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

□

Proposition 9.6. $K \vdash (\Box A \wedge \Box B) \to \Box(A \wedge B)$

Proof.

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

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

$$(p \to q) \to ((q \to r) \to (p \to r))$$
$$(p \to (q \to r)) \to ((p \land q) \to r) \qquad \square$$

Proposition 9.7. $K \vdash \neg\Box p \to \Diamond\neg p$

Proof.

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

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

$$(p \to q) \to (\neg q \to \neg p)$$
$$(p \to q) \to ((q \to r) \to (p \to r)). \qquad \square$$

9.3 Derived Rules

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

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

and applying MP twice. Often we want to replace a sub-formula by a formula we know to be equivalent, e.g., $\Diamond A$ by $\neg\Box\neg A$, or $\neg\neg A$ by A. So rather than write out the actual derivation, it is more convenient to simply record why the intermediate steps are derivable. For this purpose, let us collect some facts about derivability.

Proposition 9.8. *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

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

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

We will indicate use of this proposition by PL.

Proposition 9.9. *If* $K \vdash A_1 \to (A_2 \to \cdots (A_{n-1} \to A_n) \dots)$ *then* $K \vdash \Box A_1 \to (\Box A_2 \to \cdots (\Box A_{n-1} \to \Box A_n) \dots)$.

Proof. By induction on n, just as in the proof of Proposition 12.3. □

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

Proposition 9.10. $K \vdash (\Box A \land \Box B) \to \Box(A \land B)$

Proof.

1.	$K \vdash A \to (B \to (A \land B))$	TAUT
2.	$K \vdash \Box A \to (\Box B \to \Box(A \land B)))$	RK, 1
3.	$K \vdash (\Box A \land \Box B) \to \Box(A \land B)$	PL, 2

□

Proposition 9.11. *If* $K \vdash A \leftrightarrow B$ *and* $K \vdash C[A/q]$ *then* $K \vdash C[B/q]$

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 9.11 by "A for B" whenever we re-write a formula $C(B)$ for $C(A)$. In other words, "A for B" abbreviates:

$\vdash C(A)$	
$\vdash A \leftrightarrow B$	
$\vdash C(B)$	by Proposition 9.11

For instance:

Proposition 9.12. $K \vdash \neg\Box p \to \Diamond\neg p$

Proof.

1.	$K \vdash \Diamond\neg p \leftrightarrow \neg\Box\neg\neg p$	DUAL
2.	$K \vdash \neg\Box\neg\neg p \to \Diamond\neg p$	PL, 1
3.	$K \vdash \neg\Box p \to \Diamond\neg p$	p for $\neg\neg p$

□

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

$K \vdash \neg\Box\neg\neg p \to \Diamond\neg p$	
$K \vdash \neg\neg p \leftrightarrow p$	TAUT
$K \vdash \neg\Box p \to \Diamond\neg p$	by Proposition 9.11

The roles of $C(q)$, A, and B in Proposition 9.11 are played here, respectively, by $\neg\Box q \to \Diamond\neg p$, $\neg\neg p$, and p.

When a formula contains a sub-formula $\neg\Diamond A$, we can replace it by $\Box\neg A$ using Proposition 9.11, since $K \vdash \neg\Diamond A \leftrightarrow \Box\neg A$. 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 $K \vdash \neg\Box p \to \Diamond\neg p$, but $K \vdash \neg\Box A \to \Diamond\neg A$ for arbitrary A.

Proposition 9.13. *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. □

9.4 More Proofs in *K*

Let's see some more examples of derivability in *K*, now using the simplified method introduced in section 9.3.

Proposition 9.14. $K \vdash \Box(A \to B) \to (\Diamond A \to \Diamond B)$

Proof.

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

Proposition 9.15. $K \vdash \Box A \to (\Diamond(A \to B) \to \Diamond B)$

Proof.

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

Proposition 9.16. $K \vdash (\Diamond A \lor \Diamond B) \to \Diamond(A \lor B)$

Proof.

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

□

Proposition 9.17. $K \vdash \Diamond(A \lor B) \to (\Diamond A \lor \Diamond B)$

Proof.

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

□

Problems

Problem 9.1. Find derivations in K for the following formulas:

1. $\Box\neg p \to \Box(p \to q)$

2. $(\Box p \lor \Box q) \to \Box(p \lor q)$

3. $\Diamond p \to \Diamond(p \lor q)$

Problem 9.2. Prove Proposition 9.11 by proving, by induction on the complexity of C, that if $K \vdash A \leftrightarrow B$ then $K \vdash C[A/q] \leftrightarrow C[B/q]$.

Problem 9.3. Show that the following derivability claims hold:

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

CHAPTER 10

Modal Tableaux

10.1 Introduction

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

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

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

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

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

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

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

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

10.2 Rules for K

The rules for the regular propositional connectives are the same as for regular propositional signed tableaux, just with prefixes added. In each case, the rule applied to a signed formula $\sigma\,S\,A$ produces new formulas that are also prefixed by σ. This should be intuitively clear: e.g., if $A \land B$ is true at (a world named by) σ, then A and B are true at σ (and not at any other world). We collect the propositional rules in Table 10.1.

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

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

for some prefix σ and formula A.

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

$\dfrac{\sigma\,\mathbb{T}\,\neg A}{\sigma\,\mathbb{F}\,A}\;\neg\mathbb{T}$	$\dfrac{\sigma\,\mathbb{F}\,\neg A}{\sigma\,\mathbb{T}\,A}\;\neg\mathbb{F}$
$\dfrac{\sigma\,\mathbb{T}\,A \land B}{\begin{array}{c}\sigma\,\mathbb{T}\,A\\ \sigma\,\mathbb{T}\,B\end{array}}\;\land\mathbb{T}$	$\dfrac{\sigma\,\mathbb{F}\,A \land B}{\sigma\,\mathbb{F}\,A \quad \mid \quad \sigma\,\mathbb{F}\,B}\;\land\mathbb{F}$
$\dfrac{\sigma\,\mathbb{T}\,A \lor B}{\sigma\,\mathbb{T}\,A \quad \mid \quad \sigma\,\mathbb{T}\,B}\;\lor\mathbb{T}$	$\dfrac{\sigma\,\mathbb{F}\,A \lor B}{\begin{array}{c}\sigma\,\mathbb{F}\,A\\ \sigma\,\mathbb{F}\,B\end{array}}\;\lor\mathbb{F}$
$\dfrac{\sigma\,\mathbb{T}\,A \to B}{\sigma\,\mathbb{F}\,A \quad \mid \quad \sigma\,\mathbb{T}\,B}\;\to\mathbb{T}$	$\dfrac{\sigma\,\mathbb{F}\,A \to B}{\begin{array}{c}\sigma\,\mathbb{T}\,A\\ \sigma\,\mathbb{F}\,B\end{array}}\;\to\mathbb{F}$

Table 10.1: Prefixed tableau rules for the propositional connectives

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

$$B_1, \dots, B_n \vdash A$$

iff there is a closed tableau for the assumptions

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

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

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

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

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

Table 10.2: The modal rules for K.

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

The rules are given in Table 10.2.

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

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

But $\Box A \nvDash \Diamond A$, so our proof system would be unsound. Likewise, $\Diamond A \nvDash \Box A$, but without the restriction that the prefix for $\Box\mathbb{F}$ must be new, this would be a closed tableau:

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

10.3 Tableaux for K

Example 10.1. We give a closed tableau that shows $\vdash (\Box A \wedge \Box B) \rightarrow \Box(A \wedge B)$.

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

Example 10.2. We give a closed tableau that shows $\vdash \Diamond(A \vee B) \rightarrow (\Diamond A \vee \Diamond B)$:

$$
\begin{array}{llll}
1. & & 1\,\mathbb{F}\;\Diamond(A \lor B) \to (\Diamond A \lor \Diamond B) & \text{Assumption} \\
2. & & 1\,\mathbb{T}\;\Diamond(A \lor B) & {\to}\mathbb{F}\,1 \\
3. & & 1\,\mathbb{F}\;\Diamond A \lor \Diamond B & {\to}\mathbb{F}\,1 \\
4. & & 1\,\mathbb{F}\;\Diamond A & {\lor}\mathbb{F}\,3 \\
5. & & 1\,\mathbb{F}\;\Diamond B & {\lor}\mathbb{F}\,3 \\
6. & & 1.1\,\mathbb{T}\;A \lor B & \Diamond\mathbb{T}\,2 \\
7. & 1.1\,\mathbb{T}\;A & 1.1\,\mathbb{T}\;B & {\lor}\mathbb{T}\,6 \\
8. & 1.1\,\mathbb{F}\;A & 1.1\,\mathbb{F}\;B & \Diamond\mathbb{F}\,4;\ \Diamond\mathbb{F}\,5 \\
 & \otimes & \otimes &
\end{array}
$$

10.4 Soundness for K

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

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

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

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

Definition 10.3. Let P be some set of prefixes, i.e., $P \subseteq (\mathbb{Z}^+)^* \setminus \{\Lambda\}$ and let M be a model. A function $f\colon P \to W$ is an *interpretation of P in M* if, whenever σ and $\sigma.n$ are both in P, then

$Rf(\sigma)f(\sigma.n)$.

Relative to an interpretation of prefixes P we can define:

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

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

Definition 10.4. Let Γ be a set of prefixed formulas, and let $P(\Gamma)$ be the set of prefixes that occur in it. If f is an interpretation of $P(\Gamma)$ in M, we say that M satisfies Γ with respect to f, $M, f \Vdash \Gamma$, if M satisfies every prefixed formula in Γ with respect to f. Γ is *satisfiable* iff there is a model M and interpretation f of $P(\Gamma)$ such that $M, f \Vdash \Gamma$.

Proposition 10.5. *If Γ contains both $\sigma\,\mathbb{T}\,A$ and $\sigma\,\mathbb{F}\,A$, for some formula A and prefix σ, then Γ is unsatisfiable.*

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

Theorem 10.6 (Soundness). *If Γ has a closed tableau, Γ is unsatisfiable.*

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

We show the following: Extending a satisfiable tableau by one of the rules of inference always results in a satisfiable tableau. This will prove the theorem: any closed tableau results by applying rules of inference to the tableau consisting only of assumptions from Γ. So if Γ were satisfiable, any tableau for it would be satisfiable. A closed tableau, however, is clearly not satisfiable, since all its branches are closed and closed branches are unsatisfiable.

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

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

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

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

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

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

5. The branch is expanded by applying $\rightarrow\mathbb{F}$ to $\sigma\, \mathbb{F}\, B \rightarrow C \in \Gamma$: This results in two new signed formulas on the

branch: $\sigma\,\mathbb{T}\,B$ and $\sigma\,\mathbb{F}\,C$. Suppose $M, f \Vdash \Gamma$, in particular $M, f(\sigma) \nVdash B \to C$. Then $M, f(\sigma) \Vdash B$ and $M, f(\sigma) \nVdash C$. This means that M, f satisfies both $\sigma\,\mathbb{T}\,B$ and $\sigma\,\mathbb{F}\,C$.

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

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

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

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

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

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

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

Corollary 10.7. *If $\Gamma \vdash A$ then $\Gamma \vDash A$.*

Proof. If $\Gamma \vdash A$ then for some $B_1, \ldots, B_n \in \Gamma$, $\Delta = \{1\,\mathbb{F}\,A, 1\,\mathbb{T}\,B_1, \ldots, 1\,\mathbb{T}\,B_n\}$ has a closed tableau. We want to show that $\Gamma \vDash A$. Suppose not, so for some M and w, $M, w \Vdash B_i$ for $i = 1, \ldots, n$, but $M, w \nVdash A$. Let $f(1) = w$; then f is an interpretation of $P(\Delta)$ into M, and M satisfies Δ with respect to f. But by Theorem 10.6, Δ is unsatisfiable since it has a closed tableau, a contradiction. So we must have $\Gamma \vdash A$ after all. □

Corollary 10.8. *If $\vdash A$ then A is true in all models.*

10.5 Rules for Other Accessibility Relations

In order to deal with logics determined by special accessibility relations, we consider the additional rules in Table 10.3.

Adding these rules results in systems that are sound and complete for the logics given in Table 10.4.

Example 10.9. We give a closed tableau that shows $S5 \vdash 5$, i.e., $\Box A \to \Box\Diamond A$.

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

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

Table 10.3: More modal rules.

10.6 Soundness for Additional Rules

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

Proposition 10.10. T$\Box$ *and* T$\Diamond$ *are sound for reflexive models.*

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

Logic	R is ...	Rules
$\mathbf{T} = \mathbf{KT}$	reflexive	T□, T◇
$\mathbf{D} = \mathbf{KD}$	serial	D□, D◇
$\mathbf{K4}$	transitive	4□, 4◇
$\mathbf{B} = \mathbf{KTB}$	reflexive, symmetric	T□, T◇ B□, B◇
$\mathbf{S4} = \mathbf{KT4}$	reflexive, transitive	T□, T◇, 4□, 4◇
$\mathbf{S5} = \mathbf{KT4B}$	reflexive, transitive, euclidean	T□, T◇, 4□, 4◇, 4r□, 4r◇

Table 10.4: Tableau rules for various modal logics.

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

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

Proposition 10.11. D□ *and* D◇ *are sound for serial models.*

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

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

Proposition 10.12. $\mathrm{B}\Box$ *and* $\mathrm{B}\Diamond$ *are sound for symmetric models.*

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

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

Proposition 10.13. $4\Box$ *and* $4\Diamond$ *are sound for transitive models.*

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

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

Proposition 10.14. $4\mathrm{r}\Box$ *and* $4\mathrm{r}\Diamond$ *are sound for euclidean models.*

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

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

Corollary 10.15. *The tableau systems given in Table 10.4 are sound for the respective classes of models.*

10.7 Simple Tableaux for *S5*

S5 is sound and complete with respect to the class of universal models, i.e., models where every world is accessible from every world. In universal models the accessibility relation doesn't matter: "there is a world w where $M, w \Vdash A$" is true if and only if there is such a w that's accessible from u. So in *S5*, we can define models as simply a set of worlds and a valuation V. This suggests that we should be able to simplify the tableau rules as well. In the general case, we take as prefixes sequences of positive integers, so that we can keep track of which such prefixes name worlds which are accessible from others: $\sigma.n$ names a world accessible from σ. But in *S5* any world is accessible from any world, so there is no need to so keep track. Instead, we can use positive integers as prefixes. The simplified rules are given in Table 10.5.

Example 10.16. We give a simplified closed tableau that shows $S5 \vdash 5$, i.e., $\Diamond A \to \Box\Diamond A$.

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

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

Table 10.5: Simplified rules for *S5*.

10.8 Completeness for *K*

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

Definition 10.17. A branch in a tableau is called complete if, whenever it contains a prefixed formula $\sigma\ S\ A$ to which a rule can be applied, it also contains

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

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

Proposition 10.18. *Every finite Γ has a tableau in which every branch is complete.*

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

Theorem 10.19 (Completeness). *If Γ has no closed tableau, Γ is satisfiable.*

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Corollary 10.20. *If $\Gamma \vDash A$ then $\Gamma \vdash A$.*

Corollary 10.21. *If A is true in all models, then $\vdash A$.*

10.9 Countermodels from Tableaux

The proof of the completeness theorem doesn't just show that if $\vDash A$ then $\vdash A$, it also gives us a method for constructing countermodels to A if $\nvDash A$. In the case of $\mathbf{K}$, this method constitutes a *decision procedure*. For suppose $\nvDash A$. Then the proof of Proposition 10.18 gives a method for constructing a complete tableau. The method in fact always terminates. The propositional rules for $\mathbf{K}$ only add prefixed formulas of lower complexity, i.e., each propositional rule need only be applied once on a branch for any signed formula $\sigma\,S\,A$. New prefixes are only generated by the $\Box\mathbb{F}$

and $\Diamond\mathbb{T}$ rules, and also only have to be applied once (and produce a single new prefix). $\Box\mathbb{T}$ and $\Diamond\mathbb{F}$ have to be applied potentially multiple times, but only once per prefix, and only finitely many new prefixes are generated. So the construction either results in a closed branch or a complete branch after finitely many stages.

Once a tableau with an open complete branch is constructed, the proof of Theorem 10.19 gives us an explict model that satisfies the original set of prefixed formulas. So not only is it the case that if $\Gamma \vDash A$, then a closed tableau exists and $\Gamma \vdash A$, if we look for the closed tableau in the right way and end up with a "complete" tableau, we'll not only know that $\Gamma \nvDash A$ but actually be able to construct a countermodel.

Example 10.22. We know that $\nvDash \Box(p \lor q) \to (\Box p \lor \Box q)$. The construction of a tableau begins with:

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

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

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

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

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

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

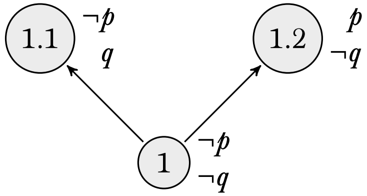


Figure 10.1: A countermodel to $\Box(p \lor q) \to (\Box p \lor \Box q)$.

tains $1.1\,\mathbb{T}\,q$). The model is pictured in Figure 10.1, and you can verify that it is a countermodel to $\Box(p \lor q) \to (\Box p \lor \Box q)$.

Problems

Problem 10.1. Find closed tableaux in K for the following formulas:

1. $\Box\neg p \to \Box(p \to q)$
2. $(\Box p \lor \Box q) \to \Box(p \lor q)$
3. $\Diamond p \to \Diamond(p \lor q)$
4. $\Box(p \land q) \to \Box p$

Problem 10.2. Complete the proof of Theorem 10.6.

Problem 10.3. Give closed tableaux that show the following:

1. $\mathsf{KT5} \vdash \mathrm{B}$;
2. $\mathsf{KT5} \vdash 4$;
3. $\mathsf{KDB4} \vdash \mathrm{T}$;
4. $\mathsf{KB4} \vdash 5$;
5. $\mathsf{KB5} \vdash 4$;

6. $\mathbf{KT} \vdash \mathsf{D}$.

Problem 10.4. Complete the proof of Proposition 10.10

Problem 10.5. Complete the proof of Proposition 10.11

Problem 10.6. Complete the proof of Proposition 10.12

Problem 10.7. Complete the proof of Proposition 10.13

Problem 10.8. Complete the proof of Proposition 10.14

Problem 10.9. Complete the proof of Theorem 10.19.

PART IV

Is this really necessary?

CHAPTER 11

Frame Definability

11.1 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, $\Box A \to A$ will be true there (in other words, the schema $\Box p \to p$ and all its substitution instances are true in M).

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

The solution is to remove the variable assignment V from the equation. If we require that $\Box p \to p$ 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 $\Box p \to p$ is false at w.

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

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

11.2 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 11.1. *Let $M = \langle W, R, V\rangle$ be a model. If R has the property on the left side of Table 11.1, 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 $A \to \Box\Diamond A$ be a given instance of B,

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

Table 11.1: Five correspondence facts.

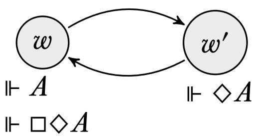


Figure 11.1: The argument from symmetry.

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 11.1). 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 11.1 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 $M, w \nVdash p$. But here we have just a single instance of T that fails in M; other instances, e.g., $\Box\neg p \to \neg p$ 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 11.2. *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: $M, u \Vdash A$ if and only if $M, v \Vdash A$ (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 11.1 fails in the case of* T *(similar arguments can be given for some—though not all—the other schemas mentioned in Theorem 11.1).*

Although we will focus on the five classical formulas D, T, B, 4, and 5, we record in Table 11.2 a few more properties of accessibility relations. The accessibility relation R is partially functional, if from every world at most one world is accessible. If it is the case that from every world exactly one world is accessible, we call it functional. (Thus the functional relations are precisely those that are both serial and partially functional). They are called "functional" because the accessibility relation operates like a (partial) function. A relation is weakly dense if whenever Ruv, there is a w "between" u and v. So weakly dense relations are in a sense the opposite of transitive relations: in a transitive relation, whenever you can reach v from u by a detour via w, you can reach v from u directly; in a weakly dense relation, whenever you can reach v from u directly, you can also reach it by a detour via some w. A relation is weakly directed if whenever you can reach worlds u and v from some world w, you can reach a single world t from both u and v—this is sometimes called the "diamond property" or "confluence."

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

Table 11.2: Five more correspondence facts.

11.3 Frames

Definition 11.3. A *frame* is a pair $\mathbf{F} = \langle W, R\rangle$ where W is a nonempty set of worlds and R a binary relation on W. A model $\mathbf{M}$ is *based on* a frame $\mathbf{F} = \langle W, R\rangle$ if and only if $\mathbf{M} = \langle W, R, V\rangle$ for some valuation V.

Definition 11.4. If $\mathbf{F}$ is a frame, we say that A is *valid in* $\mathbf{F}$, $\mathbf{F} \vDash A$, if $\mathbf{M} \Vdash A$ for every model $\mathbf{M}$ based on $\mathbf{F}$.

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

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: $\mathcal{F} \vDash A$ iff $\mathcal{C} \vDash A$ 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.

11.4 Frame Definability

Even though the converse implications of Theorem 11.1 fail, they hold if we replace "model" by "frame": for the properties considered in Theorem 11.1, 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 11.5. If $\mathcal{F}$ is a class of frames, we say A *defines* $\mathcal{F}$ iff $F \vDash A$ for all and only frames $F \in \mathcal{F}$.

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

Theorem 11.6. *If the formula on the right side of Table 11.1 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., $F \vDash \Box p \to \Diamond p$. 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 $M \Vdash \Box A$ and $M, w \nVdash \Diamond A$ for any A, including p. But then $M, w \nVdash \Box p \to \Diamond p$, contradicting the assumption that $F \vDash \Box p \to \Diamond p$.

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

u such that Rwu: $M, u \Vdash p$, and hence $M, w \Vdash \Box p$. But by hypothesis $\Box p \to p$ 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., $p \to \Box\Diamond p$ 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 $M, u \nVdash p \to \Box\Diamond p$, and so B is not valid in F.

4. Suppose 4 is valid in $F = \langle W, R\rangle$, i.e., $F \vDash \Box p \to \Box\Box p$, 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, $\Box p \to \Box\Box p$, 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., $F \nvDash \Diamond p \to \Box\Diamond p$. Suppose there are worlds u, v, $w \in W$ such that Rwu and Rwv but not Ruv. Define V such that for all worlds z, $z \in V(p)$ if and only if it is *not* the case that Ruz. Let $M = \langle W, R, V\rangle$. Then by hypothesis $M, v \Vdash p$ and since Rwv 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, $\Diamond p \to \Box\Diamond p$, fails at w. □

You'll notice a difference between the proof for D and the other cases: no mention was made of the valuation V. In effect, we proved that if $\mathfrak{M} \Vdash \mathsf{D}$ then $\mathfrak{M}$ is serial. So D defines the class of serial *models*, not just frames.

Corollary 11.7. *Any model where* D *is true is serial.*

Corollary 11.8. *Each formula on the right side of Table 11.1 defines the class of frames which have the property on the left side.*

Proof. In Theorem 11.1, 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 $\mathfrak{F}$ has the property on the left, the formula on the right is valid in $\mathfrak{F}$. In Theorem 11.6, we proved the converse implications: if a formula on the right is valid in $\mathfrak{F}$, $\mathfrak{F}$ has the property on the left. □

Theorem 11.6 also shows that the properties can be combined: for instance if both B and 4 are valid in $\mathfrak{F}$ 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 $\mathfrak{M}, w \Vdash \mathsf{T}$ then $\mathfrak{M}, w \Vdash \mathsf{D}$.) We record some such relationships.

Proposition 11.9. *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.*

11.5 First-order Definability

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

Definition 11.10. 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 $\boldsymbol{F} = \langle W, R\rangle \in \mathcal{F}$ iff $\mathbf{M} \vDash A$ in the first-order structure $\mathbf{M}$ with $|\mathbf{M}| = W$ and $Q^{\mathbf{M}} = R$.

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

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

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

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

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

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

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

Now consider the set of formulas

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

Every finite subset of Γ is satisfiable: Let k be largest such that A_k is in the subset, $|\mathbf{M}_k| = \{1, \dots, k\}$, $a_i^{\mathbf{M}_k} = i$, and $Q^{\mathbf{M}_k} =<$. Since $<$ on $\{1, \dots, k\}$ is transitive and converse well-founded, $\mathbf{M}_k \vDash F$. $\mathbf{M}_k \vDash A_i$ by construction, for all $i \leq k$. By the Compactness Theorem for first-order logic, Γ is satisfiable in some structure $\mathbf{M}$. By hypothesis, since $\mathbf{M} \vDash F$, the relation $Q^{\mathbf{M}}$ is converse well-founded. But clearly, $a_1^{\mathbf{M}}$, $a_2^{\mathbf{M}}$, ... 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.

11.6 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 11.11. 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 11.12. *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. $\Box$

Proposition 11.12 is the semantic counterpart to Proposition 12.13, 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 11.13. *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 11.14. *A formula A is valid in all frames $\mathbf{F} = \langle W, R\rangle$ where R is an equivalence relation, if and only if it is valid in all frames $\mathbf{F} = \langle W, R\rangle$ where R is universal. Hence, the logic of universal frames is just* **S5**.

Proof. It's immediate to verify that a universal relation R on W is an equivalence. Hence, if A is valid in all frames where R is an equivalence it is valid in all universal frames. For the other direction, we argue contrapositively: suppose B is a formula that fails at a world w in a model $\mathbf{M} = \langle W, R, V\rangle$ based on a frame $\langle W, R\rangle$, where R is an equivalence on W. So $\mathbf{M}, w \nVdash B$. Define a model $\mathbf{M}' = \langle W', R', V'\rangle$ 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 $\mathbf{M}'$ is represented by the shaded area in Figure 11.2.) It is easy to see that R and R' agree on W'. Then one can show by induction on formulas that for all $w' \in W'$: $\mathbf{M}', w' \Vdash A$ if and only if $\mathbf{M}, w' \Vdash A$ for each A (this

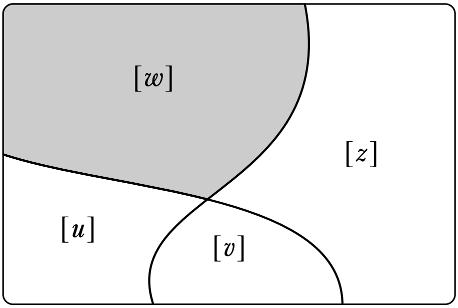


Figure 11.2: A partition of W in equivalence classes.

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

11.7 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 11.15. The *standard translation* $\mathrm{ST}_x(A)$ is inductively defined as follows:

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

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

3. $A \equiv \neg B$: $\mathrm{ST}_x(A) = \neg\mathrm{ST}_x(B)$.

4. $A \equiv (B \land C)$: $\mathrm{ST}_x(A) = (\mathrm{ST}_x(B) \land \mathrm{ST}_x(C))$.

5. $A \equiv (B \lor C)$: $\mathrm{ST}_x(A) = (\mathrm{ST}_x(B) \lor \mathrm{ST}_x(C))$.

6. $A \equiv (B \to C)$: $\mathrm{ST}_x(A) = (\mathrm{ST}_x(B) \to \mathrm{ST}_x(C))$.

7. $A \equiv \Box B$: $\mathrm{ST}_x(A) = \forall y\, (Q(x,y) \to \mathrm{ST}_y(B))$.

8. $A \equiv \Diamond B$: $\mathrm{ST}_x(A) = \exists y\, (Q(x,y) \land \mathrm{ST}_y(B))$.

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

Proposition 11.16. *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*

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

Proof. By induction on A. □

Proposition 11.17. *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*

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

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

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

Proof. $F' \vDash A'$ iff for every structure M' where $P_i^{M'} \subseteq W$ for $i = 1, \dots, n$, and for every s with $s(x) \in W$, $M', s \vDash \mathrm{ST}_x(A)$. By Proposition 11.16, that is the case iff for all models M based on F and every world $w \in W$, $M, w \Vdash A$, i.e., $F \vDash A$. □

Definition 11.18. 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 $M \vDash A$ in the structure M with $|M| = W$ and $P^M = R$.

Corollary 11.19. *If a class of frames is definable by a formula A, the corresponding class of accessibility relations is definable by a monadic second-order sentence.*

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

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

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

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

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

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

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

Problems

Problem 11.1. Complete the proof of Theorem 11.1.

Problem 11.2. Prove the claims in Proposition 11.2.

