Boxes and Diamonds

An Open Introduction to Modal Logic
Ann Arbor remix

Summer 2020

Boxes and Diamonds

The Open Logic Project

Instigator

Editorial Board

Contributors

Boxes and Diamonds

An Open Introduction to Modal Logic
Ann Arbor remix

Remixed by Brian Weatherson; based on a version by Richard Zach

Summer 2020

The Open Logic Project would like to acknowledge the generous support of the Faculty of Arts and 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 Universalis ADF Standard by LaTeX.

Contents

PART I

Propositional Logic

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 be true or false. Whenever the "truth value" of the propositional variable in a formula are determined, so is the truth value of any formulas formed from them using propositional connectives. We say that propositional logic is *truth functional*, because its semantics is given by functions of truth values. In particular, in propositional logic we leave out of consideration any further determination of truth and falsity, e.g., whether something is necessarily true rather than just contingently true, or whether something is known to be true, or whether something is true now rather than was true or will be true. We only consider two truth values true ($\mathbb{T}$) and false ($\mathbb{F}$), and so exclude from discussion the possibility that a statement may be neither true nor false, or only half true. We also concentrate only on connectives where the truth value of a formula built from them is completely determined by the truth values of its parts (and not, say, on its meaning). In particular, whether the truth value of conditionals in English is truth functional in this sense is contentious. The material conditional $\rightarrow$ is; other logics deal with conditionals that are not truth functional.

In order to develop the theory and metatheory of truth-functional propositional logic, we must first define the syntax and semantics of its expressions. We will describe one way of constructing formulas from propositional variables using the connectives. Alternative definitions are possible. Other systems will chose different symbols, will select different sets of connectives as primitive, will use parentheses differently (or even not at all, as in the case of so-called Polish notation). What all approaches have in common, though, is that the formation rules define the set of formulas *inductively*. If done properly, every expression can result essentially in only one way according to the formation rules. The inductive definition resulting in expressions that are *uniquely readable* means we can give meanings to these expressions using the same method—inductive definition.

Giving the meaning of expressions is the domain of semantics. The central concept in semantics for propositonal logic is that of satisfaction in a valuation . A valuation v assigns truth values $\mathbb{T}$, $\mathbb{F}$ to the propositional variables. Any valuation determines a truth value $\overline{v}(A)$ for any formula A. A formula is satisfied in a valuation v iff $\overline{v}(A) = \mathbb{T}$—we write this as $v \vDash A$. This relation can also be defined by induction on the structure of A, using the truth functions for the logical connectives to define, say, satisfaction of $A \land B$ 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.

In addition to the primitive connectives introduced above, we also use the following *defined* symbols: $\leftrightarrow$ (biconditional), truth $\top$

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 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 x is a variable , then $\forall x\, A$ is a formula.
8. If A is a formula and x is a variable , then $\exists x\, A$ is a formula.
9. Nothing else is a formula.

The definitions of the set of terms and that of formulas are *inductive definitions*. Essentially, we construct the set of formulas in infinitely many stages. In the initial stage, we pronounce all atomic formulas to be formulas; this corresponds to the first few cases of the definition, i.e., the cases for $\bot$, p_i. "Atomic formula" thus means any formula of this form.

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

Definition 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 of all the atomic formulas and is such that*

1. *it holds for $\neg A$ whenever it holds for A;*
2. *if holds for and $(A \wedge B)$ whenever it holds for A and B;*
3. *if holds for and $(A \vee B)$ whenever it holds for A and B;*
4. *if holds for and $(A \to B)$ whenever it holds for A and B;*

then P holds of 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. $\mathrm{Frm}(\mathcal{L}_0)$ is the smallest such class, so $\mathrm{Frm} \subseteq S$. So $\mathrm{Frm} = S$, and every formula has propery P. □

Proposition 1.5. *Any formula in* $\mathrm{Frm}(\mathcal{L}_0)$ *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* $\mathrm{Frm}(\mathcal{L}_0)$ *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 B in* $\mathrm{Frm}(\mathcal{L}_0)$.

4. $(B \wedge C)$ *for some formulas* B *and* C.

5. $(B \vee C)$ *for some formulas* B *and* C.

6. $(B \to C)$ *for some formulas* B *and* C.

Moreover, such parsing is unique.

Proof. By induction on A. For instance, suppose that A has two distinct readings as $(B \to C)$ and $(B' \to C')$. Then B and B' must be the same (or else one would be a proper initial segment of the other); so if the two readings of A are distinct it must be because C and C' are distinct readings of the same sequence of symbols, which is impossible by the inductive hypothesis. □

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

1.4 Valuations and Satisfaction

Definition 1.9 (Valuations). Let $\{\mathbb{T}, \mathbb{F}\}$ be the set of the two truth values, "true" and "false." A *valuation* for $\mathcal{L}_0$ is a function v assigning either $\mathbb{T}$ or $\mathbb{F}$ to the propositional variables of the language, i.e., $v\colon \mathrm{At}_0 \to \{\mathbb{T}, \mathbb{F}\}$.

Definition 1.10. Given a valuation v, define the evaluation function $\overline{v}(:)\mathrm{Frm}(\mathcal{L}_0) \to \{\mathbb{T}, \mathbb{F}\}$ 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 valuation clauses correspond to the following truth tables:

A	B	$A \wedge B$	$A \vee B$	$A \to B$	$A \leftrightarrow B$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{F}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$	$\mathbb{F}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$	$\mathbb{T}$	$\mathbb{T}$

Theorem 1.11 (Local Determination). *Suppose that v_1 and v_2 are valuations that agree on the propositional letters occurring in A, i.e., $\overline{v_1}(p_n) = \overline{v_2}(p_n)$ whenever p_n occurs in A. Then they also agree on any A, i.e., $\overline{v_1}(A) = \overline{v_2}(A)$.*

Proof. By induction on A. □

Definition 1.12 (Satisfaction). Using the evaluation function, we can define the notion of *satisfaction of a formula A by a valuation v*, $v \vDash A$, inductively as follows. (We write $v \nvDash A$ to mean "not $v \vDash A$.")

1. $A \equiv \bot$: $v \nvDash \bot$.
2. $A \equiv p_i$: $\boldsymbol{M} \vDash p_i$ iff $\overline{v}(p_i) = \mathbb{T}$.
3. $A \equiv \neg B$: $v \vDash \neg B$ iff $v \nvDash B$.
4. $A \equiv (B \wedge C)$: $v \vDash (B \wedge C)$ iff $v \vDash B$ and $v \vDash C$.
5. $A \equiv (B \vee C)$: $v \vDash (B \vee C)$ iff $v \vDash A$ or $v \vDash B$ (or both).
6. $A \equiv (B \to C)$: $v \vDash (B \to C)$ 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 1.13. *$v \vDash A$ iff $\overline{v}(A) = \mathbb{T}$.*

Proof. By induction on A. □

1.5 Semantic Notions

We define the following semantic notions:

Definition 1.14.