Problem 11.3. Let $\mathbf{M} = \langle W, R, V \rangle$ be a model. Show that if R satisfies the left-hand properties of Table 11.2, every instance of the corresponding right-hand formula is true in $\mathbf{M}$.

Problem 11.4. Show that if the formula on the right side of Table 11.2 is valid in a frame $\mathbf{F}$, then $\mathbf{F}$ 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 11.5. Prove Proposition 11.9.

Problem 11.6. Prove Proposition 11.12 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 12

More Axiomatic Derivations

12.1 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 12.1. A *modal logic* is a set Σ of modal formulas which

1. contains all tautologies, and

2. is closed under substitution, i.e., if $A \in \Sigma$, and $D_1, \dots, D_n$ 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 $A \to B \in \Sigma$, then $B \in \Sigma$.

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

Definition 12.2. A modal logic Σ is *normal* if it contains

$$\Box(p \to q) \to (\Box p \to \Box q), \tag{K}$$

$$\Diamond p \leftrightarrow \neg\Box\neg p \tag{DUAL}$$

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

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

Proposition 12.3. *Every normal modal logic is closed under rule* RK,

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

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

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

Assume

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

By the induction hypothesis, we have

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

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

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

Using modus ponens and suitable tautological instances we get

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

Proposition 12.4. *Every normal modal logic Σ contains $\neg\Diamond\bot$.*

Proposition 12.5. *Let $A_1, \ldots, A_n$ be formulas. Then there is a smallest modal logic Σ containing all instances of $A_1, \ldots, A_n$.*

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

Definition 12.6. The smallest normal modal logic containing $A_1, \ldots, A_n$ is called a *modal system* and denoted by $\boldsymbol{K}A_1 \ldots A_n$. The smallest normal modal logic is denoted by $\boldsymbol{K}$.

12.2 Derivations and Modal Systems

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

Definition 12.7. Given a modal system $KA_1 \dots A_n$ and a formula B we say that B is *derivable* in $KA_1 \dots A_n$, written $KA_1 \dots A_n \vdash B$, 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 Σ-derivation of B.

Proposition 12.8. $KA_1 \dots A_n = \{B : KA_1 \dots A_n \vdash B\}$.

Proof. We use induction on the length of derivations to show that $\{B : KA_1 \dots A_n \vdash B\} \subseteq KA_1 \dots A_n$.

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 $KA_1 \dots A_n$ contains these as well, so $B \in KA_1 \dots A_n$.

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 $C \rightarrow B \in KA_1 \dots A_n$ by induction hypothesis. But every modal logic is closed under modus ponens, so $B \in KA_1 \dots A_n$. If $B \equiv \Box C$ follows from C by NEC, then $C \in KA_1 \dots A_n$ by induction hypothesis. But every normal modal logic is closed under NEC, so $B \in KA_1 \dots A_n$.

The converse inclusion follows by showing that $\Sigma = \{B : KA_1 \dots A_n \vdash B\}$ is a normal modal logic containing all the instances of $A_1, \dots, A_n$, and the observation that $KA_1 \dots A_n$ is, by definition, the smallest such logic.

1. Every tautology B is a tautological instance, so $KA_1 \dots A_n \vdash B$, so Σ contains all tautologies.

2. If $KA_1 \dots A_n \vdash C$ and $KA_1 \dots A_n \vdash C \rightarrow B$, then $KA_1 \dots A_n \vdash B$: Combine the derivation of C with that of $C \rightarrow B$, and

add the line B. The last line is justified by MP. So Σ is closed under modus ponens.

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

4. We have $KA_1 \ldots A_n \vdash \mathrm{K}$, so $K \in \Sigma$.

5. We have $KA_1 \ldots A_n \vdash$ DUAL, so DUAL $\in \Sigma$.

6. If $KA_1 \ldots A_n \vdash C$, the additional line $\Box C$ is justified by NEC. Consequently, Σ is closed under NEC. Thus, Σ is normal. □

12.3 Dual Formulas

Definition 12.9. Each of the formulas T, B, 4, and 5 has a *dual*, denoted by a subscripted diamond, as follows:

$$p \to \Diamond p \tag{$\mathrm{T}_\Diamond$}$$

$$\Diamond\Box p \to p \tag{$\mathrm{B}_\Diamond$}$$

$$\Diamond\Diamond p \to \Diamond p \tag{$4_\Diamond$}$$

$$\Diamond\Box p \to \Box p \tag{$5_\Diamond$}$$

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

Proposition 12.10. *For each formula A in Definition 12.9: $\mathbf{K}A = \mathbf{K}A_\Diamond$.*

Proof. Exercise. □

12.4 Proofs in Modal Systems

We now come to proofs in systems of modal logic other than $\mathbf{K}$.

Proposition 12.11. *The following provability results obtain:*

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

Proof. We exhibit proofs for each.

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

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

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

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

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

1.	***KDB4*** $\vdash \Diamond\Box A \to A$	$\mathrm{B}_\Diamond$
2.	***KDB4*** $\vdash \Box\Box A \to \Diamond\Box A$	D with $\Box A$ for p
3.	***KDB4*** $\vdash \Box\Box A \to A$	PL1, 2
4.	***KDB4*** $\vdash \Box A \to \Box\Box A$	4
5.	***KDB4*** $\vdash \Box A \to A$	PL, 1, 4.

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

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

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

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

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

1.	***KT*** $\vdash \Box A \to A$	T
2.	***KT*** $\vdash A \to \Diamond A$	$\mathrm{T}_\Diamond$
3.	***KT*** $\vdash \Box A \to \Diamond A$	PL, 1, 2

$\Box$

Definition 12.12. 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 11.12).

Proposition 12.13. *KTB4 = KT5 = KDB4 = KDB5.*

Proof. Exercise. □

12.5 Soundness

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

Theorem 12.14 (Soundness Theorem). *If every instance of A_1, ..., A_n is valid in the classes of models $\mathscr{C}_1$, ..., $\mathscr{C}_n$, respectively, then $\mathbf{K}A_1 \dots A_n \vdash B$ implies that B is valid in the class of models $\mathscr{C}_1 \cap \dots \cap \mathscr{C}_n$.*

12.6 Showing Systems are Distinct

In section 12.4 we saw how to prove that two systems of modal logic are in fact the same system. Theorem 12.14 allows us to show that two modal systems Σ and Σ' are distinct, by finding a formula A such that $\Sigma' \vdash A$ that fails in a model of Σ.

Proposition 12.15. *KD ⊊ KT*

Proof. This is the syntactic counterpart to the semantic fact that all reflexive relations are serial. To show **KD** ⊆ **KT** we need to see that **KD** $\vdash B$ implies **KT** $\vdash B$, which follows from **KT** $\vdash$ **D**, as shown in Proposition 12.11(6). To show that the inclusion is proper, by Soundness (Theorem 12.14), it suffices to exhibit a model of **KD** where T, i.e., $\Box p \rightarrow p$, fails (an easy task left as an exercise), for then by Soundness **KD** $\nvdash \Box p \rightarrow p$. □

Proposition 12.16. *$KB \neq K4$.*

Proof. We construct a symmetric model where some instance of 4 fails; since obviously the instance is derivable for $\mathbf{K4}$ but not in $\mathbf{KB}$, it will follow $\mathbf{K4} \nsubseteq \mathbf{KB}$. Consider the symmetric model M of Figure 12.1. Since the model is symmetric, K and B are true in M (by Proposition 8.19 and Theorem 11.1, respectively). However, $M, w_1 \nVdash \Box p \to \Box\Box p$. □

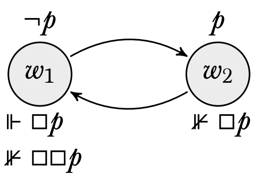


Figure 12.1: A symmetric model falsifying an instance of 4.

Theorem 12.17. *$KTB \nvdash 4$ and $KTB \nvdash 5$.*

Proof. By Theorem 11.1 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 12.2. Then $M, w_1 \nVdash \Box p \to \Box\Box p$, so 4 fails at w_1. Similarly, $M, w_2 \nVdash \Diamond\neg p \to \Box\Diamond\neg p$, so the instance of 5 with $A = \neg p$ fails at w_2. □

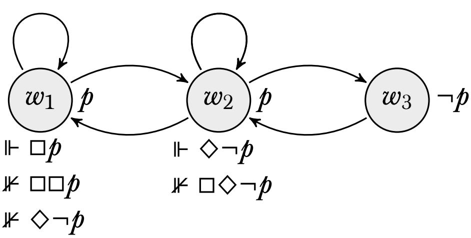


Figure 12.2: The model for Theorem 12.17.

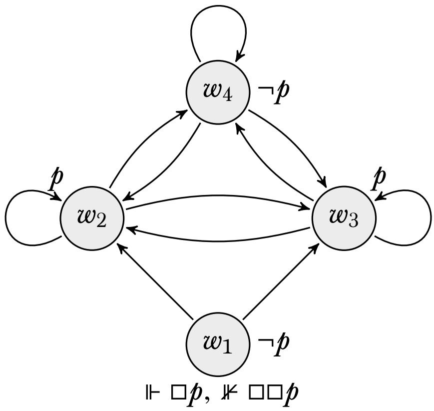


Figure 12.3: The model for Theorem 12.18.

Theorem 12.18. *KD5* $\neq$ *KT4* $=$ *S4*.

Proof. By Theorem 11.1 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 12.3, and notice that $M, w_1 \nVdash \Box p \to \Box\Box p$. □

12.7 Derivability from a Set of Formulas

In section 12.4 we defined a notion of provability of a formula in a system Σ. We now extend this notion to provability in Σ from formulas in a set Γ.

Definition 12.19. A formula A is derivable in a system Σ from a set of formulas Γ, written $\Gamma \vdash_\Sigma A$ if and only if there are B_1, …, $B_n \in \Gamma$ such that $\Sigma \vdash B_1 \to (B_2 \to \cdots (B_n \to A) \cdots)$.

12.8 Properties of Derivability

Proposition 12.20. *Let Σ be a modal system and Γ 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 $\Gamma \vdash_\Sigma B \to A$;*

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

The proof is an easy exercise. Part (5) of Proposition 12.20 gives us that, for instance, if $\Gamma \vdash_\Sigma A \lor B$ 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 12.21. A set Γ is *deductively closed* relatively to a system Σ if and only if $\Gamma \vdash_\Sigma A$ implies $A \in \Gamma$.

12.9 Consistency

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

Definition 12.22. A set Γ is *consistent* relatively to a system Σ or, as we will say, Σ-consistent, if and only if $\Gamma \nvdash_\Sigma \bot$.

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

Proposition 12.23. *Let Γ be a set of formulas. Then:*

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

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

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

Proof. These facts follow easily using classical propositional logic. We give the argument for (3). Proceed contrapositively and suppose neither $\Gamma \cup \{A\}$ nor $\Gamma \cup \{\neg A\}$ is Σ-consistent. Then by (2), both $\Gamma, A \vdash_{\Sigma} \bot$ and $\Gamma, \neg A \vdash_{\Sigma} \bot$. By the deduction theorem $\Gamma \vdash_{\Sigma} A \rightarrow \bot$ and $\Gamma \vdash_{\Sigma} \neg A \rightarrow \bot$. But $(A \rightarrow \bot) \rightarrow ((\neg A \rightarrow \bot) \rightarrow \bot)$ is a tautological instance, hence by Proposition 12.20(5), $\Gamma \vdash_{\Sigma} \bot$. □

Problems

Problem 12.1. Prove Proposition 12.4.

Problem 12.2. Prove Proposition 12.10.

Problem 12.3. Prove Proposition 12.13.

Problem 12.4. Give an alternative proof of Theorem 12.18 using a model with 3 worlds.

Problem 12.5. Provide a single reflexive transitive model showing that both ***KT4*** $\nvdash$ B and ***KT4*** $\nvdash$ 5.

CHAPTER 13

Completeness and Canonical Models

13.1 Introduction

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

Completeness is the converse of soundness: that $\mathbf{K}$ is complete means that if a formula A is valid, $\vdash A$, for instance. Proving completeness is a lot harder to do than proving soundness. It is useful, first, to consider the contrapositive: $\mathbf{K}$ is complete iff whenever $\nvdash A$, there is a countermodel, i.e., a model $\mathbf{M}$ such that $\mathbf{M} \nVdash A$. Equivalently (negating A), we could prove that whenever $\nvdash \neg A$, there is a model of A. In the construction of such a model, we can use information contained in A. When we find models for specific formulas we often do the same: e.g., if we want to find a countermodel to $p \to \Box q$, we know that it has to contain a world where p is true and $\Box q$ is false. And a world where $\Box q$ is false means there has to be a world accessible from it where q is false. And that's all we need to know: which worlds make the propositional variables true, and which worlds are accessible from which worlds.

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

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

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

13.2 Complete Σ-Consistent Sets

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

Definition 13.1. A set Γ is *complete Σ-consistent* if and only if it is Σ-consistent and for every A, either $A \in \Gamma$ or $\neg A \in \Gamma$.

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

Proposition 13.2. *Suppose Γ is complete Σ-consistent. Then:*

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

Proof.

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

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

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

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

5. Exercise.

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

7. Exercise.

13.3 Lindenbaum's Lemma

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

Theorem 13.3 (Lindenbaum's Lemma). *If Γ is Σ-consistent then there is a complete Σ-consistent set Δ extending Γ.*

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

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

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

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

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

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

So suppose Δ were Σ-inconsistent. Then $\Delta \vdash_\Sigma \bot$, i.e., there are $A_1, \dots, A_k \in \Delta$ such that $\Sigma \vdash A_1 \to (A_2 \to \cdots (A_k \to \bot) \dots)$. 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 Δ_n. This would mean $\Delta_n \vdash_\Sigma \bot$, i.e., Δ_n is Σ-inconsistent.

To show that each Δ_n is Σ-consistent, we use a simple induction on n. $\Delta_0 = \Gamma$, and we assumed Γ was Σ-consistent. So the claim holds for $n = 0$. Now suppose it holds for n, i.e., Δ_n is Σ-consistent. Δ_{n+1} is either $\Delta_n \cup \{A_n\}$ if that is Σ-consistent, otherwise it is $\Delta_n \cup \{\neg A_n\}$. In the first case, Δ_{n+1} is clearly Σ-consistent. However, by Proposition 12.23(3), either $\Delta_n \cup \{A_n\}$ or $\Delta_n \cup \{\neg A_n\}$ is consistent, so Δ_{n+1} is consistent in the other case as well. □

Corollary 13.4. *$\Gamma \vdash_\Sigma A$ if and only if $A \in \Delta$ for each complete Σ-consistent set Δ extending Γ (including when $\Gamma = \emptyset$, in which case we get another characterization of the modal system Σ.)*

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

13.4 Modalities and Complete Consistent Sets

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

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

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

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

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

Definition 13.5. If Γ is a set of formulas, let

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

and

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

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

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

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

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

Lemma 13.6. *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 $\Sigma \vdash B_1 \to (B_2 \to \cdots (B_n \to A) \cdots)$. Since Σ is normal, by rule RK, $\Sigma \vdash \Box B_1 \to (\Box B_2 \to \cdots (\Box B_n \to \Box A) \cdots)$, where obviously $\Box B_1, \dots, \Box B_k \in \Box\Gamma$. Hence, by definition, $\Box\Gamma \vdash_{\Sigma} \Box A$. □

Lemma 13.7. *If $\Box^{-1}\Gamma \vdash_{\Sigma} A$ then $\Gamma \vdash_{\Sigma} \Box A$.*

Proof. Suppose $\Box^{-1}\Gamma \vdash_{\Sigma} A$; then by Lemma 13.6, $\Box\Box^{-1}\Gamma \vdash \Box A$. But since $\Box\Box^{-1}\Gamma \subseteq \Gamma$, also $\Gamma \vdash_{\Sigma} \Box A$ by monotonicity. □

Proposition 13.8. *If Γ is complete Σ-consistent, then $\Box A \in \Gamma$ if and only if for every complete Σ-consistent Δ such that $\Box^{-1}\Gamma \subseteq \Delta$, it holds that $A \in \Delta$.*

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

For the "if" direction, we prove the contrapositive: Suppose $\Box A \notin \Gamma$. Since Γ is complete Σ-consistent, it is deductively closed, and hence $\Gamma \nvdash_\Sigma \Box A$. By Lemma 13.7, $\Box^{-1}\Gamma \nvdash_\Sigma A$. By Proposition 12.23(2), $\Box^{-1}\Gamma \cup \{\neg A\}$ is Σ-consistent. By Lindenbaum's Lemma, there is a complete Σ-consistent set Δ such that $\Box^{-1}\Gamma \cup \{\neg A\} \subseteq \Delta$. By consistency, $A \notin \Delta$. □

Lemma 13.9. *Suppose Γ and Δ are complete Σ-consistent. Then $\Box^{-1}\Gamma \subseteq \Delta$ if and only if $\Diamond\Delta \subseteq \Gamma$.*

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

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

Proposition 13.10. *If Γ is complete Σ-consistent, then $\Diamond A \in \Gamma$ if and only if for some complete Σ-consistent Δ such that $\Diamond\Delta \subseteq \Gamma$, it holds that $A \in \Delta$.*

Proof. Suppose Γ is complete Σ-consistent. $\Diamond A \in \Gamma$ iff $\neg\Box\neg A \in \Gamma$ by DUAL and closure. $\neg\Box\neg A \in \Gamma$ iff $\Box\neg A \notin \Gamma$ by Proposition 13.2(4) since Γ is complete Σ-consistent. By Proposition 13.8, $\Box\neg A \notin \Gamma$ iff, for some complete Σ-consistent Δ with

$\Box^{-1}\Gamma \subseteq \Delta$, $\neg A \notin \Delta$. Now consider any such Δ. By Lemma 13.9, $\Box^{-1}\Gamma \subseteq \Delta$ iff $\Diamond\Delta \subseteq \Gamma$. Also, $\neg A \notin \Delta$ iff $A \in \Delta$ by Proposition 13.2(4). So $\Diamond A \in \Gamma$ iff, for some complete Σ-consistent Δ with $\Diamond\Delta \subseteq \Gamma$, $A \in \Delta$. □

13.5 Canonical Models

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

Definition 13.11. Let Σ be a normal modal logic. The *canonical model* for Σ is $M^\Sigma = \langle W^\Sigma, R^\Sigma, V^\Sigma\rangle$, where:

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

13.6 The Truth Lemma

The canonical model M^Σ is defined in such a way that $M^\Sigma, \Delta \Vdash A$ iff $A \in \Delta$. For propositional variables, the definition of V^Σ yields this directly. We have to verify that the equivalence holds for all formulas, however. We do this by induction. The inductive step involves proving the equivalence for formulas involving propositional operators (where we have to use Proposition 13.2) and the modal operators (where we invoke the results of section 13.4).

Proposition 13.12 (Truth Lemma). *For every formula A, $M^\Sigma, \Delta \Vdash A$ if and only if $A \in \Delta$.*

Proof. By induction on A.

1. $A \equiv \bot$: $M^\Sigma, \Delta \nVdash \bot$ by Definition 8.7, and $\bot \notin \Delta$ by Proposition 13.2(3).

2. $A \equiv p$: $M^\Sigma, \Delta \Vdash p$ iff $\Delta \in V^\Sigma(p)$ by Definition 8.7. Also, $\Delta \in V^\Sigma(p)$ iff $p \in \Delta$ by definition of V^Σ.

3. $A \equiv \neg B$: $M^\Sigma, \Delta \Vdash \neg B$ iff $M^\Sigma, \Delta \nVdash B$ (Definition 8.7) iff $B \notin \Delta$ (by inductive hypothesis) iff $\neg B \in \Delta$ (by Proposition 13.2(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 8.7) iff $B \in \Delta$ or $C \in \Delta$ (by inductive hypothesis) iff $B \vee C \in \Delta$ (by Proposition 13.2(6)).

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

7. $A \equiv \Box B$: First suppose that $M^\Sigma, \Delta \Vdash \Box B$. By Definition 8.7, for every Δ' such that $R^\Sigma \Delta\Delta'$, $M^\Sigma, \Delta' \Vdash B$. By inductive hypothesis, for every Δ' such that $R^\Sigma \Delta\Delta'$, $B \in \Delta'$. By definition of R^Σ, for every Δ' such that $\Box^{-1}\Delta \subseteq \Delta'$, $B \in \Delta'$. By Proposition 13.8, $\Box B \in \Delta$.

   Now assume $\Box B \in \Delta$. Let $\Delta' \in W^\Sigma$ be such that $R^\Sigma \Delta\Delta'$, i.e., $\Box^{-1}\Delta \subseteq \Delta'$. Since $\Box B \in \Delta$, $B \in \Box^{-1}\Delta$. Consequently, $B \in \Delta'$. By inductive hypothesis, $M^\Sigma, \Delta' \Vdash B$. Since Δ' is arbitrary with $R^\Sigma \Delta\Delta'$, for all $\Delta' \in W^\Sigma$ such that $R^\Sigma \Delta\Delta'$, $M^\Sigma, \Delta' \Vdash B$. By Definition 8.7, $M^\Sigma, \Delta \Vdash \Box B$.

8. $A \equiv \Diamond B$: Exercise. □

13.7 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 Σ are exactly the Σ-derivable ones. Models with this property are said to *determine* Σ.

Definition 13.13. A model $\mathbf{M}$ *determines* a normal modal logic Σ precisely when $\mathbf{M} \Vdash A$ if and only if $\Sigma \vdash A$, for all formulas A.

Theorem 13.14 (Determination). *$\mathbf{M}^\Sigma \Vdash A$ if and only if $\Sigma \vdash A$.*

Proof. If $\mathbf{M}^\Sigma \Vdash A$, then for every complete Σ-consistent Δ, we have $\mathbf{M}^\Sigma, \Delta \Vdash A$. Hence, by the Truth Lemma, $A \in \Delta$ for every complete Σ-consistent Δ, whence by Corollary 13.4 (with $\Gamma = \emptyset$), $\Sigma \vdash A$.

Conversely, if $\Sigma \vdash A$ then by Proposition 13.2(1), every complete Σ-consistent Δ contains A, and hence by the Truth Lemma, $\mathbf{M}^\Sigma, \Delta \Vdash A$ for every $\Delta \in W^\Sigma$, i.e., $\mathbf{M}^\Sigma \Vdash A$. □

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

Corollary 13.15. *The basic modal logic* $\mathbf{K}$ *is complete with respect to the class of all models, i.e., if $\vDash A$ then* $\mathbf{K} \vdash A$.

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

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

13.8 Frame Completeness

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

Theorem 13.16. *If a normal modal logic Σ contains one of the formulas on the left-hand side of Table 13.1, then the canonical model for Σ has the corresponding property on the right-hand side.*

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

Table 13.1: Basic correspondence facts.

Proof. We take each of these up in turn.

Suppose Σ contains D, and let $\Delta \in W^\Sigma$; we need to show that there is a Δ' such that $R^\Sigma \Delta\Delta'$. It suffices to show that $\Box^{-1}\Delta$ is Σ-consistent, for then by Lindenbaum's Lemma, there is a complete Σ-consistent set $\Delta' \supseteq \Box^{-1}\Delta$, and by definition of R^Σ we have $R^\Sigma \Delta\Delta'$. So, suppose for contradiction that $\Box^{-1}\Delta$ is *not* Σ-consistent, i.e., $\Box^{-1}\Delta \vdash_\Sigma \bot$. By Lemma 13.7, $\Delta \vdash_\Sigma \Box\bot$, and since Σ contains D, also $\Delta \vdash_\Sigma \Diamond\bot$. But Σ is normal, so $\Sigma \vdash \neg\Diamond\bot$ (Proposition 12.4), whence also $\Delta \vdash_\Sigma \neg\Diamond\bot$, against the consistency of Δ.

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

Now suppose Σ contains B, and suppose $R^\Sigma \Delta\Delta'$ for Δ, $\Delta' \in W^\Sigma$. We need to show that $R^\Sigma \Delta'\Delta$, i.e., $\Box^{-1}\Delta' \subseteq \Delta$. By Lemma 13.9, this is equivalent to $\Diamond\Delta \subseteq \Delta'$. So suppose $A \in \Delta$. By B, also $\Box\Diamond A \in \Delta$. By the hypothesis that $R^\Sigma \Delta\Delta'$, we have that $\Box^{-1}\Delta \subseteq \Delta'$, and hence $\Diamond A \in \Delta'$, as required.

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

Now suppose Σ contains 5, suppose $R^{\Sigma}\Delta_1\Delta_2$ and $R^{\Sigma}\Delta_1\Delta_3$. We need to show $R^{\Sigma}\Delta_2\Delta_3$. The first hypothesis gives $\Box^{-1}\Delta_1 \subseteq \Delta_2$, and the second hypothesis is equivalent to $\Diamond\Delta_3 \subseteq \Delta_2$, by Lemma 13.9. To show $R^{\Sigma}\Delta_2\Delta_3$, by Lemma 13.9, it suffices to show $\Diamond\Delta_3 \subseteq \Delta_2$. So let $\Diamond A \in \Diamond\Delta_3$, i.e., $A \in \Delta_3$. By the second hypothesis $\Diamond A \in \Delta_1$ and by 5, $\Box\Diamond A \in \Delta_1$ as well. But now the first hypothesis gives $\Diamond A \in \Delta_2$, as desired. □

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

Theorem 13.17. *Let $\mathscr{C}_D$, $\mathscr{C}_T$, $\mathscr{C}_B$, $\mathscr{C}_4$, and $\mathscr{C}_5$ 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 $KA_1 \dots A_n$ is determined by the class of models $\mathscr{C} = \mathscr{C}_{A_1} \cap \dots \cap \mathscr{C}_{A_n}$.*

Proposition 13.18. *Let Σ be a normal modal logic; then:*

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

2. *If Σ contains the schema $\Diamond A \leftrightarrow \Box A$ then the canonical model for Σ is functional.*

3. *If Σ contains the schema $\Box\Box A \to \Box A$ then the canonical model for Σ is weakly dense.*

(see Table 11.2 for definitions of these frame properties).

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

2. This follows immediately from part (1) and the seriality proof in Theorem 13.16.

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

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

It suffices to show that Γ is Σ-consistent, for then by Lindenbaum's Lemma it can be extended to a complete Σ-consistent set Δ_3 such that $\Box^{-1}\Delta_1 \subseteq \Delta_3$ and $\Diamond\Delta_2 \subseteq \Delta_3$, i.e., $R^\Sigma \Delta_1\Delta_3$ and $R^\Sigma \Delta_3\Delta_2$ (by Lemma 13.9).

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

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

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

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

$$A_1, \dots, A_n \vdash_\Sigma (\Diamond B_1 \wedge \dots \wedge \Diamond B_m) \to \bot$$

by the deduction theorem

Proposition 12.20(4), and TAUT

$$A_1, \dots, A_n \vdash_\Sigma \Diamond(B_1 \wedge \dots \wedge B_m) \to \bot$$

since Σ is normal

$$A_1, \dots, A_n \vdash_\Sigma \neg\Diamond(B_1 \wedge \dots \wedge B_m)$$

by PL

$$A_1, \dots, A_n \vdash_\Sigma \Box\neg(B_1 \wedge \dots \wedge B_m)$$

$\Box\neg$ for $\neg\Diamond$

$$\Box A_1, \dots, \Box A_n \vdash_\Sigma \Box\Box\neg(B_1 \wedge \dots \wedge B_m)$$

by Lemma 13.6

$$\Box A_1, \dots, \Box A_n \vdash_\Sigma \Box\neg(B_1 \wedge \dots \wedge B_m)$$

by schema $\Box\Box A \to \Box A$

$$\Delta_1 \vdash_\Sigma \Box\neg(B_1 \wedge \dots \wedge B_m)$$

by monotonicity, Proposition 12.20(1)

$$\Box\neg(B_1 \wedge \dots \wedge B_m) \in \Delta_1$$

by deductive closure;

$$\neg(B_1 \wedge \dots \wedge B_m) \in \Delta_2$$

since $R^\Sigma \Delta_1 \Delta_2$. □

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

Problems

Problem 13.1. Complete the proof of Proposition 13.2.

Problem 13.2. Show that if Γ is complete Σ-consistent, then $\Diamond A \in \Gamma$ if and only if there is a complete Σ-consistent Δ such that $\Box^{-1}\Gamma \subseteq \Delta$ and $A \in \Delta$. *Do this without using Lemma 13.9.*

Problem 13.3. Complete the proof of Proposition 13.12.