1. A formula A is *satisfiable* if for some v, $v \vDash A$; it is *unsatisfiable* if for no v, $v \vDash A$;
2. A formula A is a *tautology* if $v \vDash A$ for all valuations v;
3. A formula A is *contingent* if it is satisfiable but not a tautology;
4. If Γ is a set of formulas, $\Gamma \vDash A$ ("Γ entails A") if and only if $v \vDash A$ for every valuation v for which $v \vDash \Gamma$.
5. If Γ is a set of formulas, Γ is *satisfiable* if there is a valuation v for which $v \vDash \Gamma$, and Γ is *unsatisfiable* otherwise.

Proposition 1.15.

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. *Monotony: if $\Gamma \subseteq \Delta$ and $\Gamma \vDash A$ then also $\Delta \vDash A$;*
5. *Transitivity: if $\Gamma \vDash A$ and $\Delta \cup \{A\} \vDash B$ then $\Gamma \cup \Delta \vDash B$;*

Proof. Exercise. □

Proposition 1.16. *$\Gamma \vDash A$ if and only if $\Gamma \cup \{\neg A\}$ is unsatisfiable;*

Proof. Exercise. □

Theorem 1.17 (Semantic Deduction Theorem)**.** *$\Gamma \vDash A \to B$ 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. Give a mathematically rigorous definition of $A[B/p]$ by induction.

Problem 1.4. Prove Proposition 1.13

Problem 1.5. Prove Proposition 1.15

Problem 1.6. Prove Proposition 1.16

Problem 1.7. Prove Theorem 1.17

CHAPTER 2

Tableaux

2.1 Tableaux

While many derivation systems operate with arrangements of sentences, tableaux operate with signed formulas. A signed formula is a pair consisting of a truth value sign ($\mathbb{T}$ or $\mathbb{F}$) and a sentence

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

A tableau consists of signed formulas arranged in a downward-branching tree. It begins with a number of *assumptions* and continues with signed formulas which result from one of the signed formulas above it by applying one of the rules of inference. Each rule allows us to add one or more signed formulas to the end of a branch, or two signed formulas side by side—in this case a branch splits into two, with the two added signed formulas forming the ends of the two branches.

A rule applied to a complex signed formula results in the addition of signed formulas which are immediate sub-formulas. They come in pairs, one rule for each of the two signs. For instance, the $\land\mathbb{T}$ rule applies to $\mathbb{T}\,A \land B$, and allows the addition of both the two signed formulas $\mathbb{T}\,A$ and $\mathbb{T}\,B$ to the end of any branch containing $\mathbb{T}\,A \land B$, and the rule $\land\mathbb{F}$ allows a branch to be split by adding $\mathbb{F}\,A$ and $\mathbb{F}\,B$ side-by-side. A tableau is closed if every one of its branches contains a matching pair of signed formulas $\mathbb{T}\,A$ and $\mathbb{F}\,A$.

The $\vdash$ relation based on tableaux is defined as follows: $\Gamma \vdash A$ iff there is some finite set $\Gamma_0 = \{B_1, \dots, B_n\} \subseteq \Gamma$ such that there is a closed tableau for the assumptions

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

For instance, here is a closed tableau that shows that $\vdash (A \land B) \to A$:

1.	$\mathbb{F}\,(A \land B) \to A$	Assumption
2.	$\mathbb{T}\, A \land B$	$\to\mathbb{F}\,1$
3.	$\mathbb{F}\, A$	$\to\mathbb{F}\,1$
4.	$\mathbb{T}\, A$	$\to\mathbb{T}\,2$
5.	$\mathbb{T}\, B$	$\to\mathbb{T}\,2$
	$\otimes$	

A set Γ is inconsistent in the tableau calculus if there is a closed tableau for assumptions

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

for some $B_i \in \Gamma$.

The sequent calculus was invented in the 1950s independently by Evert Beth and Jaakko Hintikka, and simplified and popularized by Raymond Smullyan. It is very easy to use, since constructing a tableau is a very systematic procedure. Because of the systematic nature of tableaux, they also lend themselves to implementation by computer. However, tableau is often hard to read and their connection to proofs are sometimes not easy to see. The approach is also quite general, and many different logics have tableau systems. Tableaux also help us to find structures that satisfy given (sets of) sentences: if the set is satisfiable, it won't have a closed tableau , i.e., any tableau will have an open branch. The satisfying structure can be "read off" an open branch, provided all rules it is possible to apply have been applied on that branch. There is also a very close connection to the sequent calculus: essentially, a closed tableau is a condensed derivation in the sequent calculus, written upside-down.

2.2 Rules and Tableaux

A tableau is a systematic survey of the possible ways a sentence can be true or false in a structure. The bulding blocks of a tableau are signed formulas: sentences plus a truth value "sign," either $\mathbb{T}$ or $\mathbb{F}$. These signed formulas are arranged in a (downward growing) tree.

Definition 2.1. A *signed formula* is a pair consisting of a truth value and a sentence , i.e., either:

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

Intuitively, we might read $\mathbb{T}\,A$ as "A might be true" and $\mathbb{F}\,A$ as "A might be false" (in some structure).

Each signed formula in the tree is either an *assumption* (which are listed at the very top of the tree), or it is obtained from a signed formula above it by one of a number of rules of inference. There are two rules for each possible main operator of the preceding formula , one for the case when the sign is $\mathbb{T}$, and one for the case where the sign is $\mathbb{F}$. Some rules allow the tree to branch, and some only add signed formulas to the branch. A rule may be (and often must be) applied not to the immediately preceding signed formula , but to any signed formula in the branch from the root to the place the rule is applied.

A branch is *closed* when it contains both $\mathbb{T}\,A$ and $\mathbb{F}\,A$. A closed tableau is one where every branch is closed. Under the intuitive interpretation, any branch describes a joint possibility, but $\mathbb{T}\,A$ and $\mathbb{F}\,A$ are not jointly possible. In other words, if a branch is closed, the possibility it describes has been ruled out. In particular, that means that a closed tableau rules out all possibilities of simultaneously making every assumption of the form $\mathbb{T}\,A$ true and every assumption of the form $\mathbb{F}\,A$ false.

A closed tableau *for* A is a closed tableau with root $\mathbb{F}\,A$. If such a closed tableau exists, all possibilities for A being false have been ruled out; i.e., A must be true in every structure.