CHAPTER 14

Modal Sequent Calculus

14.1 Introduction

The sequent calculus for propositional logic can be extended by additional rules that deal with $\Box$ and $\Diamond$. For instance, for **K**, we have **LK** plus:

$$\frac{\Gamma \Rightarrow \Delta, A}{\Box\Gamma \Rightarrow \Diamond\Delta, \Box A}\,\Box \qquad \frac{A, \Gamma \Rightarrow \Delta}{\Diamond A, \Box\Gamma \Rightarrow \Diamond\Delta}\,\Diamond$$

For extensions of **K**, additional rules have to be added as well.

Not every modal logic has such a sequent calculus. Even **S5**, which is semantically simple (it can be defined without using accessibility relations at all) is not known to have a sequent calculus that results from **LK** which is complete without the rule Cut. However, it has a cut-free complete *hypersequent* calculus.

14.2 Rules for K

The rules for the regular propositional connectives are the same as for regular sequent calculus **LK**. Axioms are also the same: any sequent of the form $A \Rightarrow A$ counts as an axiom.

For the modal operators $\Box$ and $\Diamond$, we have the following additional rules:

$$\dfrac{\Gamma \Rightarrow \Delta, A}{\Box\Gamma \Rightarrow \Diamond\Delta, \Box A}\,\Box \qquad \dfrac{A, \Gamma \Rightarrow \Delta}{\Diamond A, \Box\Gamma \Rightarrow \Diamond\Delta}\,\Diamond$$

Here, $\Box\Gamma$ means the sequence of formulas resulting from Γ by putting $\Box$ in front of every formula in Γ and $\Diamond\Delta$ is the sequence of formulas resulting from Δ by putting $\Diamond$ in front of every formula in Δ. Γ and Δ may be empty; in that case the corresponding part $\Box\Gamma$ and $\Diamond\Delta$ of the conclusion sequent is empty as well.

The restriction of adding a $\Box$ on the right and $\Diamond$ on the left to a single formula A is necessary. If we allowed to add $\Box$ to any number of formulas on the right or to add $\Diamond$ to any number of formulas on the left we would be able to derive:

$$\dfrac{\dfrac{\dfrac{A \Rightarrow A}{\Rightarrow A, \neg A}\,\neg\mathrm{R}}{\Rightarrow \Box A, \Box\neg A}\,\Box *}{\Rightarrow \Box A \lor \Box\neg A}\,\lor\mathrm{R} \qquad \dfrac{\dfrac{\dfrac{\dfrac{A \Rightarrow A}{\neg A, A \Rightarrow}\,\neg\mathrm{L}}{\Diamond\neg A, \Diamond A \Rightarrow}\,\Diamond *}{\Diamond A \Rightarrow \neg\Diamond\neg A}\,\neg\mathrm{R}}{\Rightarrow \Diamond A \to \neg\Diamond\neg A}\,\to\mathrm{R}$$

But $\Box A \lor \Box\neg A$ and $\Diamond A \to \neg\Diamond\neg A$ are not valid in $\mathbf{K}$.

If we allowed side formulas in addition to A in the premise, and allowed the $\Box$ rule to add $\Box$ to only A on the right, or allowed the $\Diamond$ rule to add $\Diamond$ to only A on the left (but do nothing to the side formulas) we would be able to derive:

$$\dfrac{\dfrac{\dfrac{\dfrac{A \Rightarrow A}{\Rightarrow A, \neg A}\,\neg\mathrm{R}}{\Rightarrow \neg A, A}\,\mathrm{XR}}{\Rightarrow \neg A, \Box A}\,\Box *}{\Rightarrow \neg A \lor \Box A}\,\lor\mathrm{R} \qquad \dfrac{\dfrac{\dfrac{\dfrac{A \Rightarrow A}{\neg A, A \Rightarrow}\,\neg\mathrm{L}}{\Diamond\neg A, A \Rightarrow}\,\Diamond *}{A \Rightarrow \neg\Diamond\neg A}\,\neg\mathrm{R}}{\Rightarrow A \to \neg\Diamond\neg A}\,\to\mathrm{R}$$

But $\neg A \lor \Box A$ (which is equivalent to $A \to \Box A$) and $A \to \neg\Diamond\neg A$ are not valid in $\mathbf{K}$.

14.3 Sequent Derivations for K

Example 14.1. We give a sequent calculus derivation that shows $\vdash (\Box A \land \Box B) \to \Box(A \land B)$.

$$
\begin{prooftree}
\AxiomC{$A \Rightarrow A$}
\doubleLine
\UnaryInfC{$B, A \Rightarrow A$}
\AxiomC{$B \Rightarrow B$}
\doubleLine
\UnaryInfC{$B, A \Rightarrow B$}
\RightLabel{$\land$R}
\BinaryInfC{$B, A \Rightarrow A \land B$}
\RightLabel{$\Box$}
\UnaryInfC{$\Box B, \Box A \Rightarrow \Box(A \land B)$}
\RightLabel{$\land$L}
\UnaryInfC{$\Box A \land \Box B, \Box A \Rightarrow \Box(A \land B)$}
\RightLabel{XL}
\UnaryInfC{$\Box A, \Box A \land \Box B \Rightarrow \Box(A \land B)$}
\RightLabel{$\land$L}
\UnaryInfC{$\Box A \land \Box B, \Box A \land \Box B \Rightarrow \Box(A \land B)$}
\RightLabel{CL}
\UnaryInfC{$\Box A \land \Box B \Rightarrow \Box(A \land B)$}
\RightLabel{$\to$R}
\UnaryInfC{$\Rightarrow (\Box A \land \Box B) \to \Box(A \land B)$}
\end{prooftree}
$$

Example 14.2. We give a sequent calculus derivation that shows $\vdash \Diamond(A \lor B) \to (\Diamond A \lor \Diamond B)$.

$$
\begin{prooftree}
\AxiomC{$A \Rightarrow A$}
\doubleLine
\UnaryInfC{$A \Rightarrow A, B$}
\AxiomC{$B \Rightarrow B$}
\doubleLine
\UnaryInfC{$B \Rightarrow A, B$}
\RightLabel{$\lor$L}
\BinaryInfC{$A \lor B \Rightarrow A, B$}
\RightLabel{$\Diamond$}
\UnaryInfC{$\Diamond(A \lor B) \Rightarrow \Diamond A, \Diamond B$}
\RightLabel{$\lor$R}
\UnaryInfC{$\Diamond(A \lor B) \Rightarrow \Diamond A, \Diamond A \lor \Diamond B$}
\RightLabel{XR}
\UnaryInfC{$\Diamond(A \lor B) \Rightarrow \Diamond A \lor \Diamond B, \Diamond A$}
\RightLabel{$\lor$R}
\UnaryInfC{$\Diamond(A \lor B) \Rightarrow \Diamond A \lor \Diamond B, \Diamond A \lor \Diamond B$}
\RightLabel{CR}
\UnaryInfC{$\Diamond(A \lor B) \Rightarrow \Diamond A \lor \Diamond B$}
\RightLabel{$\to$R}
\UnaryInfC{$\Rightarrow \Diamond(A \lor B) \to (\Diamond A \lor \Diamond B)$}
\end{prooftree}
$$

Here is a derivation of DUAL.

$$
\begin{prooftree}
\AxiomC{$A \Rightarrow A$}
\RightLabel{$\lnot$R}
\UnaryInfC{$\lnot A, A \Rightarrow$}
\RightLabel{$\Diamond$}
\UnaryInfC{$\Diamond\lnot A, \Box A \Rightarrow$}
\RightLabel{$\lnot$R}
\UnaryInfC{$\Box A \Rightarrow \lnot\Diamond\lnot A$}
\RightLabel{$\to$R}
\UnaryInfC{$\Rightarrow \Box A \to \lnot\Diamond\lnot A$}
\AxiomC{$A \Rightarrow A$}
\RightLabel{$\lnot$R}
\UnaryInfC{$\Rightarrow A, \lnot A$}
\RightLabel{XR}
\UnaryInfC{$\Rightarrow \lnot A, A$}
\RightLabel{$\Box$}
\UnaryInfC{$\Rightarrow \Diamond\lnot A, \Box A$}
\RightLabel{XR}
\UnaryInfC{$\Rightarrow \Box A, \Diamond\lnot A$}
\RightLabel{$\lnot$R}
\UnaryInfC{$\lnot\Diamond\lnot A \Rightarrow \Box A$}
\RightLabel{$\to$R}
\UnaryInfC{$\Rightarrow \lnot\Diamond\lnot A \to \Box A$}
\RightLabel{$\land$R}
\BinaryInfC{$\Rightarrow \Box A \leftrightarrow \lnot\Diamond\lnot A$}
\end{prooftree}
$$

$$\frac{A, \Gamma \Rightarrow \Delta}{\Box A, \Gamma \Rightarrow \Delta}\,\mathrm{T}\Box \qquad \frac{\Gamma \Rightarrow \Delta, A}{\Gamma \Rightarrow \Delta, \Diamond A}\,\mathrm{T}\Diamond$$

$$\frac{\Gamma \Rightarrow \Delta}{\Box\Gamma \Rightarrow \Diamond\Delta}\,\mathrm{D}$$

$$\frac{\Gamma, \Diamond\Pi \Rightarrow \Box\Delta, \Lambda, A}{\Box\Gamma, \Pi \Rightarrow \Delta, \Diamond\Lambda, \Box A}\,\mathrm{B}\Box \qquad \frac{A, \Diamond\Gamma, \Pi \Rightarrow \Box\Lambda, \Delta}{\Diamond A, \Gamma, \Box\Pi \Rightarrow \Lambda, \Diamond\Delta}\,\mathrm{B}\Diamond$$

$$\frac{\Box\Gamma \Rightarrow \Diamond\Delta, A}{\Box\Gamma \Rightarrow \Diamond\Delta, \Box A}\,4\Box \qquad \frac{A, \Box\Gamma \Rightarrow \Diamond\Delta}{\Diamond A, \Box\Gamma \Rightarrow \Diamond\Delta}\,4\Diamond$$

$$\frac{\Box\Gamma, \Diamond\Pi \Rightarrow \Box\Delta, \Diamond\Lambda, A}{\Box\Gamma, \Diamond\Pi \Rightarrow \Box\Delta, \Diamond\Lambda, \Box A}\,5\Box \qquad \frac{A, \Diamond\Gamma, \Box\Pi \Rightarrow \Diamond\Delta, \Box\Lambda}{\Diamond A, \Diamond\Gamma, \Box\Pi \Rightarrow \Diamond\Delta, \Box\Lambda}\,5\Diamond$$

Table 14.1: More modal rules.

14.4 Rules for Other Accessibility Relations

In order to deal with logics determined by special accessibility relations, we consider the additional rules in Table 14.1.

Adding these rules results in systems that are sound and complete for the logics given in Table 14.2.

Example 14.3. We give a sequent derivation that shows $\mathbf{K4} \vdash 4$, i.e., $\Box A \to \Box\Box A$.

$$\dfrac{\dfrac{\Box A \Rightarrow \Box A}{\Box A \Rightarrow \Box\Box A}\,4\Box}{\Rightarrow \Box A \to \Box\Box A}\,{\to}\mathrm{R}$$

Logic	R is ...	Rules
$T = KT$	reflexive	$\Box$, T$\Box$, T$\Diamond$
$D = KD$	serial	$\Box$, D
$K4$	transitive	$\Box$, 4$\Box$, 4$\Diamond$
$B = KTB$	reflexive, symmetric	$\Box$, T$\Box$, T$\Diamond$ B$\Box$, B$\Diamond$
$S4 = KT4$	reflexive, transitive	$\Box$, T$\Box$, T$\Diamond$ 4$\Box$, 4$\Diamond$
$S5 = KT5$	reflexive, transitive, euclidean	$\Box$, T$\Box$, T$\Diamond$ 5$\Box$, 5$\Diamond$

Table 14.2: Sequent rules for various modal logics.

Example 14.4. We give a sequent derivation that shows $S5 \vdash 5$, i.e., $\Diamond A \to \Box\Diamond A$.

$$\dfrac{\dfrac{\Diamond A \Rightarrow \Diamond A}{\Diamond A \Rightarrow \Box\Diamond A}\,5\Box}{\Rightarrow \Diamond A \to \Box\Diamond A}\,{\to}\mathrm{R}$$

Example 14.5. The sequent calculus for $S5$ is not complete without the Cut rule; e.g., $\Diamond\Box A \to A$, which is valid in $S5$, has no proof without Cut. Here is a derivation using Cut:

$$\dfrac{\dfrac{\dfrac{\Box A \Rightarrow \Box A}{\Diamond\Box A \Rightarrow \Box A}\,5\Diamond \qquad \dfrac{A \Rightarrow A}{\Box A \Rightarrow A}\,\mathrm{T}\Box}{\Diamond\Box A \Rightarrow A}\,\mathrm{Cut}}{\Rightarrow \Diamond\Box A \to A}\,{\to}\mathrm{R}$$

Problems

Problem 14.1. Find sequent calculus proofs in K for the following formulas:

1. $\Box\neg p \to \Box(p \to q)$

2. $(\Box p \lor \Box q) \to \Box(p \lor q)$

3. $\Diamond p \to \Diamond(p \lor q)$

4. $\Box(p \land q) \to \Box p$

Problem 14.2. Give sequent derivations that show the following:

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

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

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

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

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

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

PART V

But you can't tell me what to think!

CHAPTER 15

Epistemic Logics

15.1 Introduction

Just as modal logic deals with *modal propositions* and the entailment relations among them, epistemic logic deals with *epistemic propositions* and the entailment relations among them. Rather than interpreting the modal operators as representing possibility and necessity, the unary connectives are interpreted in epistemic or doxastic ways, to model knowledge and belief. For example, we might want to express claims like the following:

1. Richard knows that Calgary is in Alberta.
2. Audrey thinks it is possible that a dog is on the couch.
3. Richard knows that Audrey knows that her class is on Tuesdays.
4. Everyone knows that a year has 12 months.

Contemporary epistemic logic is often traced to Jaako Hintikka's *Knowledge and Belief*, from 1962, and it was written at a time when possible worlds semantics were becoming increasingly more used in logic. In fact, epistemic logics use most of the same semantic tools as other modal logics, but will interpret them differently. The main change is in what we take the *accessibility relation* to represent. In epistemic logics, they represent some form of *epistemic possibility*. We'll see that the epistemic notion that we're modelling will affect the constraints that we want to place on the accessibility relation. And we'll also see what happens to correspondence theory when it is given an epistemic interpretation. You'll notice that the examples above mention two agents: Richard and Audrey, and the relationship between the things that each one knows. The epistemic logics we'll consider will be multi-agent logics, in which such things can be expressed. In contrast, a single-agent epistemic logic would only talk about what one individual knows or believes.

15.2 The Language of Epistemic Logic

Definition 15.1. Let G be a set of agent-symbols. The basic language of multi-agent epistemic logic contains

1. The propositional constant for falsity $\bot$.

2. A countably infinite set of propositional variables: p_0, p_1, p_2, ...

3. The propositional connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional).

4. The knowledge operator K_a where $a \in G$.

If we are only concerned with the knowledge of a single agent in our system, we can drop the reference to the set G, and individual agents. In that case, we only have the basic operator K.

Definition 15.2. *Formulas* of the epistemic language are inductively defined as follows:

1. $\bot$ is an atomic formula.

2. Every propositional variable p_i is an (atomic) formula.

3. If A is a formula, then $\neg A$ is a formula.

4. If A and B are formulas, then $(A \land B)$ is a formula.

5. If A and B are formulas, then $(A \lor B)$ is a formula.

6. If A and B are formulas, then $(A \to B)$ is a formula.

7. If A is a formula and $a \in G$, then $\mathsf{K}_a A$ is a formula.

8. Nothing else is a formula.

If a formula A does not contain K_a, we say it is *modal-free.*

Definition 15.3. While the K operator is intended to symbolize individual knowledge, E, often read as "everybody knows," symbolizes group knowledge. Where $G' \subseteq G$, we define $\mathsf{E}_{G'} A$ as an abbreviation for

$$\bigwedge_{b \in G'} \mathsf{K}_b A.$$

We can also define an even stronger sense of knowledge, namely *common knowledge* among a group of agents G. When a piece of information is common knowledge among a group of agents, it means that for every combination of agents in that group, they all know that each other knows that each other knows ... ad infinitum. This is significantly stronger than group knowledge, and it is easy to come up with relational models in which a formula is group knowledge, but not common knowledge. We will use $\mathsf{C}_G A$ to symbolize "it is common knowledge among G that A."

15.3 Relational Models

The basic semantic concept for epistemic logics is the same as that of ordinary modal logics. Relational models still consist of a set of worlds, and an assignment that determines which propositional variables count as "true" at which worlds. And if we are only dealing with a single agent, we have a single accessibility relation as usual. However, if we have a multi-agent epistemic logic, then our single accessibility relation becomes a set of accessibility relations, one for each a in our set of agent symbols G.

A *relational model* consists of a set of worlds, which are related by binary accessibility relations—one for each agent—together with an assignment which determines which propositional variables are true at which worlds.

Definition 15.4. A *model* for the multi-agent epistemic language is a triple $M = \langle W, R, V \rangle$, where

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

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

The mechanics are just like the mechanics for normal modal logic, just with more accessibility relations added in. For a given agent, we will generally interpret their accessibility relation as representing something about their informational states. For example, we often treat $R_a ww'$, as expressing that w' is consistent with a's information at w. Or to put it another way, at w, they cannot tell the difference between world w and world w'.

15.4 Truth at a World

Just as with normal modal logic, every epistemic model determines which formulas count as true at which worlds in it. We use the same notation "model M makes formula A true at world w" for the basic notion of relational semantics. The relation is defined inductively and is identical to the normal modal case for all non-modal operators.

Definition 15.5. *Truth of a formula A at w* in a M, in symbols: $M, w \Vdash A$, is defined inductively as follows:

1. $A \equiv \bot$: Never $M, w \Vdash \bot$.
2. $M, w \Vdash p$ iff $w \in V(p)$
3. $A \equiv \neg B$: $M, w \Vdash A$ iff $M, w \nVdash B$.
4. $A \equiv (B \wedge C)$: $M, w \Vdash A$ iff $M, w \Vdash B$ and $M, w \Vdash C$.
5. $A \equiv (B \vee C)$: $M, w \Vdash A$ iff $M, w \Vdash B$ or $M, w \Vdash C$ (or both).
6. $A \equiv (B \to C)$: $M, w \Vdash A$ iff $M, w \nVdash B$ or $M, w \Vdash C$.
7. $A \equiv \mathsf{K}_a B$: $M, w \Vdash A$ iff $M, w' \Vdash B$ for all $w' \in W$ with $R_a w w'$

Here's where we need to think about restrictions on our accessibility relations, though. After all, by clause (7), a formula $\mathsf{K}_a B$ is true at w whenever there are no w' with $R_a w w'$. This is the same clause as in normal modal logic; when a world has no successors, all $\Box$-formulas are vacuously true there. This seems extremely counterintuitive if we think about K as representing *knowledge*. After all, we tend to think that there are *no* circumstances under which an agent might know both A and $\neg A$ at the same time.

One solution is to ensure that our accessibility relation in epistemic logic will always be *reflexive*. This roughly corresponds to the idea that the actual world is consistent with an agent's

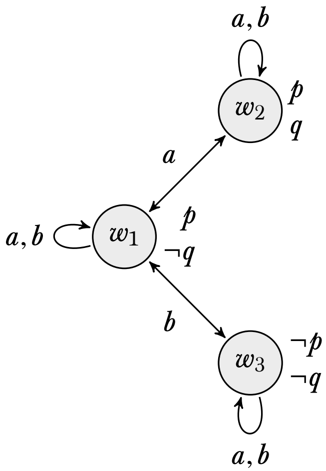


Figure 15.1: A simple epistemic model.

information. In fact, epistemic logics typically use S5, but others might use weaker systems depending on what exactly they want the K_a relation to represent.

Now that we have given our basic definition of truth at a world, the other semantic concepts from normal modal logic, such as modal validity and entailment, simply carry over, applied to this new way of thinking about the interpretation for the modal operators.

We are now also in a position to give truth conditions for the common knowledge operator C_G. Recall from appendix B.6 that the *transitive closure* R^+ of a relation R is defined as

$$R^+ = \bigcup_{n \in \mathbb{N}} R^n,$$

where

$$\begin{aligned} R^0 &= R \text{ and} \\ R^{n+1} &= \{\langle x, z\rangle : \exists y (R^n xy \land Ryz)\}. \end{aligned}$$

If R is ...	*then ... is true in M:*
	$\mathsf{K}(p \to q) \to (\mathsf{K}p \to \mathsf{K}q)$ (Closure)
reflexive: $\forall w Rww$	$\mathsf{K}p \to p$ (Veridicality)
transitive: $\forall u \forall v \forall w((Ruv \land Rvw) \to Ruw)$	$\mathsf{K}p \to \mathsf{K}\mathsf{K}p$ (Positive Introspection)
euclidean: $\forall w \forall u \forall v((Rwu \land Rwv) \to Ruv)$	$\neg\mathsf{K}p \to \mathsf{K}\neg\mathsf{K}p$ (Negative Introspection)

Table 15.1: Four epistemic principles.

Then, where G is a group of agents, we define $R_G = (\bigcup_{b \in G} R_b)^+$ to be the transitive closure of the union of all agents' accessibility relations.

Definition 15.6. If $G' \subseteq G$, we let $M, w \Vdash \mathsf{C}_{G'}A$ iff for every w' such that $R_{G'}ww'$, $M, w' \Vdash A$.

15.5 Accessibility Relations and Epistemic Principles

Given what we already know about frame correspondence in normal modal logics, we might want to see what the characteristic formulas look like given epistemic interpretations. We have already said that epistemic logics are typically interpreted in S5. So let's take a look at how various epistemic principles are represented, and consider how they correspond to various frame conditions.

Recall from normal modal logic, that different modal formulas characterized different properties of accessibility relations. This table picks out a few that correspond to particular epistemic principles.

Veridicality, corresponding to the T axiom, is often treated as the most uncontroversial of these principles, as it represents that claim that if a formula is known, then it must be true. Closure, as well as Positive and Negative Introspection are much more contested.

Closure, corresponding to the K axiom, represents the idea that an agent's knowledge is closed under implication. This might seem plausible to us in some cases. For instance, I might know that if I am in Victoria, then I am on Vancouver Island. Barring odd skeptical scenarios, I do know that I am in Victoria, and this should also suggest that I know I am on Vancouver Island. So in this case, the logical closure of my knowledge might seem relatively intuitive. On the other hand, we do not always think through the consequences of our knowledge, and so this might lead to less intuitive results in other cases.

Positive Introspection, sometimes known as the KK-principle, is sometimes articulated as the statement that if I know something, then I know that I know. It is the epistemic counterpart of the 4 axiom. Correspondingly, negative introspection is articulated as the statement that if I *don't* know something, then I know that I don't know it, which is the counterpart of the 5 axiom. Both of these seem to admit of relatively ordinary counterexamples, in which I am unsure whether or not I know something that I do in fact know.

15.6 Bisimulations

One remaining question that we might have about the expressive power of our epistemic language has to do with the relationship between models and the formulas that hold in them. We have seen from our frame correspondence results that when certain formulas are valid in a frame, they will also ensure that those frames satisfy certain properties. But does our modal language, for example, allow us to distinguish between a world at which there is a reflexive arrow, and an infinite chain of worlds, each of which leads to the next? That is, is there any formula A that might hold at only one of these two worlds?

Bisimulation is a relationship that we can define between relational models to say that they have effectively the same structure. And as we will see, it will capture a sense of equivalence between models that can be captured in our epistemic language.

Definition 15.7 (Bisimulation). Let $M_1 = \langle W_1, R_1, V_1 \rangle$ and $M_2 = \langle W_2, R_2, V_2 \rangle$ be two relational models. And let $\mathcal{R} \subseteq W_1 \times W_2$ be a binary relation. We say that $\mathcal{R}$ is a *bisimulation* when for every $\langle w_1, w_2 \rangle \in \mathcal{R}$, we have:

1. $w_1 \in V_1(p)$ iff $w_2 \in V_2(p)$ for all propositional variables p.
2. For all agents $a \in A$ and worlds $v_1 \in W_1$, if $R_{1_a} w_1 v_1$ then there is some $v_2 \in W_2$ such that $R_{2_a} w_2 v_2$, and $\langle v_1, v_2 \rangle \in \mathcal{R}$.
3. For all agents $a \in A$ and worlds $v_2 \in W_2$, if $R_{2_a} w_2 v_2$ then there is some $v_1 \in W_1$ such that $R_{1_a} w_1 v_1$, and $\langle v_1, v_2 \rangle \in \mathcal{R}$.

When there is a bisimulation between M_1 and M_2 that links worlds w_1 and w_2, we can also write $\langle M_1, w_1 \rangle \leftrightarroweq \langle M_2, w_2 \rangle$, and call $\langle M_1, w_1 \rangle$ and $\langle M_2, w_2 \rangle$ *bisimilar*.

The different clauses in the bisimulation relation ensure different things. Clause 1 ensures that bisimilar worlds will satisfy the same modal-free formulas, since it ensures agreement on all propositional variables. The other two clauses, sometimes referred to as “forth” and “back,” respectively, ensure that the accessibility relations will have the same structure.

Theorem 15.8. *If $\langle M_1, w_1 \rangle \leftrightarroweq \langle M_2, w_2 \rangle$, then for every formula A, we have that $M_1, w_1 \Vdash A$ iff $M_2, w_2 \Vdash A$.*

Even though the two models pictured in Figure 15.2 aren’t quite the same as each other, there is a bisimulation linking worlds w_1 and v_1. This bisimulation will also link both w_2 and w_3 to v_2, with the idea being that there is nothing expressible in our modal language that can really distinguish between them.

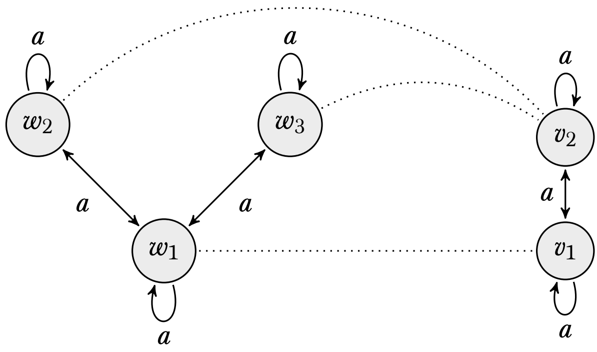


Figure 15.2: Two bisimilar models.

The situation would be different if w_2 and w_3 satisfied different propositional variables, however.

15.7 Public Announcement Logic

Dynamic epistemic logics allow us to represent the ways in which agents' knowledge changes over time, or as they gain new information. Many of these represent changes in knowlege using informational *events* or *updates*. The most basic kind of update is a public announcement in which some formula is truthfully announced and all of the agents witness this taking place together. To do this, we expand the language as follows

Definition 15.9. Let G be a set of agent-symbols. The basic language of multi-agent epistemic logic with public announcements contains

1. The propositional constant for falsity $\bot$.

2. A countably infinite set of propositional variables: p_0, p_1,

$p_2, \ldots$

3. The propositional connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional)

4. The knowledge operator K_a where $a \in G$.

5. The public announcement operator $[B]$ where B is a formula.

The public announcement operator functions as a box operator, and our inductive definition of the language is given accordingly:

Definition 15.10. *Formulas* of the epistemic 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 and $a \in G$, then $\mathsf{K}_a A$ is a formula.

8. If A and B are formulas, then $[A]B$ is a formula.

9. Nothing else is a formula.

The intended reading of the formula $[A]B$ is “After A is truthfully announced, B holds. It will sometimes also be useful to talk about common knowledge in the context of public announcements, so the language may also include the common knowledge operator $\mathsf{C}_G A$.

15.8 Semantics of Public Announcement Logic

Relational models for public announcement logics are the same as they were in epistemic logics. However, the semantics for the public announcement operator are something new.