2.3 Propositional Rules

Rules for ¬

$$\frac{\mathbb{T}\,\neg A}{\mathbb{F}\,A}\;\neg\mathbb{T} \qquad\qquad \frac{\mathbb{F}\,\neg A}{\mathbb{T}\,A}\;\neg\mathbb{F}$$

Rules for ∧

$$\frac{\mathbb{T}\,A \land B}{\begin{array}{c}\mathbb{T}\,A\\ \mathbb{T}\,B\end{array}}\;\land\mathbb{T} \qquad\qquad \frac{\mathbb{F}\,A \land B}{\mathbb{F}\,A \quad \mid \quad \mathbb{F}\,B}\;\land\mathbb{F}$$

Rules for ∨

$$\frac{\mathbb{T}\,A \lor B}{\mathbb{T}\,A \quad \mid \quad \mathbb{T}\,B}\;\lor\mathbb{T} \qquad\qquad \frac{\mathbb{F}\,A \lor B}{\begin{array}{c}\mathbb{F}\,A\\ \mathbb{F}\,B\end{array}}\;\lor\mathbb{F}$$

Rules for →

$$\frac{\mathbb{T}\,A \to B}{\mathbb{F}\,A \quad \mid \quad \mathbb{T}\,B}\;\to\mathbb{T} \qquad\qquad \frac{\mathbb{F}\,A \to B}{\begin{array}{c}\mathbb{T}\,A\\ \mathbb{F}\,B\end{array}}\;\to\mathbb{F}$$

The Cut Rule

$$\frac{}{\mathbb{T}\,A \quad \mid \quad \mathbb{F}\,A}\;\text{Cut}$$

The Cut rule is not applied "to" a previous signed formula; rather, it allows every branch in a tableau to be split in two, one branch containing $\mathbb{T}\,A$, the other $\mathbb{F}\,A$. It is not necessary—any set of signed formulas with a closed tableau has one not using Cut—but it allows us to combine tableaux in a convenient way.

2.4 Tableaux

We've said what an assumption is, and we've given the rules of inference. Tableaux are inductively generated from these: each tableau either is a single branch consisting of one or more assumptions, or it results from a tableau by applying one of the rules of inference on a branch.

Definition 2.2 (Tableau). A tableau for assumptions $S_1 A_1, \ldots, S_n A_n$ (where each S_i is either $\mathbb{T}$ or $\mathbb{F}$) is a tree of signed formulas satisfying the following conditions:

1. The n topmost signed formulas of the tree are $S_i A_i$, one below the other.

2. Every signed formula in the tree that is not one of the assumptions results from a correct application of an inference rule to a signed formula in the branch above it.

A branch of a tableau is *closed* iff it contains both $\mathbb{T}\,A$ and $\mathbb{F}\,A$, and *open* otherwise. A tableau in which every branch is closed is a *closed tableau* (for its set of assumptions). If a tableau is not closed, i.e., if it contains at least one open branch, it is *open*.

Example 2.3. Every set of assumptions on its own is a tableau but it will generally not be closed. (Obviously, it is closed only if the assumptions already contain a pair of signed formulas $\mathbb{T}\,A$ and $\mathbb{F}\,A$.)

From a tableau (open or closed) we can obtain a new, larger one by applying one of the rules of inference to a signed formula A in it. The rule will append one or more signed formulas to the end of any branch containing the occurrence of A to which we apply the rule.

For instance, consider the assumption $\mathbb{T}\,A \land \neg A$. Here is the (open) tableau consisting of just that assumption:

1.	$\mathbb{T}\,A \land \neg A$	Assumption

We obtain a new tableau from it by applying the $\land\mathbb{T}$ rule to the assumption. That rule allows us to add two new lines to the tableau , $\mathbb{T}\, A$ and $\mathbb{T}\,\neg A$:

1.	$\mathbb{T}\, A \land \neg A$	Assumption
2.	$\mathbb{T}\, A$	$\land\mathbb{T}\, 1$
3.	$\mathbb{T}\,\neg A$	$\land\mathbb{T}\, 1$

When we write down tableaux, we record the rules we've applied on the right (e.g., $\land\mathbb{T}1$ means that the signed formula on that line is the result of applying the $\land\mathbb{T}$ rule to the signed formula on line 1). This new tableau now contains additional signed formulas, but to only one ($\mathbb{T}\,\neg A$) can we apply a rule (in this case, the $\neg\mathbb{T}$ rule). This results in the closed tableau

1.	$\mathbb{T}\, A \land \neg A$	Assumption
2.	$\mathbb{T}\, A$	$\land\mathbb{T}\, 1$
3.	$\mathbb{T}\,\neg A$	$\land\mathbb{T}\, 1$
4.	$\mathbb{F}\, A$	$\neg\mathbb{T}\, 3$
	$\otimes$	

2.5 Examples of Tableaux

Example 2.4. Let's find a closed tableau for the sentence $(A \wedge B) \to A$.

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

1.	$\mathbb{F}\,(A \wedge B) \to A$	Assumption

There is only one assumption, so only one signed formula to which we can apply a rule. (For every signed formula , there is always at most one rule that can be applied: it's the rule for the corresponding sign and main operator of the sentence.) In this case, this means, we must apply $\to\mathbb{F}$.

1.	$\mathbb{F}\,(A \wedge B) \to A$ ✓	Assumption
2.	$\mathbb{T}\,A \wedge B$	$\to\mathbb{F}\,1$
3.	$\mathbb{F}\,A$	$\to\mathbb{F}\,1$

To keep track of which signed formulas we have applied their corresponding rules to, we write a checkmark next to the sentence. However, *only* write a checkmark if the rule has been applied to all open branches. Once a signed formula has had the corresponding rule applied in every open branch, we will not have to return to it and apply the rule again. In this case, there is only one branch, so the rule only has to be applied once. (Note that checkmarks are only a convenience for constructing tableaux and are not officially part of the syntax of tableaux.)

There is one new signed formula to which we can apply a rule: the $\mathbb{T}\,A \wedge B$ on line 3. Applying the $\wedge\mathbb{T}$ rule results in:

1.	$\mathbb{F}\,(A \wedge B) \to A$ ✓	Assumption
2.	$\mathbb{T}\,A \wedge B$ ✓	$\to\mathbb{F}\,1$
3.	$\mathbb{F}\,A$	$\to\mathbb{F}\,1$
4.	$\mathbb{T}\,A$	$\wedge\mathbb{T}\,2$
5.	$\mathbb{T}\,B$	$\wedge\mathbb{T}\,2$
	$\otimes$	

Since the branch now contains both $\mathbb{T}\,A$ (on line 4) and $\mathbb{F}\,A$ (on line 3), the branch is closed. Since it is the only branch, the tableau is closed. We have found a closed tableau for $(A \land B) \to A$.