Definition 15.11. *Truth of a formula A at w* in a $M = \langle W, R, V\rangle$, in symbols: $M, w \Vdash A$, is defined inductively as follows:

1. $A \equiv \bot$: Never $M, w \Vdash \bot$.
2. $M, w \Vdash p$ iff $w \in V(p)$
3. $A \equiv \neg B$: $M, w \Vdash A$ iff $M, w \nVdash B$.
4. $A \equiv (B \wedge C)$: $M, w \Vdash A$ iff $M, w \Vdash B$ and $M, w \Vdash C$.
5. $A \equiv (B \vee C)$: $M, w \Vdash A$ iff $M, w \Vdash B$ or $M, w \Vdash C$ (or both).
6. $A \equiv (B \rightarrow C)$: $M, w \Vdash A$ iff $M, w \nVdash B$ or $M, w \Vdash C$.
7. $A \equiv \mathsf{K}_a B$: $M, w \Vdash A$ iff $M, w' \Vdash B$ for all $w' \in W$ with $R_a w w'$
8. $A \equiv [B]C$: $M, w \Vdash A$ iff $M, w \Vdash B$ implies $M \mid B, w \Vdash C$

   Where $M \mid B = \langle W', R', V'\rangle$ is defined as follows:

   a) $W' = \{u \in W : M, u \Vdash B\}$. So the worlds of $M \mid B$ are the worlds in M at which B holds.

   b) $R'_a = R_a \cap (W' \times W')$. Each agent's accessibility relation is simply restricted to the worlds that remain in W'.

   c) $V'(p) = \{u \in W' : u \in V(p)\}$. Similarly, the propositional valuations at worlds remain the same, representing the idea that informational events will not change the truth value of propositional variables.

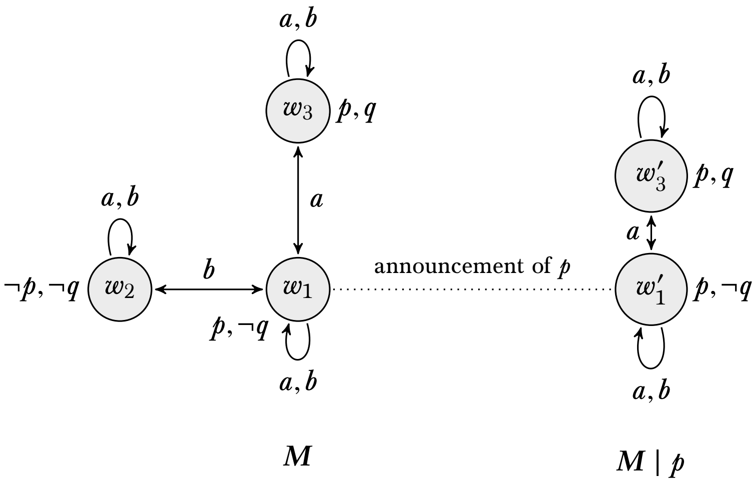


Figure 15.3: Before and after the public announcement of p.

What is distinctive, then, about public announcement logics, is that the truth of a formula at M can sometimes only be decided by referring to a model other than M itself.

Notice also that our semantics treats the announcement operator as a $\Box$ operator, and so if a formula A cannot be truthfully announced at a world, then $[A]B$ will hold there trivially, just as all $\Box$ formulas hold at endpoints.

We can see the public announcement of a formula as shrinking a model, or restricting it to the worlds at which the formula was true. Figure 15.3 gives an example of the effects of publicly announcing p. One notable thing about that model is that agent b learns that p as a result of the announcement, while agent a does not (since a already knew that p was true).

More formally, we have $M, w_1 \Vdash \neg \mathsf{K}_b p$ but $M \mid p, w_1' \Vdash \mathsf{K}_b p$. This implies that $M, w_1 \Vdash [p]\mathsf{K}_b p$. But we have some even stronger claims that we can make about the result of the announcement. In fact, it is the case that $M, w_1 \Vdash [p]\mathsf{C}_{\{a,b\}} p$. In other words, after p is announced, it becomes *common knowledge*.

We might wonder, though, whether this holds in the general case, and whether a truthful announcement of A will *always* result in A becoming common knowledge. It may be surprising that the answer is no. And in fact, it is possible to truthfully announce formulas that will no longer be true once they are announced. For example, consider the effects of announcing $p \land \lnot \mathsf{K}_b p$ at w_1 in Figure 15.3. In fact, $\boldsymbol{M} \mid p$ and $\boldsymbol{M} \mid (p \land \lnot \mathsf{K}_b p)$ are the same model. However, as we have already noted, $\boldsymbol{M} \mid p, w_1' \Vdash \mathsf{K}_b p$. Therefore, $\boldsymbol{M} \mid (p \land \lnot \mathsf{K}_b p), w_1' \Vdash \lnot(p \land \lnot \mathsf{K}_b p)$, so this is a formula that becomes false once it has been announced.

PART VI

Is this going to go on forever?

CHAPTER 16

Temporal Logics

16.1 Introduction

Temporal logics deal with claims about things that will or have been the case. Arthur Prior is credited as the originator of temporal logic, which he called *tense logic.* Our treatment of temporal logic here will largely follow Prior's original modal treatment of introducing temporal operators into the basic framework of propositional logic, which treats claims as generally lacking in tense.

For example, in propositional logic, I might talk about a dog, Beezie, who sometimes sits and sometimes doesn't sit, as dogs are wont to do. It would be contradictory in classical logic to claim that Beezie is sitting and also that Beezie is not sitting. But obviously both can be true, just not at the same time; adding temporal operators to the language can allow us to express that claim relatively easily. The addition of temporal operators also allows us to account for the validity of inferences like the one from "Beezie will get a treat or a ball" to "Beezie will get a treat or Beezie will get a ball."

However, a lot of philosophical issues arise with temporal logic that might lead us to adopt one framework of temporal logic over another. For example, a future contingent is a statement about the future that is neither necessary nor impossible. If we say “Richard will go to the grocery store tomorrow," we are expressing a claim about something that has not yet happened, and whose truth value is contestable. In fact, it is contestable whether that claim can even be *assigned* a truth value in the first place. If we are strict determinists, then perhaps we can be comfortable with the idea that this sentence is in fact true or false, even before the event in question is supposed to take place—it just may be that we do not know its truth value yet. In contrast, we might believe in a genuinely open future, in which the truth values of future contingents are undetermined.

As it turns out, a lot of these commitments about the structure and nature of time are built in to our choices of models and frameworks of temporal logics. For example, we might ask ourselves whether we should construct models in which time is linear, branching or even circular. We might have to make decisions about whether our temporal models will have beginning and end points, and whether time is to be represented using discrete instants or as a continuum.

16.2 Semantics for Temporal Logic

Definition 16.1. The basic language of temporal logic contains

1. The propositional constant for falsity $\bot$.
2. A countably infinite set of propositional variables: p_0, p_1, p_2, ...
3. The propositional connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional).
4. Past operators P and H.
5. Future operators F and G.

Later on, we will discuss the potential addition of other kinds of modal operators.

Definition 16.2. *Formulas* of the temporal language are inductively defined as follows:

1. $\bot$ is an atomic formula.
2. Every propositional variable p_i is an (atomic) formula.
3. If A 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 $\mathsf{P}A$, $\mathsf{H}A$, FA, $\mathsf{G}A$ are all formulas.
8. Nothing else is a formula.

The semantics of temporal logics are given in terms of relational models, as with other kinds of intensional logics.

Definition 16.3. A *model* for temporal language is a triple $M = \langle T, \prec, V \rangle$, where

1. T is a nonempty set, interpreted as points in time.
2. $\prec$ is a binary relation on T.
3. V is a function assigning to each propositional variable p a set $V(p)$ of points in time.

When $t \prec t'$ holds, we say that t *precedes* t'. When $t \in V(p)$ we say p is *true at* t.

For now, you will notice that we do not impose any conditions on our precedence relation $\prec$. This means that at present, there are no restrictions on the structure of our temporal models, so we could have models in which time is linear, branching, circular, or has any structure whatsoever.

Just as with normal modal logic, every temporal model determines which formulas count as true at which points in it. We use the same notation "model M makes formula A true at point t" for the basic notion of relational semantics. The relation is defined inductively and is identical to the normal modal case for all non-modal operators.

Definition 16.4. *Truth of a formula A at t* in a M, in symbols: $M, t \Vdash A$, is defined inductively as follows:

1. $A \equiv \bot$: Never $M, t \Vdash \bot$.
2. $M, t \Vdash p$ iff $t \in V(p)$
3. $A \equiv \neg B$: $M, t \Vdash A$ iff $M, t \nVdash B$.
4. $A \equiv (B \wedge C)$: $M, t \Vdash A$ iff $M, t \Vdash B$ and $M, t \Vdash C$.
5. $A \equiv (B \vee C)$: $M, t \Vdash A$ iff $M, t \Vdash B$ or $M, t \Vdash C$ (or both).
6. $A \equiv (B \rightarrow C)$: $M, t \Vdash A$ iff $M, t \nVdash B$ or $M, t \Vdash C$.
7. $A \equiv \mathsf{P}B$: $M, t \Vdash A$ iff $M, t' \Vdash B$ for some $t' \in T$ with $t' \prec t$
8. $A \equiv \mathsf{H}B$: $M, t \Vdash A$ iff $M, t' \Vdash B$ for every $t' \in T$ with $t' \prec t$
9. $A \equiv \mathsf{F}B$: $M, t \Vdash A$ iff $M, t' \Vdash B$ for some $t' \in T$ with $t \prec t'$
10. $A \equiv \mathsf{G}B$: $M, t \Vdash A$ iff $M, t' \Vdash B$ for every $t' \in T$ with $t \prec t'$

Based on the semantics, you might be able to see that the operators P and H are duals, as well as the operators F and G, such that we could define $\mathsf{H}A$ as $\neg\mathsf{P}\neg A$, and the same with G and F.

If $\prec$ *is ...*	*then ... is true in* M:
transitive: $\forall u \forall v \forall w((u \prec v \land v \prec w) \to u \prec w)$	$\mathsf{FF}p \to \mathsf{F}p$
linear: $\forall w \forall v(w \prec v \lor w = v \lor v \prec w)$	$(\mathsf{FP}p \lor \mathsf{PF}p) \to (\mathsf{P}p \lor p \lor \mathsf{F}p)$
dense: $\forall w \forall v(w \prec v \to \exists u(w \prec u \land u \prec v))$	$\mathsf{F}p \to \mathsf{FF}p$
unbounded (past): $\forall w \exists v(v \prec w)$	$\mathsf{H}p \to \mathsf{P}p$
unbounded (future): $\forall w \exists v(w \prec v)$	$\mathsf{G}p \to \mathsf{F}p$

Table 16.1: Some temporal frame correspondence properties.

16.3 Properties of Temporal Frames

Given that our temporal models do not impose any conditions on the relation $\prec$, the only one of our familiar axioms that holds in all models is K, or its analogues K_{G} and K_{H}:

$$\mathsf{G}(p \to q) \to (\mathsf{G}p \to \mathsf{G}q) \tag{K_{G}}$$

$$\mathsf{H}(p \to q) \to (\mathsf{H}p \to \mathsf{H}q) \tag{K_{H}}$$

However, if we want our models to impose stricter conditions on how time is represented, for instance by ensuring that $\prec$ is a linear order, then we will end up with other validities in our models.

Several of the properties from Table 16.1 might seem like desirable features for a model that is intended to represent time. However, it is worth noting that, even though we can impose whichever conditions we like on the $\prec$ relation, not all conditions correspond to formulas that can be expressed in the language of temporal logic. For example, irreflexivity, or the idea that

$\forall w \neg(w \prec w)$, does not have a corresponding formula in temporal logic.

16.4 Additional Operators for Temporal Logic

In addition to the unary operators for past and future, temporal logics also sometimes include binary operators S and U, intended to symbolize "since" and "until". This means adding S and U into the language of temporal logic and adding the following clause into the definition of a temporal formula:

> If A and B are formulas, then $(\mathsf{S}AB)$ and $(\mathsf{U}AB)$ are both formulas.

The semantics for these operators are then given as follows:

Definition 16.5. *Truth of a formula A at t* in a M:

1. $A \equiv \mathsf{S}BC$: $M, t \Vdash A$ iff $M, t' \Vdash B$ for some $t' \in T$ with $t' \prec t$, and for all s with $t' \prec s \prec t$, $M, s \Vdash C$
2. $A \equiv \mathsf{U}BC$: $M, t \Vdash A$ iff $M, t' \Vdash B$ for some $t' \in T$ with $t \prec t'$, and for all s with $t \prec s \prec t'$, $M, s \Vdash C$

The intuitive reading of $\mathsf{S}BC$ is "Since B was the case, C has been the case." And the intuitive reading of $\mathsf{U}BC$ is "Until B will be the case, C will be the case."

16.5 Possible Histories

The relational models of temporal logic that we have been using are extremely flexible, since we do not have to place any restrictions on the accessibility relation. This means that temporal models can branch in the past and in the future, but we might want to consider a more "modal" conception of branching, in which we consider sequences of events as possible histories. This does not necessarily require changing our language, though we might also add our "ordinary" modal operators $\Box$ and $\Diamond$, and we could also consider adding epistemic accessibility relations to represent changes in agents' knowledge over time.

Definition 16.6. A *possible histories model* for the temporal language is a triple $M = \langle T, C, V \rangle$, where

1. T is a nonempty set, interpreted as states in time.

2. C is a set of computational paths, or *possible histories* of a system. In other words, C is a set of sequences σ of states $s_1, s_2, s_3, \ldots$, where every $s_i \in T$.

3. V is a function assigning to each propositional variable p a set $V(p)$ of points in time.

To make things simpler, we will also generally assume that when a history is in C, then so are all of its suffixes. For example, if s_1, s_2, s_3 is a sequence in C, then so are s_2, s_3 and s_3. Also, when two states s_i and s_j appear in a sequence σ, we say that $s_i \prec_\sigma s_j$ when $i < j$. When $t \in V(p)$ we say p is *true at* t.

The one relevant change is that when we evaluate the truth of a formula at a point in time t in a model M, we do so relative to a history σ, in which t appears as a state. We do not need to change any of the semantics for propositional variables or for truth-functional connectives, though. All of those are exactly as they were in Definition 16.4, since none of those will make reference to σ. However, we now redefine our future operator F and add our $\Diamond$ operator with respect to these histories.

Definition 16.7. *Truth of a formula A at t, σ* in M, in symbols: $M, t, \sigma \Vdash A$:

1. $A \equiv \mathsf{F}B$: $M, t, \sigma \Vdash A$ iff $M, t', \sigma \Vdash B$ for some $t' \in T$ such

that $t \prec_\sigma t'$.

2. $A \equiv \Diamond B$: $M, t, \sigma \Vdash A$ iff $M, t, \sigma' \Vdash B$ for some $\sigma' \in C$ in which t occurs.

Other temporal and modal operators can be defined similarly. However, we can now represent claims that combine tense and modality. For example, we might symbolize "p will not occur, but it might have occurred" using the formula $\neg \mathsf{F} p \land \Diamond \mathsf{F} p$. This would hold at a point and a history at which p does not become true at a successor state, but there is an alternative history at which p will become true.

PART VII

What if things were different?

CHAPTER 17

Introduction

17.1 The Material Conditional

In its simplest form in English, a conditional is a sentence of the form "If ... then ...," where the ... are themselves sentences, such as "If the butler did it, then the gardener is innocent." In introductory logic courses, we earn to symbolize conditionals using the $\rightarrow$ connective: symbolize the parts indicated by ..., e.g., by formulas A and B, and the entire conditional is symbolized by $A \rightarrow B$.

The connective $\rightarrow$ is *truth-functional*, i.e., the truth value—$\mathbb{T}$ or $\mathbb{F}$—of $A \rightarrow B$ is determined by the truth values of A and B: $A \rightarrow B$ is true iff A is false or B is true, and false otherwise. Relative to a truth value assignment v, we define $v \vDash A \rightarrow B$ iff $v \nvDash A$ or $v \vDash B$. The connective $\rightarrow$ with this semantics is called the *material conditional*.

This definition results in a number of elementary logical facts. First of all, the deduction theorem holds for the material conditional:

$$\text{If } \Gamma, A \vDash B \text{ then } \Gamma \vDash A \rightarrow B \tag{17.1}$$

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

$$A \rightarrow B \vDash \neg A \lor B \tag{17.2}$$

$$\neg A \lor B \vDash A \rightarrow B \tag{17.3}$$

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

$$B \vDash A \rightarrow B \tag{17.4}$$
$$\neg A \vDash A \rightarrow B \tag{17.5}$$

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

$$\neg(A \rightarrow B) \vDash A \land \neg B \tag{17.6}$$
$$A \land \neg B \vDash \neg(A \rightarrow B) \tag{17.7}$$

The material conditional supports modus ponens:

$$A, A \rightarrow B \vDash B \tag{17.8}$$

The material conditional agglomerates:

$$A \rightarrow B, A \rightarrow C \vDash A \rightarrow (B \land C) \tag{17.9}$$

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

$$A \rightarrow B \vDash (A \land C) \rightarrow B \tag{17.10}$$

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

$$A \rightarrow B, B \rightarrow C \vDash A \rightarrow C \tag{17.11}$$

The material conditional is equivalent to its contrapositive:

$$A \rightarrow B \vDash \neg B \rightarrow \neg A \tag{17.12}$$
$$\neg B \rightarrow \neg A \vDash A \rightarrow B \tag{17.13}$$

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.

17.2 Paradoxes of the Material Conditional

One of the first to criticize the use of $A \rightarrow B$ as a way to symbolize "if … then …" statements of English was C. I. Lewis. Lewis was criticizing the use of the material condition in Whitehead and Russell's *Principia Mathematica*, who pronounced $\rightarrow$ as "implies." Lewis rightly complained that if $\rightarrow$ meant "implies," then any false proposition p implies that p implies q, since $p \rightarrow (p \rightarrow q)$ is true if p is false, and that any true proposition q implies that p implies q, since $q \rightarrow (p \rightarrow q)$ is true if q is true.

Logicians of course know that *implication*, i.e., logical entailment, is not a connective but a relation between formulas or statements. So we should just not read $\rightarrow$ as "implies" to avoid confusion.[1] As long as we don't, the particular worry that Lewis had simply does not arise: p does not "imply" q even if we think of p as standing for a false English sentence. To determine if $p \vDash q$ we must consider *all* valuations, and $p \nvDash q$ even when we use p to symbolize a sentence which happens to be false.

But there is still something odd about "if … then…" statements such as Lewis's

If the moon is made of green cheese, then $2 + 2 = 4$.

and about the inferences

[1]Reading "$\rightarrow$" as "implies" is still widely practised by mathematicians and computer scientists, although philosophers try to avoid the confusions Lewis highlighted by pronouncing it as "only if."

> The moon is not made of green cheese. Therefore, if the moon is made of green cheese, then $2+2=4$.
>
> $2+2=4$. Therefore, if the moon is made of green cheese, then $2+2=4$.

Yet, if "if …then …" were just $\rightarrow$, the sentence would be unproblematically true, and the inferences unproblematically valid.

Another example of concerns the tautology $(A \rightarrow B) \vee (B \rightarrow A)$. This would suggest that if you take two indicative sentences S and T from the newspaper at random, the sentence "If S then T, or if T then S" should be true.

17.3 The Strict Conditional

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

We have:

$$A \strictif B \vDash \neg A \vee B \text{ but:} \tag{17.14}$$

$$\neg A \vee B \nvDash A \strictif B \tag{17.15}$$

$$B \nvDash A \strictif B \tag{17.16}$$

$$\neg A \nvDash A \strictif B \tag{17.17}$$

$$\neg(A \rightarrow B) \nvDash A \wedge \neg B \text{ but:} \tag{17.18}$$

$$A \wedge \neg B \vDash \neg(A \strictif B) \tag{17.19}$$

However, the strict conditional still supports modus ponens:

$$A, A \strictif B \vDash B \tag{17.20}$$

The strict conditional agglomerates:

$$A \strictif B, A \strictif C \vDash A \strictif (B \wedge C) \tag{17.21}$$

Antecedent strengthening holds for the strict conditional:

$$A \strictif B \vDash (A \wedge C) \strictif B \tag{17.22}$$

The strict conditional is also transitive:

$$A \strictif B, B \strictif C \vDash A \strictif C \tag{17.23}$$

Finally, the strict conditional is equivalent to its contrapositive:

$$A \strictif B \vDash \neg B \strictif \neg A \tag{17.24}$$

$$\neg B \strictif \neg A \vDash A \strictif B \tag{17.25}$$

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

$$\Box\neg A \vDash A \strictif B \tag{17.26}$$

$$\Box B \vDash A \strictif B \tag{17.27}$$

$$A \strictif B \vDash \Box(A \strictif B) \tag{17.28}$$

$$\neg(A \strictif B) \vDash \Box\neg(A \strictif B) \tag{17.29}$$

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

17.4 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 17.1. Give *S5*-counterexamples to the entailment relations which do not hold for the strict conditional, i.e., for:

1. $\neg p \nvDash \Box(p \rightarrow q)$

2. $q \nvDash \Box(p \to q)$

3. $\neg\Box(p \to q) \nvDash p \land \neg q$

4. $\nvDash \Box(p \to q) \lor \Box(q \to p)$

Problem 17.2. Show that the valid entailment relations hold for the strict conditional by giving *S5*-proofs of:

1. $\Box(A \to B) \vDash \neg A \lor B$

2. $A \land \neg B \vDash \neg\Box(A \to B)$

3. $A, \Box(A \to B) \vDash B$

4. $\Box(A \to B), \Box(A \to C) \vDash \Box(A \to (B \land C))$

5. $\Box(A \to B) \vDash \Box((A \land C) \to B)$

6. $\Box(A \to B), \Box(B \to C) \vDash \Box(A \to C)$

7. $\Box(A \to B) \vDash \Box(\neg B \to \neg A)$

8. $\Box(\neg B \to \neg A) \vDash \Box(A \to B)$

Problem 17.3. Give proofs in *S5* of:

1. $\Box\neg A \vDash A \prec B$

2. $A \prec B \vDash \Box(A \prec B)$

3. $\neg(A \prec B) \vDash \Box\neg(A \prec B)$

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

CHAPTER 18

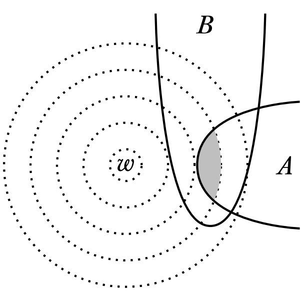
Minimal Change Semantics

18.1 Introduction

Stalnaker and Lewis proposed accounts of counterfactual conditionals such as "If the match were struck, it would light." Their accounts were proposals for how to properly understand the truth conditions for such sentences. The idea behind both proposals is this: to evaluate whether a counterfactual conditional is true, we have to consider those possible worlds which are minimally different from the way the world actually is to make the antecedent true. If the consequent is true in these possible worlds, then the counterfactual is true. For instance, suppose I hold a match and a matchbook in my hand. In the actual world I only look at them and ponder what would happen if I were to strike the match. The minimal change from the actual world where I strike the match is that where I decide to act and strike the match. It is minimal in that nothing else changes: I don't also jump in the air, striking the match doesn't also light my hair on fire, I don't suddenly lose all strength in my fingers, I am not simultaneously doused with water in a SuperSoaker ambush, etc. In that alternative possibility, the match lights. Hence, it's true that if I were to strike the match, it would light.

This intuitive account can be paired with formal semantics for logics of counterfactuals. Lewis introduced the symbol "$\Box\!\!\rightarrow$" for the counterfactual while Stalnaker used the symbol "$>$". We'll use $\Box\!\!\rightarrow$, and add it as a binary connective to propositional logic. So, we have, in addition to formulas of the form $A \rightarrow B$ also formulas of the form $A \mathrel{\Box\!\!\rightarrow} B$. The formal semantics, like the relational semantics for modal logic, is based on models in which formulas are evaluated at worlds, and the satisfaction condition defining $M, w \Vdash A \mathrel{\Box\!\!\rightarrow} B$ 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, $A \mathrel{\Box\!\!\rightarrow} B$ is satisfied at w if B is true at all closest A-worlds.

18.2 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 18.1. A *sphere model* is a triple $M = \langle W, O, V\rangle$ where W is a non-empty set of worlds, $V\colon \mathrm{At}_0 \to \wp(W)$ is a valuation, and $O\colon W \to \wp(\wp(W))$ assigns to each world w a *system of spheres* O_w. For each w, O_w is a set of sets of worlds, and must satisfy:

1. O_w is *centered* on w: $\{w\} \in O_w$.
2. O_w is *nested*: whenever $S_1, S_2 \in O_w$, $S_1 \subseteq S_2$ or $S_2 \subseteq S_1$, i.e., O_w is linearly ordered by $\subseteq$.
3. O_w is closed under non-empty unions.
4. O_w is closed under non-empty intersections.

The intuition behind O_w is that the worlds "around" w are stratified according to how far away they are from w. The innermost sphere is just w by itself, i.e., the set $\{w\}$: w is closer to w than the worlds in any other sphere. If $S \subsetneq S'$, then the worlds in $S' \setminus S$ are further way from w than the worlds in S: $S' \setminus S$ is the "layer" between the S and the worlds outside of S'. In particular, we have to think of the spheres as containing all the worlds within their outer surface; they are not just the individual layers.

The diagram in Figure 18.1 corresponds to the sphere model with $W = \{w, w_1, \ldots, w_7\}$, $V(p) = \{w_5, w_6, w_7\}$. The innermost sphere $S_1 = \{w\}$. The closest worlds to w are w_1, w_2, w_3, so the

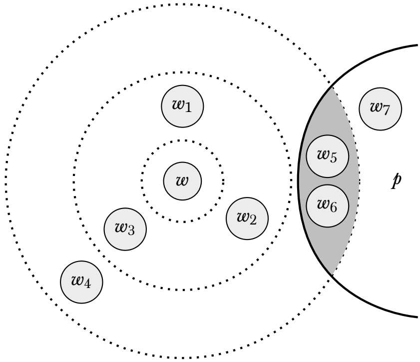


Figure 18.1: Diagram of a sphere model

next larger sphere is $S_2 = \{w, w_1, w_2, w_3\}$. The worlds further out are w_4, w_5, w_6, so the outermost sphere is $S_3 = \{w, w_1, \dots, w_6\}$. 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 $\mathbf{M}$, $\mathbf{M}, w \Vdash A$, we expand the definition for modal formulas to include a clause for $B \mathrel{\Box\!\!\rightarrow} C$:

Definition 18.2. $\mathbf{M}, w \Vdash B \mathrel{\Box\!\!\rightarrow} C$ iff either

1. For all $u \in \bigcup O_w$, $\mathbf{M}, u \nVdash B$, or
2. For some $S \in O_w$,
   a) $\mathbf{M}, u \Vdash B$ for some $u \in S$, and
   b) for all $v \in S$, either $\mathbf{M}, v \nVdash B$ or $\mathbf{M}, v \Vdash C$.

According to this definition, $\mathbf{M}, w \Vdash B \mathrel{\Box\!\!\rightarrow} C$ 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

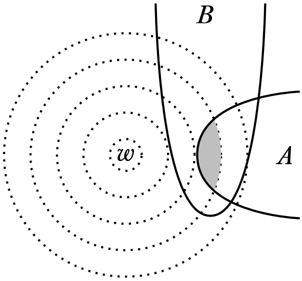


Figure 18.2: Non-vacuously true counterfactual

that we didn't require in the definition that S is the *innermost* B-admitting sphere, contrary to what one might expect from the intuitive explanation. But if the condition in (2) is satisfied for some sphere S, then it is also satisfied for all spheres S contains, and hence in particular for the innermost sphere.

Note also that the definition of sphere models does not require that there *is* an innermost B-admitting sphere: we may have an infinite sequence $S_1 \supsetneq S_2 \supsetneq \cdots \supsetneq \{w\}$ of B-admitting spheres, and hence no innermost B-admitting spheres. In that case, $M, w \Vdash B \mathrel{\Box\!\!\rightarrow} C$ iff $B \rightarrow C$ holds throughout the spheres S_i, S_{i+1}, …, for some i.

18.3 Truth and Falsity of Counterfactuals

A counterfactual $A \mathrel{\Box\!\!\rightarrow} B$ is (non-vacuously) true if the closest A-worlds are all B-worlds, as depicted in Figure 18.2. 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 18.3).