Example 2.5. Now let's find a closed tableau for $(\neg A \lor B) \to (A \to B)$.

We begin with the corresponding assumption:

$$
\begin{array}{lcl}
1. & \mathbb{F}\,(\neg A \lor B) \to (A \to B) & \text{Assumption}
\end{array}
$$

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

$$
\begin{array}{lcl}
1. & \mathbb{F}\,(\neg A \lor B) \to (A \to B) \checkmark & \text{Assumption} \\
2. & \mathbb{T}\,\neg A \lor B & \to\mathbb{F}\,1 \\
3. & \mathbb{F}\,(A \to B) & \to\mathbb{F}\,1
\end{array}
$$

We now have a choice as to whether to apply $\lor\mathbb{T}$ to line 2 or $\to\mathbb{F}$ to line 3. It actually doesn't matter which order we pick, as long as each signed formula has its corresponding rule applied in every branch. So let's pick the first one. The $\lor\mathbb{T}$ rule allows the tableau to branch, and the two conclusions of the rule will be the new signed formulas added to the two new branches. This results in:

$$
\begin{array}{lcl}
1. & \mathbb{F}\,(\neg A \lor B) \to (A \to B) \checkmark & \text{Assumption} \\
2. & \mathbb{T}\,\neg A \lor B \checkmark & \to\mathbb{F}\,1 \\
3. & \mathbb{F}\,(A \to B) & \to\mathbb{F}\,1 \\
4. & \mathbb{T}\,\neg A \qquad \mathbb{T}\,B & \lor\mathbb{T}\,2
\end{array}
$$

We have not applied the $\to\mathbb{F}$ rule to line 3 yet: let's do that now. To save time, we apply it to both branches. Recall that we write a checkmark next to a signed formula only if we have applied the corresponding rule in every open branch. So it's a good idea to apply a rule at the end of every branch that contains the signed formula the rule applies to. That way we won't have to return to that signed formula lower down in the various branches.

1.	$\mathbb{F}\,(\neg A \lor B) \to (A \to B)$ ✓		Assumption
2.	$\mathbb{T}\,\neg A \lor B$ ✓		$\to\mathbb{F}\,1$
3.	$\mathbb{F}\,(A \to B)$ ✓		$\to\mathbb{F}\,1$
4.	$\mathbb{T}\,\neg A$	$\mathbb{T}\,B$	$\lor\mathbb{T}\,2$
5.	$\mathbb{T}\,A$	$\mathbb{T}\,A$	$\to\mathbb{F}\,3$
6.	$\mathbb{F}\,B$	$\mathbb{F}\,B$	$\to\mathbb{F}\,3$
		$\otimes$	

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

1.	$\mathbb{F}\,(\neg A \lor B) \to (A \to B)$ ✓		Assumption
2.	$\mathbb{T}\,\neg A \lor B$ ✓		$\to\mathbb{F}\,1$
3.	$\mathbb{F}\,(A \to B)$ ✓		$\to\mathbb{F}\,1$
4.	$\mathbb{T}\,\neg A$	$\mathbb{T}\,B$	$\lor\mathbb{T}\,2$
5.	$\mathbb{T}\,A$	$\mathbb{T}\,A$	$\to\mathbb{F}\,3$
6.	$\mathbb{F}\,B$	$\mathbb{F}\,B$	$\to\mathbb{F}\,3$
7.	$\mathbb{F}\,A$	$\otimes$	$\neg\mathbb{T}\,4$
	$\otimes$		

Example 2.6. We can give tableaux for any number of signed formulas as assumptions. Often it is also necessary to apply more than one rule that allows branching; and in general a tableau can have any number of branches. For instance, consider a tableau for $\{\mathbb{T}\,A \lor (B \land C), \mathbb{F}\,(A \lor B) \land (A \lor C)\}$. We start by applying the $\lor\mathbb{T}$ to the first assumption:

1.	$\mathbb{T}\,A \lor (B \land C)$ ✓		Assumption
2.	$\mathbb{F}\,(A \lor B) \land (A \lor C)$		Assumption
3.	$\mathbb{T}\,A$	$\mathbb{T}\,B \land C$	$\lor\mathbb{T}\,1$

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

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

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

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

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

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

Note that we moved the result of applying $\vee\mathbb{F}$ a second time below for clarity. In this instance it would not have been needed, since the justifications would have been the same.

Two branches remain open, and $\mathbb{T}\, B \wedge C$ on line 3 remains unchecked. We apply $\wedge\mathbb{T}$ to it to obtain a closed tableau :

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

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

$$
\begin{array}{llcccl}
1. & & \multicolumn{2}{c}{\mathbb{T}\, A \lor (B \land C) \checkmark} & & \text{Assumption} \\
2. & & \multicolumn{2}{c}{\mathbb{F}\, (A \lor B) \land (A \lor C) \checkmark} & & \text{Assumption} \\
3. & \multicolumn{2}{c}{\mathbb{F}\, A \lor B \checkmark} & \multicolumn{2}{c}{\mathbb{F}\, A \lor C \checkmark} & \land\mathbb{F}\, 2 \\
4. & \multicolumn{2}{c}{\mathbb{F}\, A} & \multicolumn{2}{c}{\mathbb{F}\, A} & \lor\mathbb{F}\, 3 \\
5. & \multicolumn{2}{c}{\mathbb{F}\, B} & \multicolumn{2}{c}{\mathbb{F}\, C} & \lor\mathbb{F}\, 3 \\
6. & \mathbb{T}\, A & \mathbb{T}\, B \land C \checkmark & \mathbb{T}\, A & \mathbb{T}\, B \land C \checkmark & \lor\mathbb{T}\, 1 \\
7. & \otimes & \mathbb{T}\, B & \otimes & \mathbb{T}\, B & \land\mathbb{T}\, 3 \\
8. & & \mathbb{T}\, C & & \mathbb{T}\, C & \land\mathbb{T}\, 3 \\
 & & \otimes & & \otimes &
\end{array}
$$

2.6 Proof-Theoretic Notions

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

Definition 2.7 (Theorems). A sentence A is a *theorem* if there is a closed tableau for $\mathbb{F}\, A$. We write $\vdash A$ if A is a theorem and $\nvdash A$ if it is not.