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, $A \mathrel{\Box\!\!\rightarrow} \neg B$ is also false (see Figure 18.4). If the closest A-worlds do not overlap with the B-worlds at all, then $A \mathrel{\Box\!\!\rightarrow} B$. But, in this case

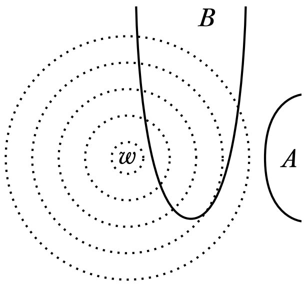


Figure 18.3: Vacuously true counterfactual

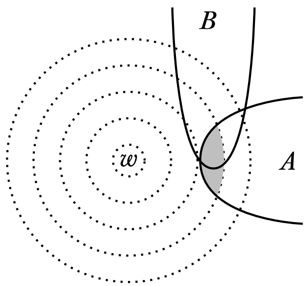


Figure 18.4: False counterfactual, false opposite

all the closest A-worlds are $\neg B$-worlds, and so $A \boxright \neg B$ is true (see Figure 18.5).

In contrast to the strict conditional, counterfactuals may be contingent. Consider the sphere model in Figure 18.6. The A-worlds closest to u are all B-worlds, so $M, u \Vdash A \boxright B$. But there are A-worlds closest to v which are not B-worlds, so $M, v \nVdash A \boxright B$.

18.4 Antecedent Strengthening

"Strengthening the antecedent" refers to the inference $A \to C \vDash (A \land B) \to C$. It is valid for the material conditional, but invalid

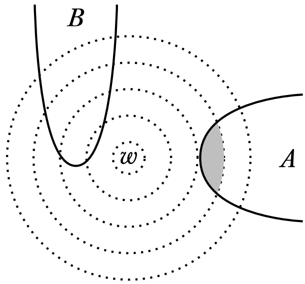


Figure 18.5: False counterfactual, true opposite

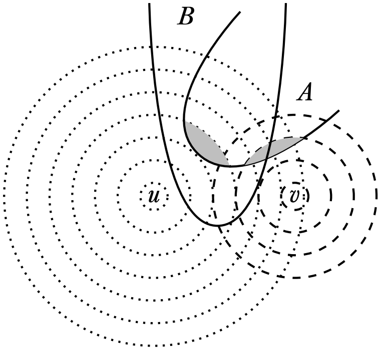


Figure 18.6: Contingent counterfactual

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:

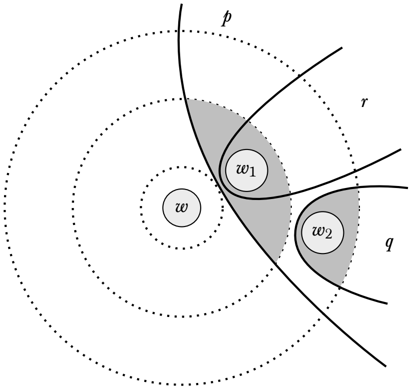


Figure 18.7: Counterexample to antecedent strengthening

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 18.3. The sphere semantics invalidates the inference, i.e., we have $p \mathrel{\Box\!\!\rightarrow} r \nvDash (p \land q) \mathrel{\Box\!\!\rightarrow} r$. 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 $M, w \Vdash p \mathrel{\Box\!\!\rightarrow} r$. There is also a $(p \land q)$-admitting sphere $S' = \{w, w_1, w_2\}$ but $M, w_2 \nVdash (p \land q) \rightarrow r$, so $M, w \nVdash (p \land q) \mathrel{\Box\!\!\rightarrow} r$ (see Figure 18.7).

18.5 Transitivity

For the material conditional, the chain rule holds: $A \rightarrow B, B \rightarrow C \vDash A \rightarrow C$. In other words, the material conditional is transitive. Is the same true for counterfactuals? Consider the following example due to Stalnaker.

> If J. Edgar Hoover had been born a Russian, he would have been a Communist.
>
> If J. Edgar Hoover were a Communist, he would have been be a traitor.
>
> Therefore, If J. Edgar Hoover had been born a Russian, he would have been be a traitor.

If Hoover had been born (at the same time he actually did), not in the United States, but in Russia, he would have grown up in the Soviet Union and become a Communist (let's assume). So the first premise is true. Likewise, the second premise, considered in isolation is true. The conclusion, however, is false: in all likelihood, Hoover would have been a fervent Communist if he had been born in the USSR, and not been a traitor (to his country). The intuitive assignment of truth values is borne out by the Stalnaker–Lewis account. The closest possible world to ours with the only change being Hoover's place of birth is the one where Hoover grows up to be a good citizen of the USSR. This is the closest possible world where the antecedent of the first premise and of the conclusion is true, and in that world Hoover is a loyal member of the Communist party, and so not a traitor. To evaluate the second premise, we have to look at a different world, however: the closest world where Hoover is a Communist, which is one where he was born in the United States, turned, and thus became a traitor.[1]

Example 18.4. The sphere semantics invalidates the inference, i.e., we have $p \mathrel{\Box\!\!\rightarrow} q, q \mathrel{\Box\!\!\rightarrow} r \nvDash p \mathrel{\Box\!\!\rightarrow} r$. 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 $M, w \Vdash p \mathrel{\Box\!\!\rightarrow} q$. There is also

[1] Of course, to appreciate the force of the example we have to take on board some metaphysical and political assumptions, e.g., that it is possible that Hoover could have been born to Russian parents, or that Communists in the US of the 1950s were traitors to their country.

a q-admitting sphere $S' = \{w, w_1\}$ and $M \nVdash q \rightarrow r$ is true at all worlds in it, so $M, w \Vdash q \mathrel{\Box\!\!\rightarrow} r$. However, the p-admitting sphere $\{w, w_1, w_2\}$ contains a world, namely w_2, where $M, w_2 \nVdash p \rightarrow r$.

18.6 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 18.5. The sphere semantics invalidates contraposition, i.e., we have $p \mathrel{\Box\!\!\rightarrow} q \nvDash \neg q \mathrel{\Box\!\!\rightarrow} \neg p$. 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 $M, w \Vdash p \mathrel{\Box\!\!\rightarrow} q$. 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 $M, w_2 \nVdash \neg q \rightarrow \neg p$.

Problems

Problem 18.1. Find a convincing, intuitive example for the failure of transitivity of counterfactuals.

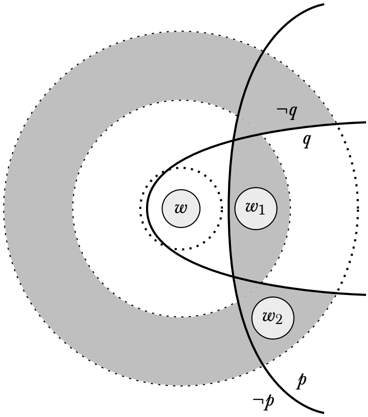


Figure 18.8: Counterexample to contraposition

Problem 18.2. Draw the sphere diagram corresponding to the counterexample in Example 18.4.

Problem 18.3. In Example 18.4, 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 VIII

How can it be true if you can't prove it?

CHAPTER 19

Introduction

19.1 Constructive Reasoning

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

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

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

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

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

Theorem 19.1. *There are irrational numbers a and b such that a^b is rational.*

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

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

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

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

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

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

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

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

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

19.2 Syntax of Intuitionistic Logic

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

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

1. A countably infinite set At_0 of propositional variables p_0, p_1, ...

2. The propositional constant for falsity $\bot$.

3. The logical connectives: $\land$ (conjunction), $\lor$ (disjunction), $\to$ (conditional)

4. Punctuation marks: (,), and the comma.

Definition 19.2 (Formula). The set $\text{Frm}(\mathcal{L}_0)$ of *formulas* of propositional intuitionistic logic is defined inductively as follows:

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

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

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

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

19.3 The Brouwer–Heyting–Kolmogorov Interpretation

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

1. We assume that we know what constitutes a construction of an atomic statement.

2. A construction of $A_1 \wedge A_2$ is a pair $\langle M_1, M_2 \rangle$ where M_1 is a construction of A_1 and M_2 is a construction of A_2.

3. A construction of $A_1 \vee A_2$ is a pair $\langle s, M \rangle$ where s is 1 and M is a construction of A_1, or s is 2 and M is a construction of A_2.

4. A construction of $A \to B$ is a function that converts a construction of A into a construction of B.

5. There is no construction for $\perp$ (absurdity).

6. $\neg A$ is defined as synonym for $A \to \perp$. That is, a construction of $\neg A$ is a function converting a construction of A into a construction of $\perp$.

Example 19.3. Take $\neg\perp$ for example. A construction of it is a function which, given any construction of $\perp$ as input, provides a construction of $\perp$ as output. Obviously, the identity function Id is such a construction: given a construction M of $\perp$, $\mathrm{Id}(M) = M$ yields a construction of $\perp$.

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

Example 19.4. Let us prove $A \to \neg\neg A$ for any proposition A, which is $A \to ((A \to \perp) \to \perp)$. The construction should be a function f that, given a construction M of A, returns a construction $f(M)$ of $(A \to \perp) \to \perp$. Here is how f constructs the construction of $(A \to \perp) \to \perp$: We have to define a function g which, when given a construction h of $A \to \perp$ as input, outputs a construction of $\perp$. We can define g as follows: apply the input h

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

Example 19.5. Let us give a construction for $\neg(A \land \neg A)$, i.e., $(A \land (A \to \bot)) \to \bot$. This is a function f which, given as input a construction M of $A \land (A \to \bot)$, yields a construction of $\bot$. A construction of a conjunction $B_1 \land B_2$ is a pair $\langle N_1, N_2 \rangle$ where N_1 is a construction of B_1 and N_2 is a construction of B_2. We can define functions p_1 and p_2 which recover from a construction of $B_1 \land B_2$ the constructions of B_1 and B_2, respectively:

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

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

Example 19.6. Let us give a construction of $((A \land B) \to C) \to (A \to (B \to C))$, i.e., a function f which turns a construction g of $(A \land B) \to C$ into a construction of $(A \to (B \to C))$. The construction g is itself a function (from constructions of $A \land B$ to constructions of C). And the output $f(g)$ is a function h_g from constructions of A to functions from constructions of B to constructions of C.

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

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

Example 19.7. To prove $\neg\neg(A \vee \neg A)$, we need a function f that transforms a construction of $\neg(A \vee \neg A)$, i.e., of $(A \vee (A \to \bot)) \to \bot$, into a construction of $\bot$. In other words, we need a function f such that $f(g)$ is a construction of $\bot$ if g is a construction of $\neg(A \vee \neg A)$.

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

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

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

As you can see, using the BHK interpretation to show the intuitionistic validity of formulas quickly becomes cumbersome and confusing. Luckily, there are better derivation systems for intuitionistic logic, and more precise semantic interpretations.

19.4 Natural Deduction

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

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

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

Conjunction

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

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

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

A construction of $A_1 \wedge A_1$ on the BHK interpretation is a pair $\langle N_1, N_2 \rangle$. So assume we have such a pair. Then we also have a construction of each conjunct: N_1 is a construction of A_1 and N_2 is a construction of A_2.

Conditional

$$\dfrac{\begin{matrix}[A]^u \\ \vdots \\ B\end{matrix}}{A \to B}\, u \;\to\text{Intro} \qquad\qquad \dfrac{A \to B \qquad A}{B} \;\to\text{Elim}$$

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

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

Disjunction

$$\dfrac{A}{A \vee B} \;\vee\text{Intro} \qquad \dfrac{B}{A \vee B} \;\vee\text{Intro}$$

$$\dfrac{A \vee B \qquad \begin{matrix}[A]^n \\ \vdots \\ C\end{matrix} \qquad \begin{matrix}[B]^n \\ \vdots \\ C\end{matrix}}{C}\, n \;\vee\text{Elim}$$

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

A_1, A_2, respectively, into constructions of C. Then $f_i(N_i)$ is a construction of C, the conclusion of $\vee$Elim.

Absurdity

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

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

Rules for $\neg$

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

$$\begin{array}{c} [A]^n \\ \vdots \\ n\ \dfrac{\bot}{\neg A}\ \neg\text{Intro} \end{array} \qquad \qquad \dfrac{\neg A \qquad A}{\bot}\ \neg\text{Elim}$$

Examples of Derivations

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

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

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

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

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

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

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

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

Proposition 19.8. *If $\Gamma \vdash A$ in intuitionistic logic, $\Gamma \vdash A$ in classical logic. In particular, if A is an intuitionistic theorem, it is also a classical theorem.*

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

19.5 Axiomatic Derivations

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

Definition 19.9 (Derivability). If Γ is a set of formulas of $\mathcal{L}$ then a *derivation* from Γ 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 modus ponens, i.e., $A_k \equiv A_j \rightarrow A_i$.

Definition 19.10 (Axioms). The set of Ax_0 of *axioms* for the intuitionistic propositional logic are all formulas of the following forms:

$$(A \wedge B) \rightarrow A \tag{19.1}$$
$$(A \wedge B) \rightarrow B \tag{19.2}$$
$$A \rightarrow (B \rightarrow (A \wedge B)) \tag{19.3}$$
$$A \rightarrow (A \vee B) \tag{19.4}$$
$$A \rightarrow (B \vee A) \tag{19.5}$$
$$(A \rightarrow C) \rightarrow ((B \rightarrow C) \rightarrow ((A \vee B) \rightarrow C)) \tag{19.6}$$
$$A \rightarrow (B \rightarrow A) \tag{19.7}$$
$$(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C)) \tag{19.8}$$
$$\bot \rightarrow A \tag{19.9}$$

Definition 19.11 (Derivability). A formula A is *derivable* from Γ, written $\Gamma \vdash A$, if there is a derivation from Γ ending in A.

Definition 19.12 (Theorems). A formula A is a *theorem* if there is a derivation of A from the empty set. We write $\vdash A$ if A is a theorem and $\nvdash A$ if it is not.

Proposition 19.13. *If $\Gamma \vdash A$ in intuitionistic logic, $\Gamma \vdash A$ in classical logic. In particular, if A is an intuitionistic theorem, it is also a classical theorem.*

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

Problems

Problem 19.1. Give derivations in intuitionistic logic of the following formulas:

1. $(\neg A \lor B) \to (A \to B)$
2. $\neg\neg\neg A \to \neg A$
3. $\neg\neg(A \land B) \leftrightarrow (\neg\neg A \land \neg\neg B)$
4. $\neg(A \lor B) \leftrightarrow (\neg A \land \neg B)$
5. $(\neg A \lor \neg B) \to \neg(A \land B)$
6. $\neg\neg(A \land B) \to (\neg\neg A \lor \neg\neg B)$

CHAPTER 20

Semantics

20.1 Introduction

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

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

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

If we define "A is known" as in epistemic logic as "true in all epistemic alternatives," then $A \land B$ is known at w if in all epistemic alternatives, both A and B are known. But since knowledge is monotonic and R is reflexive, that means that $A \wedge B$ is known at w iff A and B are known at w. For the same reason, $A \vee B$ is known at w iff at least one of them is known. So for $\wedge$ and $\vee$, the truth conditions of the connectives coincide with those in classical logic.

The truth conditions for the conditional, however, differ from classical logic. $A \rightarrow B$ is known at w iff at no w' with Rww', A is known without B also being known. This is not the same as the condition that A is unknown or B is known at w. For if we know neither A nor B at w, there might be a future epistemic state w' with Rww' such that at w', A is known without also coming to know B.

We know $\neg A$ only if there is no possible future epistemic state in which we know A. Here the idea is that if A were knowable, then in some possible future epistemic state A becomes known. Since we can't know $\bot$, in that future epistemic state, we would know A but not know $\bot$.

On this interpretation the principle of excluded middle fails. For there are some A which we don't yet know, but which we might come to know. For such a formula A, both A and $\neg A$ are unknown, so $A \vee \neg A$ is not known. But we do know, e.g., that $\neg(A \wedge \neg A)$. For no future state in which we know both A and $\neg A$ is possible, and we know this independently of whether or not we know A or $\neg A$.

Relational models are not the only available semantics for intuitionistic logic. The topological semantics is another: here propositions are interpreted as open sets in a topological space, and the connectives are interpreted as operations on these sets (e.g., $\wedge$ corresponds to intersection).

20.2 Relational models

In order to give a precise semantics for intuitionistic propositional logic, we have to give a definition of what counts as a model relative to which we can evaluate formulas. On the basis of such a definition it is then also possible to define semantics notions such as validity and entailment. One such semantics is given by relational models.

Definition 20.1. A relational model for intuitionistic propositional logic is a triple $\mathbf{M} = \langle W, R, V\rangle$, where

1. W is a non-empty set,
2. R is a partial order (i.e., a reflexive, antisymmetric, and transitive binary relation) on W, and
3. V is a function assigning to each propositional variable p a subset of W, such that
4. V is monotone with respect to R, i.e., if $w \in V(p)$ and Rww', then $w' \in V(p)$.

Definition 20.2. We define the notion of A *being true at* w *in* $\mathbf{M}$, $\mathbf{M}, w \Vdash A$, inductively as follows:

1. $A \equiv p$: $\mathbf{M}, w \Vdash A$ iff $w \in V(p)$.
2. $A \equiv \bot$: not $\mathbf{M}, w \Vdash A$.
3. $A \equiv \neg B$: $\mathbf{M}, w \Vdash A$ iff for no w' such that Rww', $\mathbf{M}, w' \Vdash B$.
4. $A \equiv B \wedge C$: $\mathbf{M}, w \Vdash A$ iff $\mathbf{M}, w \Vdash B$ and $\mathbf{M}, w \Vdash C$.
5. $A \equiv B \vee C$: $\mathbf{M}, w \Vdash A$ iff $\mathbf{M}, w \Vdash B$ or $\mathbf{M}, w \Vdash C$ (or both).
6. $A \equiv B \rightarrow C$: $\mathbf{M}, w \Vdash A$ iff for every w' such that Rww', not $\mathbf{M}, w' \Vdash B$ or $\mathbf{M}, w' \Vdash C$ (or both).

We write $\mathbf{M}, w \nVdash A$ if not $\mathbf{M}, w \Vdash A$. If Γ is a set of formulas, $\mathbf{M}, w \Vdash \Gamma$ means $\mathbf{M}, w \Vdash B$ for all $B \in \Gamma$.

Proposition 20.3. *Truth at worlds is monotonic with respect to R, i.e., if $\mathfrak{M}, w \Vdash A$ and Rww', then $\mathfrak{M}, w' \Vdash A$.*

Proof. Exercise. □

20.3 Semantic Notions

Definition 20.4. We say A is *true in the model* $\mathfrak{M} = \langle W, R, V \rangle$, $\mathfrak{M} \Vdash A$, iff $\mathfrak{M}, w \Vdash A$ for all $w \in W$. A is *valid*, $\vDash A$, iff it is true in all models. We say a set of formulas Γ *entails* A, $\Gamma \vDash A$, iff for every model $\mathfrak{M}$ and every w such that $\mathfrak{M}, w \Vdash \Gamma$, $\mathfrak{M}, w \Vdash A$.

Proposition 20.5.

1. *If $\mathfrak{M}, w \Vdash \Gamma$ and $\Gamma \vDash A$, then $\mathfrak{M}, w \Vdash A$.*
2. *If $\mathfrak{M} \Vdash \Gamma$ and $\Gamma \vDash A$, then $\mathfrak{M} \Vdash A$.*

Proof.

1. Suppose $\mathfrak{M} \Vdash \Gamma$. Since $\Gamma \vDash A$, we know that if $\mathfrak{M}, w \Vdash \Gamma$, then $\mathfrak{M}, w \Vdash A$. Since $\mathfrak{M}, u \Vdash \Gamma$ for all every $u \in W$, $\mathfrak{M}, w \Vdash \Gamma$. Hence $\mathfrak{M}, w \Vdash A$.
2. Follows immediately from (1). □

Definition 20.6. Suppose $\mathfrak{M}$ is a relational model and $w \in W$. The *restriction* $\mathfrak{M}_w = \langle W_w, R_w, V_w \rangle$ of $\mathfrak{M}$ to w is given by:

$$
\begin{aligned}
W_w &= \{u \in W : Rwu\}, \\
R_w &= R \cap (W_w)^2, \text{ and} \\
V_w(p) &= V(p) \cap W_w.
\end{aligned}
$$

Proposition 20.7. *$\mathfrak{M}, w \Vdash A$ iff $\mathfrak{M}_w \Vdash A$.*

Proposition 20.8. *Suppose for every model M such that $M \Vdash \Gamma$, $M \Vdash A$. Then $\Gamma \vDash A$.*

Proof. Suppose that $M, w \Vdash \Gamma$. By the Proposition 20.7 applied to every $B \in \Gamma$, we have $M_w \Vdash \Gamma$. By the assumption, we have $M_w \Vdash A$. By Proposition 20.7 again, we get $M, w \Vdash A$. □

20.4 Topological Semantics

Another way to provide a semantics for intuitionistic logic is using the mathematical concept of a topology.

Definition 20.9. Let X be a set. A *topology on X* is a set $\mathcal{O} \subseteq \wp(X)$ that satisfies the properties below. The elements of $\mathcal{O}$ are called the *open sets* of the topology. The set X together with $\mathcal{O}$ is called a *topological space.*

1. The empty set and the entire space are open: $\emptyset$, $X \in \mathcal{O}$.
2. Open sets are closed under finite intersections: if U, $V \in \mathcal{O}$ then $U \cap V \in \mathcal{O}$
3. Open sets are closed under arbitrary unions: if $U_i \in \mathcal{O}$ for all $i \in I$, then $\bigcup\{U_i : i \in I\} \in \mathcal{O}$.

We may write X for a topology if the collection of open sets can be inferred from the context; note that, still, only after X is endowed with open sets can it be called a topology.

Definition 20.10. A *topological model* of intuitionistic propositional logic is a triple $\mathbf{X} = \langle X, \mathcal{O}, V \rangle$ where $\mathcal{O}$ is a topology on X and V is a function assigning an open set in $\mathcal{O}$ to each propositional variable.

Given a topological model $\mathbf{X}$, we can define $[\![A]\!]_{\mathbf{X}}$ inductively as follows:

1. $[\![\bot]\!]_X = \emptyset$
2. $[\![p]\!]_X = V(p)$
3. $[\![A \wedge B]\!]_X = [\![A]\!]_X \cap [\![B]\!]_X$
4. $[\![A \vee B]\!]_X = [\![A]\!]_X \cup [\![B]\!]_X$
5. $[\![A \to B]\!]_X = \mathrm{Int}((X \setminus [\![A]\!]_X) \cup [\![B]\!]_X)$

Here, $\mathrm{Int}(V)$ is the function that maps a set $V \subseteq X$ to its *interior*, that is, the union of all open sets it contains. In other words,

$$\mathrm{Int}(V) = \bigcup \{U : U \subseteq V \text{ and } U \in \mathcal{O}\}.$$

Note that the interior of any set is always open, since it is a union of open sets. Thus, $[\![A]\!]_X$ is always an open set.

Although topological semantics is highly abstract, there are ways to think about it that might motivate it. Suppose that the elements, or "points," of X are points at which statements can be evaluated. The set of all points where A is true is the proposition expressed by A. Not every set of points is a potential proposition; only the elements of $\mathcal{O}$ are. $A \vDash B$ iff B is true at every point at which A is true, i.e., $[\![A]\!]_X \subseteq [\![B]\!]_X$, for all X. The absurd statement $\bot$ is never true, so $[\![\bot]\!]_X = \emptyset$.

How must the propositions expressed by $B \wedge C$, $B \vee C$, and $B \to C$ be related to those expressed by B and C for the intuitionistically valid laws to hold, i.e., so that $A \vdash B$ iff $[\![A]\!]_X \subset [\![B]\!]_X$? We require $\bot \vdash A$ for any A, which is satisfied because $\emptyset \subseteq U$ for all U. Since $B \wedge C \vdash B$, we require that $[\![B \wedge C]\!]_X \subseteq [\![B]\!]_X$, and similarly $[\![B \wedge C]\!]_X \subseteq [\![C]\!]_X$. The largest set satisfying $W \subseteq U$ and $W \subseteq V$ is $U \cap V$. Conversely, $B \vdash B \vee C$ and $C \vdash B \vee C$, and so we require that $[\![B]\!]_X \subseteq [\![B \vee C]\!]_X$ and $[\![C]\!]_X \subseteq [\![B \vee C]\!]_X$. The smallest set W such that $U \subseteq W$ and $V \subseteq W$ is $U \cup V$.

The definition for $\to$ is tricky: $A \to B$ expresses the weakest proposition that, combined with A, entails B. That $A \to B$ combined with A entails B is clear from $(A \to B) \wedge A \vdash B$. So $[\![A \to B]\!]_X$

should be the greatest open set such that $[\![A \to B]\!]_X \cap [\![A]\!]_X \subset [\![B]\!]_X$, leading to our definition.

Problems

Problem 20.1. Show that according to Definition 20.2, $M, w \Vdash \lnot A$ iff $M, w \Vdash A \to \bot$.

Problem 20.2. Prove Proposition 20.3.

Problem 20.3. Prove Proposition 20.7.

PART IX

Wait, hear me out: what if it's both true and false?

CHAPTER 21

Paraconsistent logics

To p and to $\neg p$, that is the question.

PART X

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, ..., 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 $\exists x(x \in A \land \varphi)$.

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* Λ among the strings over A, for every alphabet A. For instance,

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

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^{ω} of infinite sequences of elements of A. An infinite sequence $a_1 a_2 a_3 a_4 \dots$ consists of a one-way infinite list of objects, each one of which is an element of A.

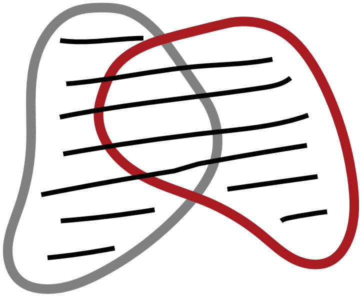


Figure A.1: The union $A \cup B$ of two sets is set of elements of A together with those of B.

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 φ, and this property can mention sets we've already defined. So for instance, if A and B are sets, the set $\{x : x \in A \vee x \in B\}$ consists of all those objects which are elements of either A or B, i.e., it's the set that combines the elements of A and B. We can visualize this as in Figure A.1, where the highlighted area indicates the elements of the two sets A and B together.

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

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

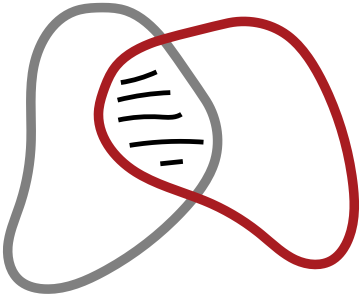


Figure A.2: The intersection $A \cap B$ of two sets is the set of elements they have in common.

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 $A_1, A_2, \ldots$

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

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

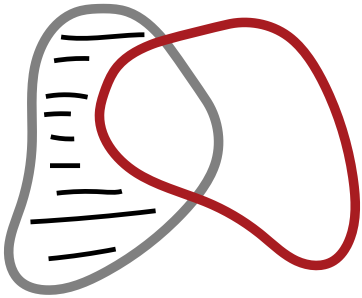


Figure A.3: The difference $A \setminus B$ of two sets is the set of those elements of A which are not also elements of B.

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:

$$A^1 = A$$
$$A^{k+1} = A^k \times A$$

Proposition A.27. *If A has n elements and B has m elements, then $A \times B$ has $n \cdot m$ elements.*

Proof. For every element x in A, there are m elements of the form $\langle x, y\rangle \in A \times B$. Let $B_x = \{\langle x, y\rangle : y \in B\}$. Since whenever $x_1 \neq x_2$, $\langle x_1, y\rangle \neq \langle x_2, y\rangle$, $B_{x_1} \cap B_{x_2} = \emptyset$. But if $A = \{x_1, \dots, x_n\}$, 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:

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

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. □

Example A.28. If A is a set, a *word* over A is any sequence of elements of A. A sequence can be thought of as an n-tuple of elements of A. For instance, if $A = \{a, b, c\}$, then the sequence "bac" can be thought of as the triple $\langle b, a, c\rangle$. Words, i.e., sequences of symbols, are of crucial importance in computer science. By convention, we count elements of A as sequences of length 1, and $\emptyset$ as the sequence of length 0. The set of *all* words over A then is

$$A^* = \{\emptyset\} \cup A \cup A^2 \cup A^3 \cup \dots$$

A.6 Russell's Paradox

Extensionality licenses the notation $\{x : \varphi(x)\}$, for *the* set of x's such that $\varphi(x)$. However, all that extensionality *really* licenses is the following thought. *If* there is a set whose members are all and only the φ's, *then* there is only one such set. Otherwise put: having fixed some φ, the set $\{x : \varphi(x)\}$ is unique, *if it exists.*

But this conditional is important! Crucially, not every property lends itself to *comprehension*. That is, some properties do *not* define sets. If they all did, then we would run into outright contradictions. The most famous example of this is Russell's Paradox.