Definition 2.8 (Derivability). A sentence A is *derivable from* a set of sentences Γ, $\Gamma \vdash A$, iff there is a finite set $\{B_1, \ldots, B_n\} \subseteq \Gamma$ and a closed tableau for the set

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

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

Definition 2.9 (Consistency). A set of sentences Γ is *inconsistent* iff there is a finite set $\{B_1, \ldots, B_n\} \subseteq \Gamma$ and a closed tableau for the set

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

If Γ is not inconsistent, we say it is *consistent.*

Proposition 2.10 (Reflexivity). *If $A \in \Gamma$, then $\Gamma \vdash A$.*

Proof. If $A \in \Gamma$, $\{A\}$ is a finite subset of Γ and the tableau

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

is closed. □

Proposition 2.11 (Monotony). *If $\Gamma \subseteq \Delta$ and $\Gamma \vdash A$, then $\Delta \vdash A$.*

Proof. Any finite subset of Γ is also a finite subset of Δ. □

Proposition 2.12 (Transitivity). *If $\Gamma \vdash A$ and $\{A\} \cup \Delta \vdash B$, then $\Gamma \cup \Delta \vdash B$.*

Proof. If $\{A\} \cup \Delta \vdash B$, then there is a finite subset $\Delta_0 = \{C_1, \ldots, C_n\} \subseteq \Delta$ such that

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

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

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

has a closed tableau.

Now consider the tableau with assumptions

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

Apply the Cut rule on A. This generates two branches, one has $\mathbb{T}\,A$ in it, the other $\mathbb{F}\,A$. Thus, on the one branch, all of

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

are available. Since there is a closed tableau for these assumptions, we can attach it to that branch; every branch through $\mathbb{T}\,A_1$ closes. On the other branch, all of

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

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

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

Proposition 2.13. *Γ is inconsistent iff $\Gamma \vdash A$ for every sentence A.*

Proof. Exercise. □

Proposition 2.14 (Compactness).

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

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

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

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

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

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

This closed tableau shows that Γ_0 is inconsistent.

□

2.7 Derivability and Consistency

We will now establish a number of properties of the derivability relation. They are independently interesting, but each will play a role in the proof of the completeness theorem.

Proposition 2.15. *If $\Gamma \vdash A$ and $\Gamma \cup \{A\}$ is inconsistent, then Γ is inconsistent.*

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

$$\{\mathbb{F}\, A, \mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n\}$$
$$\{\mathbb{T}\, \neg A, \mathbb{T}\, C_1, \ldots, \mathbb{T}\, C_m\}$$

have closed tableaux. Using the Cut rule on A we can combine these into a single closed tableau that shows $\Gamma_0 \cup \Gamma_1$ is inconsistent. Since $\Gamma_0 \subseteq \Gamma$ and $\Gamma_1 \subseteq \Gamma$, $\Gamma_0 \cup \Gamma_1 \subseteq \Gamma$, hence Γ is inconsistent. □

Proposition 2.16. *$\Gamma \vdash A$ iff $\Gamma \cup \{\neg A\}$ is inconsistent.*

Proof. First suppose $\Gamma \vdash A$, i.e., there is a closed tableau for

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

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

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

On the other hand, if there is a closed tableau for the latter, we can turn it into a closed tableau of the former by removing every formula that results from $\neg\mathbb{T}$ applied to the first assumption $\mathbb{T}\, \neg A$ as well as that assumption, and adding the assumption $\mathbb{F}\, A$. For if a branch was closed before because it contained the conclusion of $\neg\mathbb{T}$ applied to $\mathbb{T}\, \neg A$, i.e., $\mathbb{F}\, A$, the corresponding branch in the new tableau is also closed. If a branch in the old tableau was closed because it contained the assumption $\mathbb{T}\, \neg A$ as well as $\mathbb{F}\, \neg A$ we can turn it into a closed branch by applying $\neg\mathbb{F}$ to $\mathbb{F}\, \neg A$ to obtain $\mathbb{T}\, A$. This closes the branch since we added $\mathbb{F}\, A$ as an assumption. □

Proposition 2.17. *If $\Gamma \vdash A$ and $\neg A \in \Gamma$, then Γ is inconsistent.*

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

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

has a closed tableau. Replace the assumption $\mathbb{F}\, A$ by $\mathbb{T}\, \neg A$, and insert the conclusion of $\neg\mathbb{T}$ applied to $\mathbb{F}\, A$ after the assumptions. Any sentence in the tableau justified by appeal to line 1 in the old tableau is now justified by appeal to line $n + 1$. So if the old tableau was closed, the new one is. It shows that Γ is inconsistent, since all assumptions are in Γ. □

Proposition 2.18. *If $\Gamma \cup \{A\}$ and $\Gamma \cup \{\neg A\}$ are both inconsistent, then Γ is inconsistent.*

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

$$\{\mathbb{T}\,A, \mathbb{T}\,B_1, \ldots, \mathbb{T}\,B_n\}$$
$$\{\mathbb{T}\,\neg A, \mathbb{T}\,C_1, \ldots, \mathbb{T}\,C_m\}$$

both have closed tableaux, we can construct a tableau that shows that Γ is inconsistent by using as assumptions $\mathbb{T}\,B_1, \ldots, \mathbb{T}\,B_n$ together with $\mathbb{T}\,C_1, \ldots, \mathbb{T}\,C_m$, followed by an application of the Cut rule, yielding two branches, one starting with $\mathbb{T}\,A$, the other with $\mathbb{F}\,A$. Add on the part below the assumptions of the first tableau on the left side. Here, every rule application is still correct, and every branch closes. On the right side, add the part below the assumptions of the second tableau, with the results of any applications of $\neg\mathbb{T}$ to $\mathbb{T}\,\neg A$ removed.

For if a branch was closed before because it contained the conclusion of $\neg\mathbb{T}$ applied to $\mathbb{T}\,\neg A$, i.e., $\mathbb{F}\,A$, as well as $\mathbb{F}\,A$, the corresponding branch in the new tableau is also closed. If a branch in the old tableau was closed because it contained the assumption $\mathbb{T}\,\neg A$ as well as $\mathbb{F}\,\neg A$ we can turn it into a closed branch by applying $\neg\mathbb{F}$ to $\mathbb{F}\,\neg A$ to obtain $\mathbb{T}\,A$. □

2.8 Derivability and the Propositional Connectives

Proposition 2.19.

1. *Both* $A \wedge B \vdash A$ *and* $A \wedge B \vdash B$.

2. $A, B \vdash A \wedge B$.