Sets may be elements of other sets—for instance, the power set of a set A is made up of sets. And so it makes sense to ask or investigate whether a set is an element of another set. Can a set be a member of itself? Nothing about the idea of a set seems to rule this out. For instance, if *all* sets form a collection of objects, one might think that they can be collected into a single set—the set of all sets. And it, being a set, would be an element of the set of all sets.

Russell's Paradox arises when we consider the property of not having itself as an element, of being *non-self-membered*. What if we suppose that there is a set of all sets that do not have themselves as an element? Does

$$R = \{x : x \notin x\}$$

exist? It turns out that we can prove that it does not.

Theorem A.29 (Russell's Paradox). *There is no set* $R = \{x : x \notin x\}$.

Proof. If $R = \{x : x \notin x\}$ exists, then $R \in R$ iff $R \notin R$, which is a contradiction. □

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 $\mathrm{Id}_A = \{\langle x, x\rangle : x \in A\}$ 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., Lnm iff $n < m$. The subset of pairs below the diagonal, i.e.,

$$G = \{\langle 1,0\rangle, \langle 2,0\rangle, \langle 2,1\rangle, \langle 3,0\rangle, \langle 3,1\rangle, \langle 3,2\rangle, \dots\},$$

is the *greater than* relation, i.e., Gnm iff $n > m$. The union of L with I, which we might call $K = L \cup I$, is the *less than or equal to* relation: Knm iff $n \leq m$. Similarly, $H = G \cup I$ is the *greater than or equal to relation*. These relations L, G, K, and H are special kinds of relations called *orders*. L and G have the property that no number bears L or G to itself (i.e., for all n, neither Lnn nor Gnn). Relations with this property are called *irreflexive*, and, if they also happen to be orders, they are called *strict orders*.

Although orders and identity are important and natural relations, it should be emphasized that according to our definition *any* subset of A^2 is a relation on A, regardless of how unnatural or contrived it seems. In particular, $\emptyset$ is a relation on any set (the *empty relation*, which no pair of elements bears), and A^2 itself is a relation on A as well (one which every pair bears), called the *universal relation*. But also something like $E = \{\langle n,m\rangle : n > 5 \text{ or } m \times n \geq 34\}$ counts as a relation.

B.2 Special Properties of Relations

Some kinds of relations turn out to be so common that they have been given special names. For instance, $\leq$ and $\subseteq$ both relate their respective domains (say, $\mathbb{N}$ in the case of $\leq$ and $\wp(A)$ in the case of $\subseteq$) in similar ways. To get at exactly how these relations are similar, and how they differ, we categorize them according to some special properties that relations can have. It turns out that (combinations of) some of these special properties are especially important: orders and equivalence relations.

Definition B.3 (Reflexivity). A relation $R \subseteq A^2$ is *reflexive* iff, for every $x \in A$, Rxx.

Definition B.4 (Transitivity). A relation $R \subseteq A^2$ is *transitive* iff, whenever Rxy and Ryz, then also Rxz.

Definition B.5 (Symmetry). A relation $R \subseteq A^2$ is *symmetric* iff, whenever Rxy, then also Ryx.

Definition B.6 (Anti-symmetry). A relation $R \subseteq A^2$ is *anti-symmetric* iff, whenever both Rxy and Ryx, then $x = y$ (or, in other words: if $x \neq y$ then either $\neg Rxy$ or $\neg Ryx$).

In a symmetric relation, Rxy and Ryx always hold together, or neither holds. In an anti-symmetric relation, the only way for Rxy and Ryx to hold together is if $x = y$. Note that this does not *require* that Rxy and Ryx holds when $x = y$, only that it isn't ruled out. So an anti-symmetric relation can be reflexive, but it is not the case that every anti-symmetric relation is reflexive. Also note that being anti-symmetric and merely not being symmetric are different conditions. In fact, a relation can be both symmetric and anti-symmetric at the same time (e.g., the identity relation is).

Definition B.7 (Connectivity). A relation $R \subseteq A^2$ is *connected* if for all $x, y \in A$, 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. □

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., $[1]_{\equiv_n}$; $\dots$; 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 $\preccurlyeq$ on $\mathbb{B}^*$: $x \preccurlyeq y$ iff $\text{len}(x) \leq \text{len}(y)$. This is a preorder (reflexive and transitive), and even connected, but not a partial order, since it is not anti-symmetric. For instance, $01 \preccurlyeq 10$ and $10 \preccurlyeq 01$, but $01 \neq 10$.

Example B.19. An important partial order is the relation $\subseteq$ on a set of sets. This is not in general a linear order, since if $a \neq b$ and we consider $\wp(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$, we see that $\{a\} \nsubseteq \{b\}$ and $\{a\} \neq \{b\}$ and $\{b\} \nsubseteq \{a\}$.

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 Id_A, i.e., adding all the pairs $\langle x, x\rangle$. (This is called the *reflexive closure* of R.) Conversely, starting from a partial order, one can get a strict order by removing Id_A. These next two results make this precise.

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

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

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

Proposition B.25. *If R is a partial order on A, then $R^- = R \setminus \mathrm{Id}_A$ 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. □

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. □

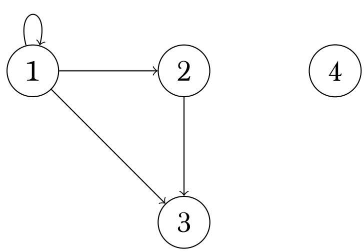
B.5 Graphs

A *graph* is a diagram in which points—called "nodes" or "vertices" (plural of "vertex")—are connected by edges. Graphs are a ubiquitous tool in discrete mathematics and in computer science. They are incredibly useful for representing, and visualizing, relationships and structures, from concrete things like networks of various kinds to abstract structures such as the possible outcomes of decisions. There are many different kinds of graphs in the literature which differ, e.g., according to whether the edges are directed or not, have labels or not, whether there can be edges from a node to the same node, multiple edges between the same nodes, etc. *Directed graphs* have a special connection to relations.

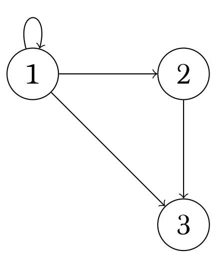
Definition B.27 (Directed graph). A *directed graph* $G = \langle V, E\rangle$ is a set of *vertices* V and a set of *edges* $E \subseteq V^2$.

According to our definition, a graph just is a set together with a relation on that set. Of course, when talking about graphs, it's only natural to expect that they are graphically represented: we can draw a graph by connecting two vertices v_1 and v_2 by an arrow iff $\langle v_1, v_2\rangle \in E$. The only difference between a relation by itself and a graph is that a graph specifies the set of vertices, i.e., a graph may have isolated vertices. The important point, however, is that every relation R on a set X can be seen as a directed graph $\langle X, R\rangle$, and conversely, a directed graph $\langle V, E\rangle$ can be seen as a relation $E \subseteq V^2$ with the set V explicitly specified.

Example B.28. The graph $\langle V, E\rangle$ with $V = \{1, 2, 3, 4\}$ and $E = \{\langle 1, 1\rangle, \langle 1, 2\rangle, \langle 1, 3\rangle, \langle 2, 3\rangle\}$ looks like this:



This is a different graph than $\langle V', E\rangle$ with $V' = \{1, 2, 3\}$, which looks like this:



B.6 Operations on Relations

It is often useful to modify or combine relations. In Proposition B.24, we considered the *union* of relations, which is just the union of two relations considered as sets of pairs. Similarly, in Proposition B.25, we considered the relative difference of relations. Here are some other operations we can perform on relations.

Definition B.29. Let R, S be relations, and A be any set.

The *inverse* of R is $R^{-1} = \{\langle y, x\rangle : \langle x, y\rangle \in R\}$.

The *relative product* of R and S is $(R \mid S) = \{\langle x, z\rangle : \exists y(Rxy \land Syz)\}$.

The *restriction* of R to A is $R{\restriction}_A = R \cap A^2$.

The *application* of R to A is $R[A] = \{y : (\exists x \in A)Rxy\}$

Example B.30. Let $S \subseteq \mathbb{Z}^2$ be the successor relation on $\mathbb{Z}$, i.e., $S = \{\langle x, y\rangle \in \mathbb{Z}^2 : x + 1 = y\}$, so that Sxy iff $x + 1 = y$.

S^{-1} is the predecessor relation on $\mathbb{Z}$, i.e., $\{\langle x, y\rangle \in \mathbb{Z}^2 : x - 1 = y\}$.

$S \mid S$ is $\{\langle x, y\rangle \in \mathbb{Z}^2 : x + 2 = y\}$

$S{\restriction}_{\mathbb{N}}$ is the successor relation on $\mathbb{N}$.

$S[\{1, 2, 3\}]$ is $\{2, 3, 4\}$.

Definition B.31 (Transitive closure). Let $R \subseteq A^2$ be a binary relation.

The *transitive closure* of R is $R^+ = \bigcup_{0<n\in\mathbb{N}} R^n$, where we recursively define $R^1 = R$ and $R^{n+1} = R^n \mid R$.

The *reflexive transitive closure* of R is $R^* = R^+ \cup \mathrm{Id}_A$.

Example B.32. Take the successor relation $S \subseteq \mathbb{Z}^2$. S^2xy iff $x + 2 = y$, S^3xy iff $x + 3 = y$, etc. So S^+xy iff $x + n = y$ for some $n \geq 1$. In other words, S^+xy iff $x < y$, and S^*xy iff $x \leq y$.

Problems

Problem B.1. List the elements of the relation $\subseteq$ on the set $\wp(\{a,b,c\})$.

Problem B.2. Give examples of relations that are (a) reflexive and symmetric but not transitive, (b) reflexive and anti-symmetric, (c) anti-symmetric, transitive, but not reflexive, and (d) reflexive, symmetric, and transitive. Do not use relations on numbers or sets.

Problem B.3. Show that $\equiv_n$ is an equivalence relation, for any $n \in \mathbb{Z}^+$, and that $\mathbb{N}/_{\equiv_n}$ has exactly n members.

Problem B.4. Give a proof of Proposition B.25.

Problem B.5. Consider the less-than-or-equal-to relation $\leq$ on the set $\{1,2,3,4\}$ as a graph and draw the corresponding diagram.

Problem B.6. Show that the transitive closure of R is in fact transitive.

APPENDIX C

Proofs

C.1 Introduction

Based on your experiences in introductory logic, you might be comfortable with a derivation system—probably a natural deduction or Fitch style derivation system, or perhaps a proof-tree system. You probably remember doing proofs in these systems, either proving a formula or show that a given argument is valid. In order to do this, you applied the rules of the system until you got the desired end result. In reasoning *about* logic, we also prove things, but in most cases we are not using a derivation system. In fact, most of the proofs we consider are done in English (perhaps, with some symbolic language thrown in) rather than entirely in the language of first-order logic. When constructing such proofs, you might at first be at a loss—how do I prove something without a derivation system? How do I start? How do I know if my proof is correct?

Before attempting a proof, it's important to know what a proof is and how to construct one. As implied by the name, a *proof* is meant to show that something is true. You might think of this in terms of a dialogue—someone asks you if something is true, say, if every prime other than two is an odd number. To answer "yes" is not enough; they might want to know *why*. In this case, you'd give them a proof.

In everyday discourse, it might be enough to gesture at an answer, or give an incomplete answer. In logic and mathematics, however, we want rigorous proof—we want to show that something is true beyond *any* doubt. This means that every step in our proof must be justified, and the justification must be cogent (i.e., the assumption you're using is actually assumed in the statement of the theorem you're proving, the definitions you apply must be correctly applied, the justifications appealed to must be correct inferences, etc.).

Usually, we're proving some statement. We call the statements we're proving by various names: propositions, theorems, lemmas, or corollaries. A proposition is a basic proof-worthy statement: important enough to record, but perhaps not particularly deep nor applied often. A theorem is a significant, important proposition. Its proof often is broken into several steps, and sometimes it is named after the person who first proved it (e.g., Cantor's Theorem, the Löwenheim–Skolem theorem) or after the fact it concerns (e.g., the completeness theorem). A lemma is a proposition or theorem that is used in the proof of a more important result. Confusingly, sometimes lemmas are important results in themselves, and also named after the person who introduced them (e.g., Zorn's Lemma). A corollary is a result that easily follows from another one.

A statement to be proved often contains assumptions that clarify which kinds of things we're proving something about. It might begin with "Let A be a formula of the form $B \rightarrow C$" or "Suppose $\Gamma \vdash A$" or something of the sort. These are *hypotheses* of the proposition, theorem, or lemma, and you may assume these to be true in your proof. They restrict what we're proving, and also introduce some names for the objects we're talking about. For instance, if your proposition begins with "Let A be a formula of the form $B \rightarrow C$," you're proving something about all formulas of a certain sort only (namely, conditionals), and it's understood that $B \rightarrow C$ is an arbitrary conditional that your proof will talk about.

C.2 Starting a Proof

But where do you even start?

You've been given something to prove, so this should be the last thing that is mentioned in the proof (you can, obviously, *announce* that you're going to prove it at the beginning, but you don't want to use it as an assumption). Write what you are trying to prove at the bottom of a fresh sheet of paper—this way you don't lose sight of your goal.

Next, you may have some assumptions that you are able to use (this will be made clearer when we talk about the *type* of proof you are doing in the next section). Write these at the top of the page and make sure to flag that they are assumptions (i.e., if you are assuming p, write "assume that p," or "suppose that p"). Finally, there might be some definitions in the question that you need to know. You might be told to use a specific definition, or there might be various definitions in the assumptions or conclusion that you are working towards. *Write these down and ensure that you understand what they mean.*

How you set up your proof will also be dependent upon the form of the question. The next section provides details on how to set up your proof based on the type of sentence.

C.3 Using Definitions

We mentioned that you must be familiar with all definitions that may be used in the proof, and that you can properly apply them. This is a really important point, and it is worth looking at in a bit more detail. Definitions are used to abbreviate properties and relations so we can talk about them more succinctly. The introduced abbreviation is called the *definiendum*, and what it abbreviates is the *definiens*. In proofs, we often have to go back to how the definiendum was introduced, because we have to exploit the logical structure of the definiens (the long version of which the defined term is the abbreviation) to get through our proof. By unpacking definitions, you're ensuring that you're getting to the heart of where the logical action is.

We'll start with an example. Suppose you want to prove the following:

Proposition C.1. *For any sets A and B, $A \cup B = B \cup A$.*

In order to even start the proof, we need to know what it means for two sets to be identical; i.e., we need to know what the "=" in that equation means for sets. Sets are defined to be identical whenever they have the same elements. So the definition we have to unpack is:

Definition C.2. Sets A and B are *identical*, $A = B$, iff every element of A is an element of B, and vice versa.

This definition uses A and B as placeholders for arbitrary sets. What it defines—the *definiendum*—is the expression "$A = B$" by giving the condition under which $A = B$ is true. This condition—"every element of A is an element of B, and vice versa"—is the *definiens.*[1] The definition specifies that $A = B$ is true if, and only if (we abbreviate this to "iff") the condition holds.

When you apply the definition, you have to match the A and B in the definition to the case you're dealing with. In our case, it means that in order for $A \cup B = B \cup A$ to be true, each $z \in A \cup B$ must also be in $B \cup A$, and vice versa. The expression $A \cup B$ in the proposition plays the role of A in the definition, and $B \cup A$ that of B. Since A and B are used both in the definition and in the statement of the proposition we're proving, but in different uses, you have to be careful to make sure you don't mix up the two. For instance, it would be a mistake to think that you could prove the proposition by showing that every element of A is an element

[1] In this particular case—and very confusingly!—when $A = B$, the sets A and B are just one and the same set, even though we use different letters for it on the left and the right side. But the ways in which that set is picked out may be different, and that makes the definition non-trivial.

of B, and vice versa—that would show that $A = B$, not that $A \cup B = B \cup A$. (Also, since A and B may be any two sets, you won't get very far, because if nothing is assumed about A and B they may well be different sets.)

Within the proof we are dealing with set-theoretic notions such as union, and so we must also know the meanings of the symbol $\cup$ in order to understand how the proof should proceed. And sometimes, unpacking the definition gives rise to further definitions to unpack. For instance, $A \cup B$ is defined as $\{z : z \in A$ or $z \in B\}$. So if you want to prove that $x \in A \cup B$, unpacking the definition of $\cup$ tells you that you have to prove $x \in \{z : z \in A$ or $z \in B\}$. Now you also have to remember that $x \in \{z : \dots z \dots\}$ iff $\dots x \dots$. So, further unpacking the definition of the $\{z : \dots z \dots\}$ notation, what you have to show is: $x \in A$ or $x \in B$. So, "every element of $A \cup B$ is also an element of $B \cup A$" really means: "for every x, if $x \in A$ or $x \in B$, then $x \in B$ or $x \in A$." If we fully unpack the definitions in the proposition, we see that what we have to show is this:

Proposition C.3. *For any sets A and B: (a) for every x, if $x \in A$ or $x \in B$, then $x \in B$ or $x \in A$, and (b) for every x, if $x \in B$ or $x \in A$, then $x \in A$ or $x \in B$.*

What's important is that unpacking definitions is a necessary part of constructing a proof. Properly doing it is sometimes difficult: you must be careful to distinguish and match the variables in the definition and the terms in the claim you're proving. In order to be successful, you must know what the question is asking and what all the terms used in the question mean—you will often need to unpack more than one definition. In simple proofs such as the ones below, the solution follows almost immediately from the definitions themselves. Of course, it won't always be this simple.

C.4 Inference Patterns

Proofs are composed of individual inferences. When we make an inference, we typically indicate that by using a word like "so," "thus," or "therefore." The inference often relies on one or two facts we already have available in our proof—it may be something we have assumed, or something that we've concluded by an inference already. To be clear, we may label these things, and in the inference we indicate what other statements we're using in the inference. An inference will often also contain an explanation of *why* our new conclusion follows from the things that come before it. There are some common patterns of inference that are used very often in proofs; we'll go through some below. Some patterns of inference, like proofs by induction, are more involved (and will be discussed later).

We've already discussed one pattern of inference: unpacking, or applying, a definition. When we unpack a definition, we just restate something that involves the definiendum by using the definiens. For instance, suppose that we have already established in the course of a proof that $D = E$ (a). Then we may apply the definition of $=$ for sets and infer: "Thus, by definition from (a), every element of D is an element of E and vice versa."

Somewhat confusingly, we often do not write the justification of an inference when we actually make it, but before. Suppose we haven't already proved that $D = E$, but we want to. If $D = E$ is the conclusion we aim for, then we can restate this aim also by applying the definition: to prove $D = E$ we have to prove that every element of D is an element of E and vice versa. So our proof will have the form: (a) prove that every element of D is an element of E; (b) every element of E is an element of D; (c) therefore, from (a) and (b) by definition of $=$, $D = E$. But we would usually not write it this way. Instead we might write something like,

> We want to show $D = E$. By definition of $=$, this amounts to showing that every element of D is an el-

ement of E and vice versa.

(a) …(a proof that every element of D is an element of E) …

(b) …(a proof that every element of E is an element of D) …

Using a Conjunction

Perhaps the simplest inference pattern is that of drawing as conclusion one of the conjuncts of a conjunction. In other words: if we have assumed or already proved that p and q, then we're entitled to infer that p (and also that q). This is such a basic inference that it is often not mentioned. For instance, once we've unpacked the definition of $D = E$ we've established that every element of D is an element of E and vice versa. From this we can conclude that every element of E is an element of D (that's the "vice versa" part).

Proving a Conjunction

Sometimes what you'll be asked to prove will have the form of a conjunction; you will be asked to "prove p and q." In this case, you simply have to do two things: prove p, and then prove q. You could divide your proof into two sections, and for clarity, label them. When you're making your first notes, you might write "(1) Prove p" at the top of the page, and "(2) Prove q" in the middle of the page. (Of course, you might not be explicitly asked to prove a conjunction but find that your proof requires that you prove a conjunction. For instance, if you're asked to prove that $D = E$ you will find that, after unpacking the definition of $=$, you have to prove: every element of D is an element of E *and* every element of E is an element of D).

Proving a Disjunction

When what you are proving takes the form of a disjunction (i.e., it is an statement of the form "p or q"), it is enough to show that one of the disjuncts is true. However, it basically never happens that either disjunct just follows from the assumptions of your theorem. More often, the assumptions of your theorem are themselves disjunctive, or you're showing that all things of a certain kind have one of two properties, but some of the things have the one and others have the other property. This is where proof by cases is useful (see below).

Conditional Proof

Many theorems you will encounter are in conditional form (i.e., show that if p holds, then q is also true). These cases are nice and easy to set up—simply assume the antecedent of the conditional (in this case, p) and prove the conclusion q from it. So if your theorem reads, "If p then q," you start your proof with "assume p" and at the end you should have proved q.

Conditionals may be stated in different ways. So instead of "If p then q," a theorem may state that "p only if q," "q if p," or "q, provided p." These all mean the same and require assuming p and proving q from that assumption. Recall that a biconditional ("p if and only if (iff) q") is really two conditionals put together: if p then q, and if q then p. All you have to do, then, is two instances of conditional proof: one for the first conditional and another one for the second. Sometimes, however, it is possible to prove an "iff" statement by chaining together a bunch of other "iff" statements so that you start with "p" an end with "q"—but in that case you have to make sure that each step really is an "iff."

Universal Claims

Using a universal claim is simple: if something is true for anything, it's true for each particular thing. So if, say, the hypothesis of your proof is $A \subseteq B$, that means (unpacking the definition of $\subseteq$), that, for every $x \in A$, $x \in B$. Thus, if you already know that $z \in A$, you can conclude $z \in B$.

Proving a universal claim may seem a little bit tricky. Usually these statements take the following form: "If x has P, then it has Q" or "All Ps are Qs." Of course, it might not fit this form perfectly, and it takes a bit of practice to figure out what you're asked to prove exactly. But: we often have to prove that all objects with some property have a certain other property.

The way to prove a universal claim is to introduce names or variables, for the things that have the one property and then show that they also have the other property. We might put this by saying that to prove something for *all* Ps you have to prove it for an *arbitrary* P. And the name introduced is a name for an arbitrary P. We typically use single letters as these names for arbitrary things, and the letters usually follow conventions: e.g., we use n for natural numbers, A for formulas, A for sets, f for functions, etc.

The trick is to maintain generality throughout the proof. You start by assuming that an arbitrary object ("x") has the property P, and show (based only on definitions or what you are allowed to assume) that x has the property Q. Because you have not stipulated what x is specifically, other that it has the property P, then you can assert that everything with P has the property Q. In short, x is a stand-in for *all* things with property P.

Proposition C.4. *For all sets A and B, $A \subseteq A \cup B$.*

Proof. Let A and B be arbitrary sets. We want to show that $A \subseteq A \cup B$. By definition of $\subseteq$, this amounts to: for every x, if $x \in A$ then $x \in A \cup B$. So let $x \in A$ be an arbitrary element of A. We have to show that $x \in A \cup B$. Since $x \in A$, $x \in A$ or $x \in B$. Thus, $x \in \{x : x \in A \lor x \in B\}$. But that, by definition of $\cup$, means $x \in A \cup B$. □

Proof by Cases

Suppose you have a disjunction as an assumption or as an already established conclusion—you have assumed or proved that p or q is true. You want to prove r. You do this in two steps: first you assume that p is true, and prove r, then you assume that q is true and prove r again. This works because we assume or know that one of the two alternatives holds. The two steps establish that either one is sufficient for the truth of r. (If both are true, we have not one but two reasons for why r is true. It is not necessary to separately prove that r is true assuming both p and q.) To indicate what we're doing, we announce that we "distinguish cases." For instance, suppose we know that $x \in B \cup C$. $B \cup C$ is defined as $\{x : x \in B \text{ or } x \in C\}$. In other words, by definition, $x \in B$ or $x \in C$. We would prove that $x \in A$ from this by first assuming that $x \in B$, and proving $x \in A$ from this assumption, and then assume $x \in C$, and again prove $x \in A$ from this. You would write "We distinguish cases" under the assumption, then "Case (1): $x \in B$" underneath, and "Case (2): $x \in C$ halfway down the page. Then you'd proceed to fill in the top half and the bottom half of the page.

Proof by cases is especially useful if what you're proving is itself disjunctive. Here's a simple example:

Proposition C.5. *Suppose $B \subseteq D$ and $C \subseteq E$. Then $B \cup C \subseteq D \cup E$.*

Proof. Assume (a) that $B \subseteq D$ and (b) $C \subseteq E$. By definition, any $x \in B$ is also $\in D$ (c) and any $x \in C$ is also $\in E$ (d). To show that $B \cup C \subseteq D \cup E$, we have to show that if $x \in B \cup C$ then $x \in D \cup E$ (by definition of $\subseteq$). $x \in B \cup C$ iff $x \in B$ or $x \in C$ (by definition of $\cup$). Similarly, $x \in D \cup E$ iff $x \in D$ or $x \in E$. So, we have to show: for any x, if $x \in B$ or $x \in C$, then $x \in D$ or $x \in E$.

> So far we've only unpacked definitions! We've reformulated our proposition without $\subseteq$ and $\cup$ and are left with trying to prove a universal conditional claim. By what we've discussed above, this is done by assuming

> that x is something about which we assume the "if" part is true, and we'll go on to show that the "then" part is true as well. In other words, we'll assume that $x \in B$ or $x \in C$ and show that $x \in D$ or $x \in E$.[2]

Suppose that $x \in B$ or $x \in C$. We have to show that $x \in D$ or $x \in E$. We distinguish cases.

Case 1: $x \in B$. By (c), $x \in D$. Thus, $x \in D$ or $x \in E$. (Here we've made the inference discussed in the preceding subsection!)

Case 2: $x \in C$. By (d), $x \in E$. Thus, $x \in D$ or $x \in E$. □

Proving an Existence Claim

When asked to prove an existence claim, the question will usually be of the form "prove that there is an x such that $\ldots x \ldots$", i.e., that some object that has the property described by "$\ldots x \ldots$". In this case you'll have to identify a suitable object show that is has the required property. This sounds straightforward, but a proof of this kind can be tricky. Typically it involves *constructing* or *defining* an object and proving that the object so defined has the required property. Finding the right object may be hard, proving that it has the required property may be hard, and sometimes it's even tricky to show that you've succeeded in defining an object at all!

Generally, you'd write this out by specifying the object, e.g., "let x be …" (where … specifies which object you have in mind), possibly proving that … in fact describes an object that exists, and then go on to show that x has the property Q. Here's a simple example.

Proposition C.6. *Suppose that $x \in B$. Then there is an A such that $A \subseteq B$ and $A \neq \emptyset$.*

Proof. Assume $x \in B$. Let $A = \{x\}$.

---

[2]This paragraph just explains what we're doing—it's not part of the proof, and you don't have to go into all this detail when you write down your own proofs.

> Here we've defined the set A by enumerating its elements. Since we assume that x is an object, and we can always form a set by enumerating its elements, we don't have to show that we've succeeded in defining a set A here. However, we still have to show that A has the properties required by the proposition. The proof isn't complete without that!

Since $x \in A$, $A \neq \emptyset$.

> This relies on the definition of A as $\{x\}$ and the obvious facts that $x \in \{x\}$ and $x \notin \emptyset$.

Since x is the only element of $\{x\}$, and $x \in B$, every element of A is also an element of B. By definition of $\subseteq$, $A \subseteq B$. □

Using Existence Claims

Suppose you know that some existence claim is true (you've proved it, or it's a hypothesis you can use), say, "for some x, $x \in A$" or "there is an $x \in A$." If you want to use it in your proof, you can just pretend that you have a name for one of the things which your hypothesis says exist. Since A contains at least one thing, there are things to which that name might refer. You might of course not be able to pick one out or describe it further (other than that it is $\in A$). But for the purpose of the proof, you can pretend that you have picked it out and give a name to it. It's important to pick a name that you haven't already used (or that appears in your hypotheses), otherwise things can go wrong. In your proof, you indicate this by going from "for some x, $x \in A$" to "Let $a \in A$." Now you can reason about a, use some other hypotheses, etc., until you come to a conclusion, p. If p no longer mentions a, p is independent of the asusmption that $a \in A$, and you've shown that it follows just from the assumption "for some x, $x \in A$."

Proposition C.7. *If $A \neq \emptyset$, then $A \cup B \neq \emptyset$.*

Proof. Suppose $A \neq \emptyset$. So for some x, $x \in A$.