Proof. 1. Both $\{\mathbb{F}\,A, \mathbb{T}\,A \wedge B\}$ and $\{\mathbb{F}\,B, \mathbb{T}\,A \wedge B\}$ have closed tableaux

$$\begin{array}{lcl}
1. & \mathbb{F}\,A & \text{Assumption} \\
2. & \mathbb{T}\,A \wedge B & \text{Assumption} \\
3. & \mathbb{T}A & \wedge\mathbb{T}\,2 \\
4. & \mathbb{T}B & \wedge\mathbb{T}\,2 \\
 & \otimes &
\end{array}$$

$$\begin{array}{lcl}
1. & \mathbb{F}\,B & \text{Assumption} \\
2. & \mathbb{T}\,A \wedge B & \text{Assumption} \\
3. & \mathbb{T}A & \wedge\mathbb{T}\,2 \\
4. & \mathbb{T}B & \wedge\mathbb{T}\,2 \\
 & \otimes &
\end{array}$$

2. Here is a closed tableau for $\{\mathbb{T}\,A, \mathbb{T}\,B, \mathbb{F}\,A \wedge B\}$:

$$\begin{array}{lccl}
1. & \multicolumn{2}{c}{\mathbb{F}\,A \wedge B} & \text{Assumption} \\
2. & \multicolumn{2}{c}{\mathbb{T}\,A} & \text{Assumption} \\
3. & \multicolumn{2}{c}{\mathbb{T}\,B} & \text{Assumption} \\
 & \diagup & \diagdown & \\
4. & \mathbb{F}\,A & \mathbb{F}\,B & \wedge\mathbb{F}\,1 \\
 & \otimes & \otimes &
\end{array}$$

□

Proposition 2.20.

1. *$A \vee B, \neg A, \neg B$ is inconsistent.*
2. *Both $A \vdash A \vee B$ and $B \vdash A \vee B$.*

Proof. 1. We give a closed tableau of $\{\mathbb{T}\, A \vee B, \mathbb{T}\, \neg A, \mathbb{T}\, \neg B\}$:

$$
\begin{array}{lccl}
1. & \multicolumn{2}{c}{\mathbb{T}\, A \vee B} & \text{Assumption} \\
2. & \multicolumn{2}{c}{\mathbb{T}\, \neg A} & \text{Assumption} \\
3. & \multicolumn{2}{c}{\mathbb{T}\, \neg B} & \text{Assumption} \\
4. & \multicolumn{2}{c}{\mathbb{F}\, A} & \neg\mathbb{T}\, 2 \\
5. & \multicolumn{2}{c}{\mathbb{F}\, B} & \neg\mathbb{T}\, 3 \\
6. & \mathbb{T}\, A & \mathbb{T}\, B & \vee\mathbb{T}\, 1 \\
 & \otimes & \otimes &
\end{array}
$$

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

$$
\begin{array}{lcl}
1. & \mathbb{F}\, A \vee B & \text{Assumption} \\
2. & \mathbb{T}\, A & \text{Assumption} \\
3. & \mathbb{F} A & \vee\mathbb{F}\, 1 \\
4. & \mathbb{F} B & \vee\mathbb{F}\, 1 \\
 & \otimes &
\end{array}
$$

$$
\begin{array}{lcl}
1. & \mathbb{F}\, A \vee B & \text{Assumption} \\
2. & \mathbb{T}\, B & \text{Assumption} \\
3. & \mathbb{F} A & \vee\mathbb{F}\, 1 \\
4. & \mathbb{F} B & \vee\mathbb{F}\, 1 \\
 & \otimes &
\end{array}
$$

□

Proposition 2.21.

1. *$A, A \to B \vdash B$.*
2. *Both $\neg A \vdash A \to B$ and $B \vdash A \to B$.*

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

1.	$\mathbb{F}\, B$	Assumption
2.	$\mathbb{T}\, A \to B$	Assumption
3.	$\mathbb{T}\, A$	Assumption
4.	$\mathbb{F}\, A$ $\quad$ $\mathbb{T}\, B$	$\to\mathbb{T}\,2$
	$\otimes$ $\quad$ $\otimes$	

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

1.	$\mathbb{F}\, A \to B$	Assumption
2.	$\mathbb{T}\,\neg A$	Assumption
3.	$\mathbb{T}\, A$	$\to\mathbb{F}\,1$
4.	$\mathbb{F}\, B$	$\to\mathbb{F}\,1$
5.	$\mathbb{F}\, A$	$\neg\mathbb{T}\,2$
	$\otimes$	

1.	$\mathbb{F}\, A \to B$	Assumption
2.	$\mathbb{T}\,\neg B$	Assumption
3.	$\mathbb{T}\, A$	$\to\mathbb{F}\,1$
4.	$\mathbb{F}\, B$	$\to\mathbb{F}\,1$
5.	$\mathbb{F}\, B$	$\neg\mathbb{T}\,2$
	$\otimes$	

$\square$

2.9 Soundness

A derivation system, such as tableaux, is *sound* if it cannot derive things that do not actually hold. Soundness is thus a kind of guaranteed safety property for derivation systems. Depending on which proof theoretic property is in question, we would like to know for instance, that

1. every derivable A is a tautology;
2. if a sentence is derivable from some others, it is also a consequence of them;
3. if a set of sentences is inconsistent, it is unsatisfiable.

These are important properties of a derivation system. If any of them do not hold, the derivation system is deficient—it would derive too much. Consequently, establishing the soundness of a derivation system is of the utmost importance.

Because all these proof-theoretic properties are defined via closed tableaux of some kind or other, proving (1)–(3) above requires proving something about the semantic properties of closed tableaux. We will first define what it means for a signed formula to be satisfied in a structure, and then show that if a tableau is closed, no structure satisfies all its assumptions. (1)–(3) then follow as corollaries from this result.

Definition 2.22. A valuation v *satisfies* a signed formula $\mathbb{T}\,A$ iff $v \vDash A$, and it satisfies $\mathbb{F}\,A$ iff $v \nvDash A$. v satisfies a set of signed formulas Γ iff it satisfies every $S\,A \in \Gamma$. Γ is *satisfiable* if there is a valuation that satisfies it, and *unsatisfiable* otherwise.

Theorem 2.23 (Soundness). *If Γ has a closed tableau , Γ is unsatisfiable.*

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

We show the following: Extending a satisfiable tableau by one of the rules of inference always results in a satisfiable tableau. This will prove the theorem: any closed tableau results by applying rules of inference to the tableau consisting only of assumptions from Γ. So if Γ were satisfiable, any tableau for it would be satisfiable. A closed tableau , however, is clearly not satisfiable: every branch contains both $\mathbb{T}\, A$ and $\mathbb{F}\, A$, and no structure can both satisfy and not satisfy A.

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

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

First, we consider the possible inferences with only one premise.

1. The branch is expanded by applying $\neg\mathbb{T}$ to $\mathbb{T}\,\neg B \in \Gamma$. Then the extended branch contains the signed formulas $\Gamma \cup \{\mathbb{F}\, B\}$. Suppose $v \vDash \Gamma$. In particular, $v \vDash \neg B$. Thus, $v \nvDash B$, i.e., v satisfies $\mathbb{F}\, B$.