> Here we first just restated the hypothesis of the proposition. This hypothesis, i.e., $A \neq \emptyset$, hides an existential claim, which you get to only by unpacking a few definitions. The definition of $=$ tells us that $A = \emptyset$ iff every $x \in A$ is also $\in \emptyset$ and every $x \in \emptyset$ is also $\in A$. Negating both sides, we get: $A \neq \emptyset$ iff either some $x \in A$ is $\notin \emptyset$ or some $x \in \emptyset$ is $\notin A$. Since nothing is $\in \emptyset$, the second disjunct can never be true, and "$x \in A$ and $x \notin \emptyset$" reduces to just $x \in A$. So $x \neq \emptyset$ iff for some x, $x \in A$. That's an existence claim. Now we use that existence claim by introducing a name for one of the elements of A:

Let $a \in A$.

> Now we've introduced a name for one of the things $\in A$. We'll continue to argue about a, but we'll be careful to only assume that $a \in A$ and nothing else:

Since $a \in A$, $a \in A \cup B$, by definition of $\cup$. So for some x, $x \in A \cup B$, i.e., $A \cup B \neq \emptyset$.

> In that last step, we went from "$a \in A \cup B$" to "for some x, $x \in A \cup B$." That doesn't mention a anymore, so we know that "for some x, $x \in A \cup B$" follows from "for some x, $x \in A$ alone." But that means that $A \cup B \neq \emptyset$. □

It's maybe good practice to keep bound variables like "x" separate from hypothetical names like a, like we did. In practice, however, we often don't and just use x, like so:

> Suppose $A \neq \emptyset$, i.e., there is an $x \in A$. By definition of $\cup$, $x \in A \cup B$. So $A \cup B \neq \emptyset$.

However, when you do this, you have to be extra careful that you use different x's and y's for different existential claims. For instance, the following is *not* a correct proof of "If $A \neq \emptyset$ and $B \neq \emptyset$ then $A \cap B \neq \emptyset$" (which is not true).

> Suppose $A \neq \emptyset$ and $B \neq \emptyset$. So for some x, $x \in A$ and also for some x, $x \in B$. Since $x \in A$ and $x \in B$, $x \in A \cap B$, by definition of $\cap$. So $A \cap B \neq \emptyset$.

Can you spot where the incorrect step occurs and explain why the result does not hold?

C.5 An Example

Our first example is the following simple fact about unions and intersections of sets. It will illustrate unpacking definitions, proofs of conjunctions, of universal claims, and proof by cases.

Proposition C.8. *For any sets A, B, and C, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$*

Let's prove it!

Proof. We want to show that for any sets A, B, and C, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

> First we unpack the definition of "=" in the statement of the proposition. Recall that proving sets identical means showing that the sets have the same elements. That is, all elements of $A \cup (B \cap C)$ are also elements of $(A \cup B) \cap (A \cup C)$, and vice versa. The "vice versa" means that also every element of $(A \cup B) \cap (A \cup C)$ must be an element of $A \cup (B \cap C)$. So in unpacking the definition, we see that we have to prove a conjunction. Let's record this:

By definition, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ iff every element of $A \cup (B \cap C)$ is also an element of $(A \cup B) \cap (A \cup C)$, and every element of $(A \cup B) \cap (A \cup C)$ is an element of $A \cup (B \cap C)$.

> Since this is a conjunction, we must prove each conjunct separately. Lets start with the first: let's prove that every element of $A \cup (B \cap C)$ is also an element of $(A \cup B) \cap (A \cup C)$.
>
> This is a universal claim, and so we consider an arbitrary element of $A \cup (B \cap C)$ and show that it must also be an element of $(A \cup B) \cap (A \cup C)$. We'll pick a variable to call this arbitrary element by, say, z. Our proof continues:

First, we prove that every element of $A \cup (B \cap C)$ is also an element of $(A \cup B) \cap (A \cup C)$. Let $z \in A \cup (B \cap C)$. We have to show that $z \in (A \cup B) \cap (A \cup C)$.

> Now it is time to unpack the definition of $\cup$ and $\cap$. For instance, the definition of $\cup$ is: $A \cup B = \{z : z \in A \text{ or } z \in B\}$. When we apply the definition to "$A \cup (B \cap C)$," the role of the "B" in the definition is now played by "$B \cap C$," so $A \cup (B \cap C) = \{z : z \in A \text{ or } z \in B \cap C\}$. So our assumption that $z \in A \cup (B \cap C)$ amounts to: $z \in \{z : z \in A \text{ or } z \in B \cap C\}$. And $z \in \{z : \dots z \dots\}$ iff $\dots z \dots$, i.e., in this case, $z \in A$ or $z \in B \cap C$.

By the definition of $\cup$, either $z \in A$ or $z \in B \cap C$.

> Since this is a disjunction, it will be useful to apply proof by cases. We take the two cases, and show that in each one, the conclusion we're aiming for (namely, "$z \in (A \cup B) \cap (A \cup C)$") obtains.

Case 1: Suppose that $z \in A$.

> There's not much more to work from based on our assumptions. So let's look at what we have to work with in the conclusion. We want to show that $z \in (A \cup B) \cap (A \cup C)$. Based on the definition of $\cap$, if we want to show that $z \in (A \cup B) \cap (A \cup C)$, we have to show that it's in both $(A \cup B)$ and $(A \cup C)$. But $z \in A \cup B$ iff $z \in A$ or $z \in B$, and we already have (as the assumption of case 1) that $z \in A$. By the same reasoning—switching C for B—$z \in A \cup C$. This argument went in the reverse direction, so let's record our reasoning in the direction needed in our proof.

Since $z \in A$, $z \in A$ or $z \in B$, and hence, by definition of $\cup$, $z \in A \cup B$. Similarly, $z \in A \cup C$. But this means that $z \in (A \cup B) \cap (A \cup C)$, by definition of $\cap$.

> This completes the first case of the proof by cases. Now we want to derive the conclusion in the second case, where $z \in B \cap C$.

Case 2: Suppose that $z \in B \cap C$.

> Again, we are working with the intersection of two sets. Let's apply the definition of $\cap$:

Since $z \in B \cap C$, z must be an element of both B and C, by definition of $\cap$.

> It's time to look at our conclusion again. We have to show that z is in both $(A \cup B)$ and $(A \cup C)$. And again, the solution is immediate.

Since $z \in B$, $z \in (A \cup B)$. Since $z \in C$, also $z \in (A \cup C)$. So, $z \in (A \cup B) \cap (A \cup C)$.

> Here we applied the definitions of $\cup$ and $\cap$ again, but since we've already recalled those definitions, and already showed that if z is in one of two sets it is in

their union, we don't have to be as explicit in what we've done.

We've completed the second case of the proof by cases, so now we can assert our first conclusion.

So, if $z \in A \cup (B \cap C)$ then $z \in (A \cup B) \cap (A \cup C)$.

Now we just want to show the other direction, that every element of $(A \cup B) \cap (A \cup C)$ is an element of $A \cup (B \cap C)$. As before, we prove this universal claim by assuming we have an arbitrary element of the first set and show it must be in the second set. Let's state what we're about to do.

Now, assume that $z \in (A \cup B) \cap (A \cup C)$. We want to show that $z \in A \cup (B \cap C)$.

We are now working from the hypothesis that $z \in (A \cup B) \cap (A \cup C)$. It hopefully isn't too confusing that we're using the same z here as in the first part of the proof. When we finished that part, all the assumptions we've made there are no longer in effect, so now we can make new assumptions about what z is. If that is confusing to you, just replace z with a different variable in what follows.

We know that z is in both $A \cup B$ and $A \cup C$, by definition of $\cap$. And by the definition of $\cup$, we can further unpack this to: either $z \in A$ or $z \in B$, and also either $z \in A$ or $z \in C$. This looks like a proof by cases again—except the "and" makes it confusing. You might think that this amounts to there being three possibilities: z is either in A, B or C. But that would be a mistake. We have to be careful, so let's consider each disjunction in turn.

By definition of $\cap$, $z \in A \cup B$ and $z \in A \cup C$. By definition of $\cup$, $z \in A$ or $z \in B$. We distinguish cases.

> Since we're focusing on the first disjunction, we haven't gotten our second disjunction (from unpacking $A \cup C$) yet. In fact, we don't need it yet. The first case is $z \in A$, and an element of a set is also an element of the union of that set with any other. So case 1 is easy:

Case 1: Suppose that $z \in A$. It follows that $z \in A \cup (B \cap C)$.

> Now for the second case, $z \in B$. Here we'll unpack the second $\cup$ and do another proof-by-cases:

Case 2: Suppose that $z \in B$. Since $z \in A \cup C$, either $z \in A$ or $z \in C$. We distinguish cases further:
  Case 2a: $z \in A$. Then, again, $z \in A \cup (B \cap C)$.

> Ok, this was a bit weird. We didn't actually need the assumption that $z \in B$ for this case, but that's ok.

Case 2b: $z \in C$. Then $z \in B$ and $z \in C$, so $z \in B \cap C$, and consequently, $z \in A \cup (B \cap C)$.

> This concludes both proofs-by-cases and so we're done with the second half.

So, if $z \in (A \cup B) \cap (A \cup C)$ then $z \in A \cup (B \cap C)$. □

C.6 Another Example

Proposition C.9. *If $A \subseteq C$, then $A \cup (C \setminus A) = C$.*

Proof. Suppose that $A \subseteq C$. We want to show that $A \cup (C \setminus A) = C$.

> We begin by observing that this is a conditional statement. It is tacitly universally quantified: the proposition holds for all sets A and C. So A and C are variables for arbitrary sets. To prove such a statement, we assume the antecedent and prove the consequent.

> We continue by using the assumption that $A \subseteq C$. Let's unpack the definition of $\subseteq$: the assumption means that all elements of A are also elements of C. Let's write this down—it's an important fact that we'll use throughout the proof.

By the definition of $\subseteq$, since $A \subseteq C$, for all z, if $z \in A$, then $z \in C$.

> We've unpacked all the definitions that are given to us in the assumption. Now we can move onto the conclusion. We want to show that $A \cup (C \setminus A) = C$, and so we set up a proof similarly to the last example: we show that every element of $A \cup (C \setminus A)$ is also an element of C and, conversely, every element of C is an element of $A \cup (C \setminus A)$. We can shorten this to: $A \cup (C \setminus A) \subseteq C$ and $C \subseteq A \cup (C \setminus A)$. (Here we're doing the opposite of unpacking a definition, but it makes the proof a bit easier to read.) Since this is a conjunction, we have to prove both parts. To show the first part, i.e., that every element of $A \cup (C \setminus A)$ is also an element of C, we assume that $z \in A \cup (C \setminus A)$ for an arbitrary z and show that $z \in C$. By the definition of $\cup$, we can conclude that $z \in A$ or $z \in C \setminus A$ from $z \in A \cup (C \setminus A)$. You should now be getting the hang of this.

$A \cup (C \setminus A) = C$ iff $A \cup (C \setminus A) \subseteq C$ and $C \subseteq (A \cup (C \setminus A)$. First we prove that $A \cup (C \setminus A) \subseteq C$. Let $z \in A \cup (C \setminus A)$. So, either $z \in A$ or $z \in (C \setminus A)$.

> We've arrived at a disjunction, and from it we want to prove that $z \in C$. We do this using proof by cases.

Case 1: $z \in A$. Since for all z, if $z \in A$, $z \in C$, we have that $z \in C$.

> Here we've used the fact recorded earlier which followed from the hypothesis of the proposition that $A \subseteq C$. The first case is complete, and we turn to

> the second case, $z \in (C \setminus A)$. Recall that $C \setminus A$ denotes the *difference* of the two sets, i.e., the set of all elements of C which are not elements of A. But any element of C not in A is in particular an element of C.

Case 2: $z \in (C \setminus A)$. This means that $z \in C$ and $z \notin A$. So, in particular, $z \in C$.

> Great, we've proved the first direction. Now for the second direction. Here we prove that $C \subseteq A \cup (C \setminus A)$. So we assume that $z \in C$ and prove that $z \in A \cup (C \setminus A)$.

Now let $z \in C$. We want to show that $z \in A$ or $z \in C \setminus A$.

> Since all elements of A are also elements of C, and $C \setminus A$ is the set of all things that are elements of C but not A, it follows that z is either in A or in $C \setminus A$. This may be a bit unclear if you don't already know why the result is true. It would be better to prove it step-by-step. It will help to use a simple fact which we can state without proof: $z \in A$ or $z \notin A$. This is called the "principle of excluded middle:" for any statement p, either p is true or its negation is true. (Here, p is the statement that $z \in A$.) Since this is a disjunction, we can again use proof-by-cases.

Either $z \in A$ or $z \notin A$. In the former case, $z \in A \cup (C \setminus A)$. In the latter case, $z \in C$ and $z \notin A$, so $z \in C \setminus A$. But then $z \in A \cup (C \setminus A)$.

> Our proof is complete: we have shown that $A \cup (C \setminus A) = C$.

□

C.7 Proof by Contradiction

In the first instance, proof by contradiction is an inference pattern that is used to prove negative claims. Suppose you want to show that some claim p is *false*, i.e., you want to show $\neg p$. The most promising strategy is to (a) suppose that p is true, and (b) show that this assumption leads to something you know to be false. "Something known to be false" may be a result that conflicts with—contradicts—p itself, or some other hypothesis of the overall claim you are considering. For instance, a proof of "if q then $\neg p$" involves assuming that q is true and proving $\neg p$ from it. If you prove $\neg p$ by contradiction, that means assuming p in addition to q. If you can prove $\neg q$ from p, you have shown that the assumption p leads to something that contradicts your other assumption q, since q and $\neg q$ cannot both be true. Of course, you have to use other inference patterns in your proof of the contradiction, as well as unpacking definitions. Let's consider an example.

Proposition C.10. *If $A \subseteq B$ and $B = \emptyset$, then A has no elements.*

Proof. Suppose $A \subseteq B$ and $B = \emptyset$. We want to show that A has no elements.

> Since this is a conditional claim, we assume the antecedent and want to prove the consequent. The consequent is: A has no elements. We can make that a bit more explicit: it's not the case that there is an $x \in A$.

A has no elements iff it's not the case that there is an x such that $x \in A$.

> So we've determined that what we want to prove is really a negative claim $\neg p$, namely: it's not the case that there is an $x \in A$. To use proof by contradiction, we have to assume the corresponding positive claim p, i.e., there is an $x \in A$, and prove a contradiction from it. We indicate that we're doing a proof by contradiction by writing "by way of contradiction, assume" or even just "suppose not," and then state the assumption p.

Suppose not: there is an $x \in A$.

> This is now the new assumption we'll use to obtain a contradiction. We have two more assumptions: that $A \subseteq B$ and that $B = \emptyset$. The first gives us that $x \in B$:

Since $A \subseteq B$, $x \in B$.

> But since $B = \emptyset$, every element of B (e.g., x) must also be an element of $\emptyset$.

Since $B = \emptyset$, $x \in \emptyset$. This is a contradiction, since by definition $\emptyset$ has no elements.

> This already completes the proof: we've arrived at what we need (a contradiction) from the assumptions we've set up, and this means that the assumptions can't all be true. Since the first two assumptions ($A \subseteq B$ and $B = \emptyset$) are not contested, it must be the last assumption introduced (there is an $x \in A$) that must be false. But if we want to be thorough, we can spell this out.

Thus, our assumption that there is an $x \in A$ must be false, hence, A has no elements by proof by contradiction. □

Every positive claim is trivially equivalent to a negative claim: p iff $\neg\neg p$. So proofs by contradiction can also be used to establish positive claims "indirectly," as follows: To prove p, read it as the negative claim $\neg\neg p$. If we can prove a contradiction from $\neg p$, we've established $\neg\neg p$ by proof by contradiction, and hence p.

In the last example, we aimed to prove a negative claim, namely that A has no elements, and so the assumption we made for the purpose of proof by contradiction (i.e., that there is an $x \in A$) was a positive claim. It gave us something to work with, namely the hypothetical $x \in A$ about which we continued to reason until we got to $x \in \emptyset$.

When proving a positive claim indirectly, the assumption you'd make for the purpose of proof by contradiction would be negative. But very often you can easily reformulate a positive claim as a negative claim, and a negative claim as a positive claim. Our previous proof would have been essentially the same had we proved "$A = \emptyset$" instead of the negative consequent "A has no elements." (By definition of $=$, "$A = \emptyset$" is a general claim, since it unpacks to "every element of A is an element of $\emptyset$ and vice versa".) But it is easily seen to be equivalent to the negative claim "not: there is an $x \in A$."

So it is sometimes easier to work with $\neg p$ as an assumption than it is to prove p directly. Even when a direct proof is just as simple or even simpler (as in the next examples), some people prefer to proceed indirectly. If the double negation confuses you, think of a proof by contradiction of some claim as a proof of a contradiction from the *opposite* claim. So, a proof by contradiction of $\neg p$ is a proof of a contradiction from the assumption p; and proof by contradiction of p is a proof of a contradiction from $\neg p$.

Proposition C.11. $A \subseteq A \cup B$.

Proof. We want to show that $A \subseteq A \cup B$.

> On the face of it, this is a positive claim: every $x \in A$ is also in $A \cup B$. The negation of that is: some $x \in A$ is $\notin A \cup B$. So we can prove the claim indirectly by assuming this negated claim, and showing that it leads to a contradiction.

Suppose not, i.e., $A \nsubseteq A \cup B$.

> We have a definition of $A \subseteq A \cup B$: every $x \in A$ is also $\in A \cup B$. To understand what $A \nsubseteq A \cup B$ means, we have to use some elementary logical manipulation on the unpacked definition: it's false that every $x \in A$ is also $\in A \cup B$ iff there is *some* $x \in A$ that is $\notin C$. (This is a place where you want to be very careful:

> many students' attempted proofs by contradiction fail because they analyze the negation of a claim like "all As are Bs" incorrectly.) In other words, $A \nsubseteq A \cup B$ iff there is an x such that $x \in A$ and $x \notin A \cup B$. From then on, it's easy.

So, there is an $x \in A$ such that $x \notin A \cup B$. By definition of $\cup$, $x \in A \cup B$ iff $x \in A$ or $x \in B$. Since $x \in A$, we have $x \in A \cup B$. This contradicts the assumption that $x \notin A \cup B$. □

Proposition C.12. *If $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$.*

Proof. Suppose $A \subseteq B$ and $B \subseteq C$. We want to show $A \subseteq C$.

> Let's proceed indirectly: we assume the negation of what we want to etablish.

Suppose not, i.e., $A \nsubseteq C$.

> As before, we reason that $A \nsubseteq C$ iff not every $x \in A$ is also $\in C$, i.e., some $x \in A$ is $\notin C$. Don't worry, with practice you won't have to think hard anymore to unpack negations like this.

In other words, there is an x such that $x \in A$ and $x \notin C$.

> Now we can use this to get to our contradiction. Of course, we'll have to use the other two assumptions to do it.

Since $A \subseteq B$, $x \in B$. Since $B \subseteq C$, $x \in C$. But this contradicts $x \notin C$. □

Proposition C.13. *If $A \cup B = A \cap B$ then $A = B$.*

Proof. Suppose $A \cup B = A \cap B$. We want to show that $A = B$.

> The beginning is now routine:

Assume, by way of contradiction, that $A \neq B$.

> Our assumption for the proof by contradiction is that $A \neq B$. Since $A = B$ iff $A \subseteq B$ an $B \subseteq A$, we get that $A \neq B$ iff $A \nsubseteq B$ *or* $B \nsubseteq A$. (Note how important it is to be careful when manipulating negations!) To prove a contradiction from this disjunction, we use a proof by cases and show that in each case, a contradiction follows.

$A \neq B$ iff $A \nsubseteq B$ or $B \nsubseteq A$. We distinguish cases.

> In the first case, we assume $A \nsubseteq B$, i.e., for some x, $x \in A$ but $\notin B$. $A \cap B$ is defined as those elements that A and B have in common, so if something isn't in one of them, it's not in the intersection. $A \cup B$ is A together with B, so anything in either is also in the union. This tells us that $x \in A \cup B$ but $x \notin A \cap B$, and hence that $A \cap B \neq A \cup B$.

Case 1: $A \nsubseteq B$. Then for some x, $x \in A$ but $x \notin B$. Since $x \notin B$, then $x \notin A \cap B$. Since $x \in A$, $x \in A \cup B$. So, $A \cap B \neq A \cup B$, contradicting the assumption that $A \cap B = A \cup B$.

Case 2: $B \nsubseteq A$. Then for some y, $y \in B$ but $y \notin A$. As before, we have $y \in A \cup B$ but $y \notin A \cap B$, and so $A \cap B \neq A \cup B$, again contradicting $A \cap B = A \cup B$. □

C.8 Reading Proofs

Proofs you find in textbooks and articles very seldom give all the details we have so far included in our examples. Authors often do not draw attention to when they distinguish cases, when they give an indirect proof, or don't mention that they use a definition. So when you read a proof in a textbook, you will often have to fill in those details for yourself in order to understand the proof. Doing this is also good practice to get the hang of the various moves you have to make in a proof. Let's look at an example.

Proposition C.14 (Absorption). *For all sets A, B,*

$$A \cap (A \cup B) = A$$

Proof. If $z \in A \cap (A \cup B)$, then $z \in A$, so $A \cap (A \cup B) \subseteq A$. Now suppose $z \in A$. Then also $z \in A \cup B$, and therefore also $z \in A \cap (A \cup B)$. □

The preceding proof of the absorption law is very condensed. There is no mention of any definitions used, no "we have to prove that" before we prove it, etc. Let's unpack it. The proposition proved is a general claim about any sets A and B, and when the proof mentions A or B, these are variables for arbitrary sets. The general claims the proof establishes is what's required to prove identity of sets, i.e., that every element of the left side of the identity is an element of the right and vice versa.

> "If $z \in A \cap (A \cup B)$, then $z \in A$, so $A \cap (A \cup B) \subseteq A$."

This is the first half of the proof of the identity: it establishes that if an arbitrary z is an element of the left side, it is also an element of the right, i.e., $A \cap (A \cup B) \subseteq A$. Assume that $z \in A \cap (A \cup B)$. Since z is an element of the intersection of two sets iff it is an element of both sets, we can conclude that $z \in A$ and also $z \in A \cup B$. In particular, $z \in A$, which is what we wanted to show. Since that's all that has to be done for the first half, we know that the rest of the proof must be a proof of the second half, i.e., a proof that $A \subseteq A \cap (A \cup B)$.

> "Now suppose $z \in A$. Then also $z \in A \cup B$, and therefore also $z \in A \cap (A \cup B)$."

We start by assuming that $z \in A$, since we are showing that, for any z, if $z \in A$ then $z \in A \cap (A \cup B)$. To show that $z \in A \cap (A \cup B)$, we have to show (by definition of "$\cap$") that (i) $z \in A$ and also (ii) $z \in A \cup B$. Here (i) is just our assumption, so there is nothing further to prove, and that's why the proof does not mention it again. For (ii), recall that z is an element of a union of sets iff it is an element of at least one of those sets. Since $z \in A$, and $A \cup B$ is the union of A and B, this is the case here. So $z \in A \cup B$. We've shown both (i) $z \in A$ and (ii) $z \in A \cup B$, hence, by definition of "$\cap$," $z \in A \cap (A \cup B)$. The proof doesn't mention those definitions; it's assumed the reader has already internalized them. If you haven't, you'll have to go back and remind yourself what they are. Then you'll also have to recognize why it follows from $z \in A$ that $z \in A \cup B$, and from $z \in A$ and $z \in A \cup B$ that $z \in A \cap (A \cup B)$.

Here's another version of the proof above, with everything made explicit:

Proof. [By definition of $=$ for sets, $A \cap (A \cup B) = A$ we have to show (a) $A \cap (A \cup B) \subseteq A$ and (b) $A \cap (A \cup B) \subseteq A$. (a): By definition of $\subseteq$, we have to show that if $z \in A \cap (A \cup B)$, then $z \in A$.] If $z \in A \cap (A \cup B)$, then $z \in A$ [since by definition of $\cap$, $z \in A \cap (A \cup B)$ iff $z \in A$ and $z \in A \cup B$], so $A \cap (A \cup B) \subseteq A$. [(b): By definition of $\subseteq$, we have to show that if $z \in A$, then $z \in A \cap (A \cup B)$.] Now suppose [(1)] $z \in A$. Then also [(2)] $z \in A \cup B$ [since by (1) $z \in A$ or $z \in B$, which by definition of $\cup$ means $z \in A \cup B$], and therefore also $z \in A \cap (A \cup B)$ [since the definition of $\cap$ requires that $z \in A$, i.e., (1), and $z \in A \cup B$), i.e., (2)]. □

C.9 I Can't Do It!

We all get to a point where we feel like giving up. But you *can* do it. Your instructor and teaching assistant, as well as your fellow students, can help. Ask them for help! Here are a few tips to help you avoid a crisis, and what to do if you feel like giving up.

To make sure you can solve problems successfully, do the following:

1. *Start as far in advance as possible.* We get busy throughout the semester and many of us struggle with procrastination, one of the best things you can do is to start your homework assignments early. That way, if you're stuck, you have time to look for a solution (that isn't crying).

2. *Talk to your classmates.* You are not alone. Others in the class may also struggle—but they may struggle with different things. Talking it out with your peers can give you a different perspective on the problem that might lead to a breakthrough. Of course, don't just copy their solution: ask them for a hint, or explain where you get stuck and ask them for the next step. And when you do get it, reciprocate. Helping someone else along, and explaining things will help you understand better, too.

3. *Ask for help.* You have many resources available to you—your instructor and teaching assistant are there for you and *want* you to succeed. They should be able to help you work out a problem and identify where in the process you're struggling.

4. *Take a break.* If you're stuck, it *might* be because you've been staring at the problem for too long. Take a short break, have a cup of tea, or work on a different problem for a while, then return to the problem with a fresh mind. Sleep on it.

Notice how these strategies require that you've started to work on the proof well in advance? If you've started the proof at 2am the day before it's due, these might not be so helpful.

This might sound like doom and gloom, but solving a proof is a challenge that pays off in the end. Some people do this as a career—so there must be something to enjoy about it. Like basically everything, solving problems and doing proofs is something that requires practice. You might see classmates who find this easy: they've probably just had lots of practice already. Try not to give in too easily.

If you do run out of time (or patience) on a particular problem: that's ok. It doesn't mean you're stupid or that you will never get it. Find out (from your instructor or another student) how it is done, and identify where you went wrong or got stuck, so you can avoid doing that the next time you encounter a similar issue. Then try to do it without looking at the solution. And next time, start (and ask for help) earlier.

C.10 Other Resources

There are many books on how to do proofs in mathematics which may be useful. Check out *How to Read and do Proofs: An Introduction to Mathematical Thought Processes* (Solow, 2013) and *How to Prove It: A Structured Approach* (Velleman, 2019) in particular. The *Book of Proof* (Hammack, 2013) and *Mathematical Reasoning* (Sandstrum, 2019) are books on proof that are freely available online. Philosophers might find *More Precisely: The Math you need to do Philosophy* (Steinhart, 2018) to be a good primer on mathematical reasoning.

There are also various shorter guides to proofs available on the internet; e.g., "Introduction to Mathematical Arguments" (Hutchings, 2003) and "How to write proofs" (Cheng, 2004).

Motivational Videos

Feel like you have no motivation to do your homework? Feeling down? These videos might help!

- `https://www.youtube.com/watch?v=ZXsQAXx_ao0`
- `https://www.youtube.com/watch?v=BQ4yd2W50No`
- `https://www.youtube.com/watch?v=StTqXEQ2l-Y`

Problems

Problem C.1. Suppose you are asked to prove that $A \cap B \neq \emptyset$. Unpack all the definitions occurring here, i.e., restate this in a way that does not mention "$\cap$", "$=$", or "$\emptyset$".

Problem C.2. Prove *indirectly* that $A \cap B \subseteq A$.

Problem C.3. Expand the following proof of $A \cup (A \cap B) = A$, where you mention all the inference patterns used, why each step follows from assumptions or claims established before it, and where we have to appeal to which definitions.

Proof. If $z \in A \cup (A \cap B)$ then $z \in A$ or $z \in A \cap B$. If $z \in A \cap B$, $z \in A$. Any $z \in A$ is also $\in A \cup (A \cap B)$. □

APPENDIX D

Induction

D.1 Introduction

Induction is an important proof technique which is used, in different forms, in almost all areas of logic, theoretical computer science, and mathematics. It is needed to prove many of the results in logic.

Induction is often contrasted with deduction, and characterized as the inference from the particular to the general. For instance, if we observe many green emeralds, and nothing that we would call an emerald that's not green, we might conclude that all emeralds are green. This is an inductive inference, in that it proceeds from many particular cases (this emerald is green, that emerald is green, etc.) to a general claim (all emeralds are green). *Mathematical* induction is also an inference that concludes a general claim, but it is of a very different kind than this "simple induction."

Very roughly, an inductive proof in mathematics concludes that all mathematical objects of a certain sort have a certain property. In the simplest case, the mathematical objects an inductive proof is concerned with are natural numbers. In that case an inductive proof is used to establish that all natural numbers have some property, and it does this by showing that

1. 0 has the property, and

2. whenever a number k has the property, so does $k + 1$.

Induction on natural numbers can then also often be used to prove general claims about mathematical objects that can be assigned numbers. For instance, finite sets each have a finite number n of elements, and if we can use induction to show that every number n has the property "all finite sets of size n are …" then we will have shown something about all finite sets.

Induction can also be generalized to mathematical objects that are *inductively defined*. For instance, expressions of a formal language such as those of first-order logic are defined inductively. *Structural induction* is a way to prove results about all such expressions. Structural induction, in particular, is very useful—and widely used—in logic.

D.2 Induction on $\mathbb{N}$

In its simplest form, induction is a technique used to prove results for all natural numbers. It uses the fact that by starting from 0 and repeatedly adding 1 we eventually reach every natural number. So to prove that something is true for every number, we can (1) establish that it is true for 0 and (2) show that whenever it is true for a number n, it is also true for the next number $n+1$. If we abbreviate "number n has property P" by $P(n)$ (and "number k has property P" by $P(k)$, etc.), then a proof by induction that $P(n)$ for all $n \in \mathbb{N}$ consists of:

1. a proof of $P(0)$, and

2. a proof that, for any k, if $P(k)$ then $P(k + 1)$.

To make this crystal clear, suppose we have both (1) and (2). Then (1) tells us that $P(0)$ is true. If we also have (2), we know in particular that if $P(0)$ then $P(0 + 1)$, i.e., $P(1)$. This follows from the general statement "for any k, if $P(k)$ then $P(k + 1)$" by putting 0 for k. So by modus ponens, we have that $P(1)$. From (2) again, now taking 1 for n, we have: if $P(1)$ then $P(2)$. Since we've just established $P(1)$, by modus ponens, we have $P(2)$. And so on. For any number n, after doing this n times, we eventually arrive at $P(n)$. So (1) and (2) together establish $P(n)$ for any $n \in \mathbb{N}$.

Let's look at an example. Suppose we want to find out how many different sums we can throw with n dice. Although it might seem silly, let's start with 0 dice. If you have no dice there's only one possible sum you can "throw": no dots at all, which sums to 0. So the number of different possible throws is 1. If you have only one die, i.e., $n = 1$, there are six possible values, 1 through 6. With two dice, we can throw any sum from 2 through 12, that's 11 possibilities. With three dice, we can throw any number from 3 to 18, i.e., 16 different possibilities. 1, 6, 11, 16: looks like a pattern: maybe the answer is $5n + 1$? Of course, $5n + 1$ is the maximum possible, because there are only $5n + 1$ numbers between n, the lowest value you can throw with n dice (all 1's) and $6n$, the highest you can throw (all 6's).

Theorem D.1. *With n dice one can throw all $5n + 1$ possible values between n and $6n$.*

Proof. Let $P(n)$ be the claim: "It is possible to throw any number between n and $6n$ using n dice." To use induction, we prove:

1. The *induction basis* $P(1)$, i.e., with just one die, you can throw any number between 1 and 6.

2. The *induction step*, for all k, if $P(k)$ then $P(k + 1)$.

(1) Is proved by inspecting a 6-sided die. It has all 6 sides, and every number between 1 and 6 shows up one on of the sides. So it is possible to throw any number between 1 and 6 using a single die.

To prove (2), we assume the antecedent of the conditional, i.e., $P(k)$. This assumption is called the *inductive hypothesis*. We use it to prove $P(k+1)$. The hard part is to find a way of thinking about the possible values of a throw of $k + 1$ dice in terms of the possible values of throws of k dice plus of throws of the extra $k+1$-st die—this is what we have to do, though, if we want to use the inductive hypothesis.

The inductive hypothesis says we can get any number between k and $6k$ using k dice. If we throw a 1 with our $(k+1)$-st die, this adds 1 to the total. So we can throw any value between $k+1$ and $6k+1$ by throwing k dice and then rolling a 1 with the $(k+1)$-st die. What's left? The values $6k+2$ through $6k+6$. We can get these by rolling k 6s and then a number between 2 and 6 with our $(k+1)$-st die. Together, this means that with $k+1$ dice we can throw any of the numbers between $k+1$ and $6(k+1)$, i.e., we've proved $P(k+1)$ using the assumption $P(k)$, the inductive hypothesis. □

Very often we use induction when we want to prove something about a series of objects (numbers, sets, etc.) that is itself defined "inductively," i.e., by defining the $(n+1)$-st object in terms of the n-th. For instance, we can define the sum s_n of the natural numbers up to n by

$$\begin{aligned} s_0 &= 0 \\ s_{n+1} &= s_n + (n+1) \end{aligned}$$

This definition gives:

$$\begin{aligned} s_0 &= 0, & & \\ s_1 &= s_0 + 1 & &= 1, \\ s_2 &= s_1 + 2 & &= 1 + 2 = 3 \\ s_3 &= s_2 + 3 & &= 1 + 2 + 3 = 6, \text{ etc.} \end{aligned}$$

Now we can prove, by induction, that $s_n = n(n+1)/2$.

Proposition D.2. $s_n = n(n+1)/2$.

Proof. We have to prove (1) that $s_0 = 0 \cdot (0+1)/2$ and (2) if $s_k = k(k+1)/2$ then $s_{k+1} = (k+1)(k+2)/2$. (1) is obvious. To prove (2), we assume the inductive hypothesis: $s_k = k(k+1)/2$. Using it, we have to show that $s_{k+1} = (k+1)(k+2)/2$.

What is s_{k+1}? By the definition, $s_{k+1} = s_k + (k+1)$. By inductive hypothesis, $s_k = k(k+1)/2$. We can substitute this into the previous equation, and then just need a bit of arithmetic of fractions:

$$\begin{aligned} s_{k+1} &= \frac{k(k+1)}{2} + (k+1) = \\ &= \frac{k(k+1)}{2} + \frac{2(k+1)}{2} = \\ &= \frac{k(k+1) + 2(k+1)}{2} = \\ &= \frac{(k+2)(k+1)}{2}. \end{aligned}$$

□

The important lesson here is that if you're proving something about some inductively defined sequence a_n, induction is the obvious way to go. And even if it isn't (as in the case of the possibilities of dice throws), you can use induction if you can somehow relate the case for $k+1$ to the case for k.

D.3 Strong Induction

In the principle of induction discussed above, we prove $P(0)$ and also if $P(k)$, then $P(k+1)$. In the second part, we assume that $P(k)$ is true and use this assumption to prove $P(k+1)$. Equivalently, of course, we could assume $P(k-1)$ and use it to prove $P(k)$—the important part is that we be able to carry out the inference from any number to its successor; that we can prove the claim in question for any number under the assumption it holds for its predecessor.

There is a variant of the principle of induction in which we don't just assume that the claim holds for the predecessor $k-1$ of k, but for all numbers smaller than k, and use this assumption to establish the claim for k. This also gives us the claim $P(n)$ for all $n \in \mathbb{N}$. For once we have established $P(0)$, we have thereby established that P holds for all numbers less than 1. And if we know that if $P(l)$ for all $l < k$, then $P(k)$, we know this in particular for $k = 1$. So we can conclude $P(1)$. With this we have proved $P(0)$ and $P(1)$, i.e., $P(l)$ for all $l < 2$, and since we have also the conditional, if $P(l)$ for all $l < 2$, then $P(2)$, we can conclude $P(2)$, and so on.

In fact, if we can establish the general conditional "for all k, if $P(l)$ for all $l < k$, then $P(k)$," we do not have to establish $P(0)$ anymore, since it follows from it. For remember that a general claim like "for all $l < k$, $P(l)$" is true if there are no $l < k$. This is a case of vacuous quantification: "all As are Bs" is true if there are no As, $\forall x\,(A(x) \rightarrow B(x))$ is true if no x satisfies $A(x)$. In this case, the formalized version would be "$\forall l\,(l < k \rightarrow P(l))$"—and that is true if there are no $l < k$. And if $k = 0$ that's exactly the case: no $l < 0$, hence "for all $l < 0$, $P(0)$" is true, whatever P is. A proof of "if $P(l)$ for all $l < k$, then $P(k)$" thus automatically establishes $P(0)$.

This variant is useful if establishing the claim for k can't be made to just rely on the claim for $k - 1$ but may require the assumption that it is true for one or more $l < k$.

D.4 Inductive Definitions

In logic we very often define kinds of objects *inductively*, i.e., by specifying rules for what counts as an object of the kind to be defined which explain how to get new objects of that kind from old objects of that kind. For instance, we often define special kinds of sequences of symbols, such as the terms and formulas of a language, by induction. For a simple example, consider strings of consisting of letters a, b, c, d, the symbol $\circ$, and brackets [and], such as "[[c$\circ$d][", "[a[]$\circ$]", "a" or "[[a$\circ$b]$\circ$d]". You probably feel that there's something "wrong" with the first two strings: the brackets don't "balance" at all in the first, and you might feel that the "$\circ$" should "connect" expressions that themselves make sense. The third and fourth string look better: for every "[" there's a closing "]" (if there are any at all), and for any $\circ$ we can find "nice" expressions on either side, surrounded by a pair of parentheses.

We would like to precisely specify what counts as a "nice term." First of all, every letter by itself is nice. Anything that's not just a letter by itself should be of the form "$[t \circ s]$" where s and t are themselves nice. Conversely, if t and s are nice, then we can form a new nice term by putting a $\circ$ between them and surround them by a pair of brackets. We might use these operations to *define* the set of nice terms. This is an *inductive definition.*

Definition D.3 (Nice terms). The set of *nice terms* is inductively defined as follows:

1. Any letter a, b, c, d is a nice term.
2. If s_1 and s_2 are nice terms, then so is $[s_1 \circ s_2]$.
3. Nothing else is a nice term.

This definition tells us that something counts as a nice term iff it can be constructed according to the two conditions (1) and (2) in some finite number of steps. In the first step, we construct all nice terms just consisting of letters by themselves, i.e.,

$$\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}$$

In the second step, we apply (2) to the terms we've constructed. We'll get

$$[\mathrm{a} \circ \mathrm{a}], [\mathrm{a} \circ \mathrm{b}], [\mathrm{b} \circ \mathrm{a}], \ldots, [\mathrm{d} \circ \mathrm{d}]$$

for all combinations of two letters. In the third step, we apply (2) again, to any two nice terms we've constructed so far. We get new nice term such as $[\mathrm{a} \circ [\mathrm{a} \circ \mathrm{a}]]$—where t is a from step 1 and s is $[\mathrm{a} \circ \mathrm{a}]$ from step 2—and $[[\mathrm{b} \circ \mathrm{c}] \circ [\mathrm{d} \circ \mathrm{b}]]$ constructed out of the two terms $[\mathrm{b} \circ \mathrm{c}]$ and $[\mathrm{d} \circ \mathrm{b}]$ from step 2. And so on. Clause (3) rules out that anything not constructed in this way sneaks into the set of nice terms.

Note that we have not yet proved that every sequence of symbols that "feels" nice is nice according to this definition. However, it should be clear that everything we can construct does in fact "feel nice": brackets are balanced, and $\circ$ connects parts that are themselves nice.

The key feature of inductive definitions is that if you want to prove something about all nice terms, the definition tells you which cases you must consider. For instance, if you are told that t is a nice term, the inductive definition tells you what t can look like: t can be a letter, or it can be $[s_1 \circ s_2]$ for some pair of nice terms s_1 and s_2. Because of clause (3), those are the only possibilities.

When proving claims about all of an inductively defined set, the strong form of induction becomes particularly important. For instance, suppose we want to prove that for every nice term of length n, the number of [in it is $< n/2$. This can be seen as a claim about all n: for every n, the number of [in any nice term of length n is $< n/2$.

Proposition D.4. *For any n, the number of [in a nice term of length n is $< n/2$.*

Proof. To prove this result by (strong) induction, we have to show that the following conditional claim is true:

> If for every $l < k$, any nice term of length l has $< l/2$ ['s, then any nice term of length k has $< k/2$ ['s.

To show this conditional, assume that its antecedent is true, i.e., assume that for any $l < k$, nice terms of length l contain $< l/2$ ['s. We call this assumption the inductive hypothesis. We want to show the same is true for nice terms of length k.

So suppose t is a nice term of length k. Because nice terms are inductively defined, we have two cases: (1) t is a letter by itself, or (2) t is $[s_1 \circ s_2]$ for some nice terms s_1 and s_2.

1. t is a letter. Then $k = 1$, and the number of [in t is 0. Since $0 < 1/2$, the claim holds.

2. t is $[s_1 \circ s_2]$ for some nice terms s_1 and s_2. Let's let l_1 be the length of s_1 and l_2 be the length of s_2. Then the length k of t is $l_1 + l_2 + 3$ (the lengths of s_1 and s_2 plus three symbols $[$, $\circ$, $]$). Since $l_1 + l_2 + 3$ is always greater than l_1, $l_1 < k$. Similarly, $l_2 < k$. That means that the induction hypothesis applies to the terms s_1 and s_2: the number m_1 of [in s_1 is $< l_1/2$, and the number m_2 of [in s_2 is $< l_2/2$.

   The number of [in t is the number of [in s_1, plus the number of [in s_2, plus 1, i.e., it is $m_1 + m_2 + 1$. Since $m_1 < l_1/2$ and $m_2 < l_2/2$ we have:

$$m_1 + m_2 + 1 < \frac{l_1}{2} + \frac{l_2}{2} + 1 = \frac{l_1 + l_2 + 2}{2} < \frac{l_1 + l_2 + 3}{2} = k/2.$$

In each case, we've shown that the number of [in t is $< k/2$ (on the basis of the inductive hypothesis). By strong induction, the proposition follows. □

D.5 Structural Induction

So far we have used induction to establish results about all natural numbers. But a corresponding principle can be used directly to prove results about all elements of an inductively defined set. This often called *structural* induction, because it depends on the structure of the inductively defined objects.

Generally, an inductive definition is given by (a) a list of "initial" elements of the set and (b) a list of operations which produce new elements of the set from old ones. In the case of nice terms, for instance, the initial objects are the letters. We only have one operation: the operations are

$$o(s_1, s_2) = [s_1 \circ s_2]$$

You can even think of the natural numbers $\mathbb{N}$ themselves as being given by an inductive definition: the initial object is 0, and the operation is the successor function $x + 1$.

In order to prove something about all elements of an inductively defined set, i.e., that every element of the set has a property P, we must:

1. Prove that the initial objects have P
2. Prove that for each operation o, if the arguments have P, so does the result.

For instance, in order to prove something about all nice terms, we would prove that it is true about all letters, and that it is true about $[s_1 \circ s_2]$ provided it is true of s_1 and s_2 individually.

Proposition D.5. *The number of* [*equals the number of*] *in any nice term t.*

Proof. We use structural induction. Nice terms are inductively defined, with letters as initial objects and the operation o for constructing new nice terms out of old ones.

1. The claim is true for every letter, since the number of [in a letter by itself is 0 and the number of] in it is also 0.
2. Suppose the number of [in s_1 equals the number of], and the same is true for s_2. The number of [in $o(s_1, s_2)$, i.e., in $[s_1 \circ s_2]$, is the sum of the number of [in s_1 and s_2 plus one. The number of] in $o(s_1, s_2)$ is the sum of the number of] in s_1 and s_2 plus one. Thus, the number of [in $o(s_1, s_2)$ equals the number of] in $o(s_1, s_2)$. □

Let's give another proof by structural induction: a proper initial segment of a string t of symbols is any string s that agrees with t symbol by symbol, read from the left, but t is longer. So, e.g., $[a \circ$ is a proper initial segment of $[a \circ b]$, but neither are $[b \circ$ (they disagree at the second symbol) nor $[a \circ b]$ (they are the same length).

Proposition D.6. *Every proper initial segment of a nice term t has more* [*'s than*] *'s.*

Proof. By induction on t:

1. t is a letter by itself: Then t has no proper initial segments.

2. $t = [s_1 \circ s_2]$ for some nice terms s_1 and s_2. If r is a proper initial segment of t, there are a number of possibilities:

   a) r is just [: Then r has one more [than it does].

   b) r is $[r_1$ where r_1 is a proper initial segment of s_1: Since s_1 is a nice term, by induction hypothesis, r_1 has more [than] and the same is true for $[r_1$.

   c) r is $[s_1$ or $[s_1 \circ$: By the previous result, the number of [and] in s_1 are equal; so the number of [in $[s_1$ or $[s_1 \circ$ is one more than the number of].

   d) r is $[s_1 \circ r_2$ where r_2 is a proper initial segment of s_2: By induction hypothesis, r_2 contains more [than]. By the previous result, the number of [and of] in s_1 are equal. So the number of [in $[s_1 \circ r_2$ is greater than the number of].

   e) r is $[s_1 \circ s_2$: By the previous result, the number of [and] in s_1 are equal, and the same for s_2. So there is one more [in $[s_1 \circ s_2$ than there are]. □

D.6 Relations and Functions

When we have defined a set of objects (such as the natural numbers or the nice terms) inductively, we can also define *relations on* these objects by induction. For instance, consider the following idea: a nice term t_1 is a subterm of a nice term t_2 if it occurs as a part of it. Let's use a symbol for it: $t_1 \sqsubseteq t_2$. Every nice term is a subterm of itself, of course: $t \sqsubseteq t$. We can give an inductive definition of this relation as follows:

Definition D.7. The relation of a nice term t_1 being a subterm of t_2, $t_1 \sqsubseteq t_2$, is defined by induction on t_2 as follows:

1. If t_2 is a letter, then $t_1 \sqsubseteq t_2$ iff $t_1 = t_2$.

2. If t_2 is $[s_1 \circ s_2]$, then $t_1 \sqsubseteq t_2$ iff $t_1 = t_2$, $t_1 \sqsubseteq s_1$, or $t_1 \sqsubseteq s_2$.

This definition, for instance, will tell us that $\mathrm{a} \sqsubseteq [\mathrm{b} \circ \mathrm{a}]$. For (2) says that $\mathrm{a} \sqsubseteq [\mathrm{b} \circ \mathrm{a}]$ iff $\mathrm{a} = [\mathrm{b} \circ \mathrm{a}]$, or $\mathrm{a} \sqsubseteq b$, or $\mathrm{a} \sqsubseteq \mathrm{a}$. The first two are false: a clearly isn't identical to $[\mathrm{b} \circ \mathrm{a}]$, and by (1), $\mathrm{a} \sqsubseteq \mathrm{b}$ iff $\mathrm{a} = \mathrm{b}$, which is also false. However, also by (1), $\mathrm{a} \sqsubseteq \mathrm{a}$ iff $\mathrm{a} = \mathrm{a}$, which is true.

It's important to note that the success of this definition depends on a fact that we haven't proved yet: every nice term t is either a letter by itself, or there are *uniquely determined* nice terms s_1 and s_2 such that $t = [s_1 \circ s_2]$. "Uniquely determined" here means that if $t = [s_1 \circ s_2]$ it isn't *also* $= [r_1 \circ r_2]$ with $s_1 \neq r_1$ or $s_2 \neq r_2$. If this were the case, then clause (2) may come in conflict with itself: reading t_2 as $[s_1 \circ s_2]$ we might get $t_1 \sqsubseteq t_2$, but if we read t_2 as $[r_1 \circ r_2]$ we might get not $t_1 \sqsubseteq t_2$. Before we prove that this can't happen, let's look at an example where it *can* happen.

Definition D.8. Define *bracketless terms* inductively by

1. Every letter is a bracketless term.

2. If s_1 and s_2 are bracketless terms, then $s_1 \circ s_2$ is a bracketless term.

3. Nothing else is a bracketless term.

Bracketless terms are, e.g., a, $\mathrm{b} \circ \mathrm{d}$, $\mathrm{b} \circ \mathrm{a} \circ \mathrm{b}$. Now if we defined "subterm" for bracketless terms the way we did above, the second clause would read

$$\text{If } t_2 = s_1 \circ s_2, \text{ then } t_1 \sqsubseteq t_2 \text{ iff } t_1 = t_2,\ t_1 \sqsubseteq s_1, \text{ or } t_1 \sqsubseteq s_2.$$

Now $\mathrm{b} \circ \mathrm{a} \circ \mathrm{b}$ is of the form $s_1 \circ s_2$ with

$$s_1 = \mathrm{b} \text{ and} \qquad\qquad s_2 = \mathrm{a} \circ \mathrm{b}.$$

It is also of the form $r_1 \circ r_2$ with

$$r_1 = \mathrm{b} \circ \mathrm{a} \text{ and} \qquad\qquad r_2 = \mathrm{b}.$$

Now is $\mathrm{a} \circ \mathrm{b}$ a subterm of $\mathrm{b} \circ \mathrm{a} \circ \mathrm{b}$? The answer is yes if we go by the first reading, and no if we go by the second.

The property that the way a nice term is built up from other nice terms is unique is called *unique readability*. Since inductive definitions of relations for such inductively defined objects are important, we have to prove that it holds.

Proposition D.9. *Suppose t is a nice term. Then either t is a letter by itself, or there are uniquely determined nice terms s_1, s_2 such that $t = [s_1 \circ s_2]$.*

Proof. If t is a letter by itself, the condition is satisfied. So assume t isn't a letter by itself. We can tell from the inductive definition that then t must be of the form $[s_1 \circ s_2]$ for some nice terms s_1 and s_2. It remains to show that these are uniquely determined, i.e., if $t = [r_1 \circ r_2]$, then $s_1 = r_1$ and $s_2 = r_2$.

So suppose $t = [s_1 \circ s_2]$ and also $t = [r_1 \circ r_2]$ for nice terms s_1, s_2, r_1, r_2. We have to show that $s_1 = r_1$ and $s_2 = r_2$. First, s_1 and r_1 must be identical, for otherwise one is a proper initial segment of the other. But by Proposition D.6, that is impossible if s_1 and r_1 are both nice terms. But if $s_1 = r_1$, then clearly also $s_2 = r_2$. □

We can also define functions inductively: e.g., we can define the function f that maps any nice term to the maximum depth of nested $[\dots]$ in it as follows:

Definition D.10. The *depth* of a nice term, $f(t)$, is defined inductively as follows:

$$f(t) = \begin{cases} 0 & \text{if } t \text{ is a letter} \\ \max(f(s_1), f(s_2)) + 1 & \text{if } t = [s_1 \circ s_2]. \end{cases}$$

For instance

$$\begin{aligned} f([\mathrm{a} \circ \mathrm{b}]) &= \max(f(\mathrm{a}), f(\mathrm{b})) + 1 = \\ &= \max(0,0) + 1 = 1, \text{ and} \\ f([[\mathrm{a} \circ \mathrm{b}] \circ \mathrm{c}]) &= \max(f([\mathrm{a} \circ \mathrm{b}]), f(\mathrm{c})) + 1 = \\ &= \max(1,0) + 1 = 2. \end{aligned}$$

Here, of course, we assume that s_1 an s_2 are nice terms, and make use of the fact that every nice term is either a letter or of the form $[s_1 \circ s_2]$. It is again important that it can be of this form in only one way. To see why, consider again the bracketless terms we defined earlier. The corresponding "definition" would be:

$$g(t) = \begin{cases} 0 & \text{if } t \text{ is a letter} \\ \max(g(s_1), g(s_2)) + 1 & \text{if } t = s_1 \circ s_2. \end{cases}$$

Now consider the bracketless term $\mathrm{a} \circ \mathrm{b} \circ \mathrm{c} \circ \mathrm{d}$. It can be read in more than one way, e.g., as $s_1 \circ s_2$ with

$$s_1 = \mathrm{a} \text{ and} \qquad s_2 = \mathrm{b} \circ \mathrm{c} \circ \mathrm{d},$$

or as $r_1 \circ r_2$ with

$$r_1 = \mathrm{a} \circ b \text{ and} \qquad r_2 = \mathrm{c} \circ \mathrm{d}.$$

Calculating g according to the first way of reading it would give

$$\begin{aligned} g(s_1 \circ s_2) &= \max(g(\mathrm{a}), g(\mathrm{b} \circ \mathrm{c} \circ \mathrm{d})) + 1 = \\ &= \max(0,2) + 1 = 3 \end{aligned}$$

while according to the other reading we get

$$g(r_1 \circ r_2) = \max(g(\mathrm{a} \circ \mathrm{b}), g(\mathrm{c} \circ \mathrm{d})) + 1 =$$

$$= \max(1,1) + 1 = 2$$

But a function must always yield a unique value; so our "definition" of g doesn't define a function at all.

Problems

Problem D.1. Define the set of supernice terms by

1. Any letter a, b, c, d is a supernice term.
2. If s is a supernice term, then so is $[s]$.
3. If s_1 and s_2 are supernice terms, then so is $[s_1 \circ s_2]$.
4. Nothing else is a supernice term.

Show that the number of [in a supernice term t of length n is $\leq n/2 + 1$.

Problem D.2. Prove by structural induction that no nice term starts with].

Problem D.3. Give an inductive definition of the function l, where $l(t)$ is the number of symbols in the nice term t.

Problem D.4. Prove by structural induction on nice terms t that $f(t) < l(t)$ (where $l(t)$ is the number of symbols in t and $f(t)$ is the depth of t as defined in Definition D.10).

APPENDIX E

The Greek Alphabet

Alpha	α	A	Nu	ν	N
Beta	β	B	Xi	ξ	Ξ
Gamma	γ	Γ	Omicron	o	O
Delta	δ	Δ	Pi	π	Π
Epsilon	ε	E	Rho	ρ	P
Zeta	ζ	Z	Sigma	σ	Σ
Eta	η	H	Tau	τ	T
Theta	θ	Θ	Upsilon	υ	Υ
Iota	ι	I	Phi	φ	Φ
Kappa	κ	K	Chi	χ	X
Lambda	λ	Λ	Psi	ψ	Ψ
Mu	μ	M	Omega	ω	Ω

Bibliography

Cheng, Eugenia. 2004. How to write proofs: A quick quide. URL http://http://eugeniacheng.com/wp-content/uploads/2017/02/cheng-proofguide.pdf.

Hammack, Richard. 2013. *Book of Proof*. Richmond, VA: Virginia Commonwealth University. URL http://www.people.vcu.edu/~rhammack/BookOfProof/BookOfProof.pdf.

Hutchings, Michael. 2003. Introduction to mathematical arguments. URL https://math.berkeley.edu/~hutching/teach/proofs.pdf.

Sandstrum, Ted. 2019. *Mathematical Reasoning: Writing and Proof*. Allendale, MI: Grand Valley State University. URL https://scholarworks.gvsu.edu/books/7/.

Solow, Daniel. 2013. *How to Read and Do Proofs*. Hoboken, NJ: Wiley.

Steinhart, Eric. 2018. *More Precisely: The Math You Need to Do Philosophy*. Peterborough, ON: Broadview, 2nd ed.

Velleman, Daniel J. 2019. *How to Prove It: A Structured Approach*. Cambridge: Cambridge University Press, 3rd ed.

About the Open Logic Project

The *Open Logic Text* is an open-source, collaborative textbook of formal meta-logic and formal methods, starting at an intermediate level (i.e., after an introductory formal logic course). Though aimed at a non-mathematical audience (in particular, students of philosophy and computer science), it is rigorous.

Coverage of some topics currently included may not yet be complete, and many sections still require substantial revision. We plan to expand the text to cover more topics in the future. We also plan to add features to the text, such as a glossary, a list of further reading, historical notes, pictures, better explanations, sections explaining the relevance of results to philosophy, computer science, and mathematics, and more problems and examples. If you find an error, or have a suggestion, please let the project team know.

The project operates in the spirit of open source. Not only is the text freely available, we provide the LaTeX source under the Creative Commons Attribution license, which gives anyone the right to download, use, modify, re-arrange, convert, and re-distribute our work, as long as they give appropriate credit. Please see the Open Logic Project website at openlogicproject.org for additional information.