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

3. The branch is expanded by applying $\wedge\mathbb{T}$ to $\mathbb{T}\, B \wedge C \in \Gamma$, which results in two new signed formulas on the branch: $\mathbb{T}\, B$ and $\mathbb{T}\, C$. Suppose $v \vDash \Gamma$, in particular $v \vDash B \wedge C$. Then $v \vDash B$ and $v \vDash C$. This means that v satisfies both $\mathbb{T}\, B$ and $\mathbb{T}\, C$.

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

5. The branch is expanded by applying $\to\mathbb{F}$ to $\mathbb{T}\, B \to C \in \Gamma$: This results in two new signed formulas on the branch: $\mathbb{T}\, B$ and $\mathbb{F}\, C$. Suppose $v \vDash \Gamma$, in particular $v \nvDash B \to C$. Then $v \vDash B$ and $v \nvDash C$. This means that v satisfies both $\mathbb{T}\, B$ and $\mathbb{F}\, C$.

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

1. The branch is expanded by applying $\wedge\mathbb{F}$ to $\mathbb{F}\, B \wedge C \in \Gamma$, which results in two branches, a left one continuing through $\mathbb{F}\, B$ and a right one through $\mathbb{F}\, C$. Suppose $v \vDash \Gamma$, in particular $v \nvDash B \wedge C$. Then $v \nvDash B$ or $v \nvDash C$. In the former case, v satisfies $\mathbb{F}\, B$, i.e., v satisfies the formulas on the left branch. In the latter, v satisfies $\mathbb{F}\, C$, i.e., v satisfies the formulas on the right branch.

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

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

4. The branch is expanded by Cut: This results in two branches, one containing $\mathbb{T}\, B$, the other containing $\mathbb{F}\, B$. Since $v \vDash \Gamma$ and either $v \vDash B$ or $v \nvDash B$, v satisfies either the left or the right branch.

□

Corollary 2.24. *If* $\vdash A$ *then* A *is a tautology.*

Corollary 2.25. *If* $\Gamma \vdash A$ *then* $\Gamma \vDash A$.

Proof. If $\Gamma \vdash A$ then for some $B_1, \ldots, B_n \in \Gamma$, $\{\mathbb{F}\, A, \mathbb{T}\, B_1, \ldots, \mathbb{T}\, B_n\}$ has a closed tableau . By Theorem 2.23, every valuation v either makes some B_i false or makes A true. Hence, if $v \vDash \Gamma$ then also $v \vDash A$. □

Corollary 2.26. *If* Γ *is satisfiable, then it is consistent.*

Proof. We prove the contrapositive. Suppose that Γ is not consistent. Then there are $B_1, \ldots, B_n \in \Gamma$ and a closed tableau for $\{\mathbb{T}\, B, \ldots, \mathbb{T}\, B\}$. By Theorem 2.23, there is no v such that $v \vDash B_i$ for all $i = 1, \ldots, n$. But then Γ is not satisfiable. □

Problems

Problem 2.1. Give closed tableaux of the following:

1. $\mathbb{F}\, \neg(A \to B) \to (A \land \neg B)$

2. $\mathbb{F}\, (A \to C) \lor (B \to C), \mathbb{T}\, (A \land B) \to C$

Problem 2.2. Prove Proposition 2.13

Problem 2.3. Prove that $\Gamma \vdash \neg A$ iff $\Gamma \cup \{A\}$ is inconsistent.

Problem 2.4. Complete the proof of Theorem 2.23.

PART II

Normal Modal Logics

CHAPTER 3

Syntax and Semantics of Normal Modal Logics

3.1 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 $A \land B$ is possibly true then A is possibly true and B is possibly true, but not conversely; and that if $A \to B$ is necessary, then if A is necessary, so is B.

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

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

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

When interpreted by the Kripke semantics, modal logic shows us what *relational structures* look like "from the inside." A relational structure is just a set equipped with a binary relation (for instance, the set of students in the class ordered by their social security number is a relational structure). But in fact relational structures come in all sorts of domains: besides relative possibility of states of the world, we can have epistemic states of some agent related by epistemic possibility, or states of a dynamical system with their state transitions, etc. Modal logic can be used to model all of these: the first give us ordinary, alethic, modal logic; the others give us epistemic logic, dynamic logic, etc.

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

3.2 The Language of Basic Modal Logic

Definition 3.1. The basic language of modal logic contains

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

Definition 3.2. *Formulas* of the basic modal language are inductively defined as follows:

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

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

Some more technical notions about the language, in particular the notion of a *substitution*, an *instance* and a *tautological in-*

stance, are described at the end of this chapter. For now, it suffices to know that these notions work roughly as you would expect; the discussion at the end shows how to turn these informal expectations into a formal theory.

3.3 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 3.3. A *model* for the basic modal language is a triple $M = \langle W, R, V \rangle$, where

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

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

The great advantage of relational semantics is that models can be represented by means of simple diagrams, such as the one in Figure 3.1. 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 3.1 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 $V(q) = \{w_2\}$.

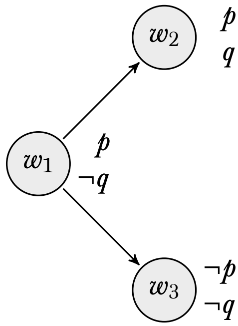


Figure 3.1: A simple model.

3.4 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 3.4. *Truth of a formula* A *at* w in a $\mathbf{M}$, in symbols: $\mathbf{M}, w \Vdash A$, is defined inductively as follows:

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

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

Proposition 3.5.

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

Proof.

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

2. Exercise.

□

3.5 Truth in a Model

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

Definition 3.6. 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 3.7. 1. *If $M \Vdash A$ then $M \nVdash \neg A$, but* not *vice-versa.*

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

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

On the other hand, if $M \nVdash \neg A$ then A is true at some world $w \in W$. It does not follow that $M, w \Vdash A$ for *every* $w \in W$. For instance, in the model of Figure 3.1, $M \nVdash \neg p$, and also $M \nVdash p$.

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

To show that the converse fails, we need to find a model M such that $M \Vdash A$ only if $M \Vdash B$, but $M \nVdash A \to B$. Consider again the model of Figure 3.1: $M \nVdash p$ and hence (vacuously) $M \Vdash p$ only if $M \Vdash q$. However, $M \nVdash p \to q$, as p is true but q false at w_1.

□

3.6 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 3.8. 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 3.9. *If A is valid in $\mathscr{C}$ it is also valid in each class $\mathscr{C}' \subseteq \mathscr{C}$.*

Proposition 3.10. *If A is valid, then so is $\Box A$.*

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

3.7 Schemas and Validity

Definition 3.11. A *schema* is a set of formulas comprising all and only the substitution instances of some modal formula C, i.e.,

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

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

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

Definition 3.12. 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 3.13. *The following schema* K *is valid*

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

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 consequent is true as well. In this case, let $M, w \Vdash \Box(A \rightarrow B)$ and $M, w \Vdash \Box A$. We need to show $M \Vdash \Box B$. So let w' be arbitrary such that Rww'. Then by the first assumption $M, w' \Vdash A \rightarrow B$ and by the second assumption $M, w' \Vdash A$. It follows that $M, w' \Vdash B$. Since w' was arbitrary, $M, w \Vdash \Box B$. □

Proposition 3.14. *The following schema* DUAL *is valid*

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

Proof. Exercise. □

Proposition 3.15. *If A and $A \rightarrow B$ are true at a world in a model then so is B. Hence, the valid formulas are closed under modus ponens.*

Proposition 3.16. *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 $\mathfrak{M} = \langle W, R, V\rangle$ is a modal model, and $B \equiv A[D_1/p_1, \dots, D_n/p_n]$ is a substitution instance of A. Define $\mathfrak{M}' = \langle W, R, V'\rangle$ by $V(p_i) = \{w : \mathfrak{M} \Vdash D_i w\}$. Then $\mathfrak{M} \Vdash Bw$ iff $\mathfrak{M}' \Vdash Aw$, for any $w \in W$. (We leave the proof as an exercise.) Now suppose that A was valid, but some substitution instance B of A was not valid. Then for some $\mathfrak{M} = \langle W, R, V\rangle$ and some $w \in W$, $\mathfrak{M} \nVdash Bw$. But then $\mathfrak{M}' \nVdash Aw$ 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 $\mathfrak{M}$, A itself is true in $\mathfrak{M}$. But it may happen that A is true in $\mathfrak{M}$ but some instance of A is false at some world in $\mathfrak{M}$. 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.

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 3.1: Valid and (or?) invalid schemas.

3.8 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 3.17. 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 $\mathbf{M} = \langle W, R, V\rangle$ and world $w \in W$, if $\mathbf{M}, w \Vdash B$ for every $B \in \Gamma$, then $\mathbf{M}, w \Vdash A$. If Γ contains a single formula B, then we write $B \vDash A$.

Example 3.18. To show that a formula entails another, we have to reason about all models, using the definition of $\mathbf{M}, w \Vdash$. For instance, to show $p \to \Diamond p \vDash \Box\neg p \to \neg p$, we might argue as follows: Consider a model $\mathbf{M} = \langle W, R, V\rangle$ and $w \in W$, and suppose $\mathbf{M}, w \Vdash p \to \Diamond p$. We have to show that $\mathbf{M}, w \Vdash \Box\neg p \to \neg p$. Suppose not. Then $\mathbf{M}, w \Vdash \Box\neg p$ and $\mathbf{M}, w \nVdash \neg p$. Since $\mathbf{M}, w \nVdash \neg p$, $\mathbf{M}, w \Vdash p$. By assumption, $\mathbf{M}, w \Vdash p \to \Diamond p$, hence $\mathbf{M}, w \Vdash \Diamond p$. By definition of $\mathbf{M}, w \Vdash \Diamond p$, there is some w' with Rww' such that $\mathbf{M}, w' \Vdash p$. Since also $\mathbf{M}, w \Vdash \Box\neg p$, $\mathbf{M}, w' \Vdash \neg p$, a contradiction.

To show that a formula B does not entail another A, we have to give a counterexample, i.e., a model $\mathbf{M} = \langle W, R, V\rangle$ where we show that at some world $w \in W$, $\mathbf{M}, w \Vdash B$ but $\mathbf{M}, w \nVdash A$. Let's

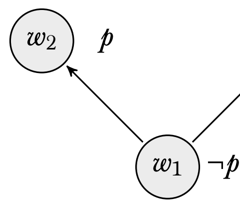


Figure 3.2: Counterexample to $p \to \Diamond p \vDash \Box p \to p$.

show that $p \to \Diamond p \nvDash \Box p \to p$. Consider the model in Figure 3.2. 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 $M', w \nVdash \Box p \to p$.

3.9 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 3.19. Where A is a modal formula all of whose propositional variables are among $p_1, \dots, p_n$, and $D_1, \dots, D_n$ are also modal formulas, we define $A[D_1/p_1, \dots, D_n/p_n]$ as the result of simultaneously substituting each D_i for p_i in A. Formally, this is a definition by induction on A:

1. $A \equiv \bot$: $\bot[D_1/p_1, \dots, D_n/p_n]$ is $\bot$.

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

3. $A \equiv p_i$: $p_i[D_1/p_1, \dots, D_n/p_n]$ is D_i.

4. $A \equiv \neg B$: $\neg B[D_1/p_1, \dots, D_n/p_n]$ is $\neg B[D_1/p_1, \dots, D_n/p_n]$.

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

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

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

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

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

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

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

Example 3.20. 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 $A[D_1/p_1][D_2/p_2]$:

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

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

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

3.10 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 3.21. A modal formula B is a *tautological instance* if and only if there is a modal-free tautology A with propositional variables $p_1, \ldots, p_n$ and formulas $D_1, \ldots, D_n$ such that $B \equiv A[D_1/p_1, \ldots, D_n/p_n]$.

Lemma 3.22. *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. $A \equiv p_i$:

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

3. $A \equiv \neg B$:

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

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

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

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

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

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

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

□

Proposition 3.23. *All tautological instances are valid.*

Proof. Contrapositively, suppose A is such that $\mathfrak{M}, w \nVdash A[D_1/p_1, \dots, D_n/p_n]$, for some model $\mathfrak{M}$ and world w. Define an assignment v such that $v(p_i) = \mathbb{T}$ if and only if $\mathfrak{M}, w \Vdash D_i$ (and v assigns arbitrary values to $q \notin \{p_1, \dots, p_n\}$). Then by Lemma 3.22, $v \nvDash A$, so A is not a tautology. □

Problems

Problem 3.1. Consider the model of Figure 3.1. Which of the following hold?

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

8. $\mathfrak{M}, w_1 \Vdash \Box q$;

9. $\mathfrak{M}, w_1 \Vdash \neg\Box\Box\neg q$.

Problem 3.2. Complete the proof of Proposition 3.5.

Problem 3.3. Let $\mathfrak{M} = \langle W, R, V\rangle$ be a model, and suppose $w_1, w_2 \in W$ are such that:

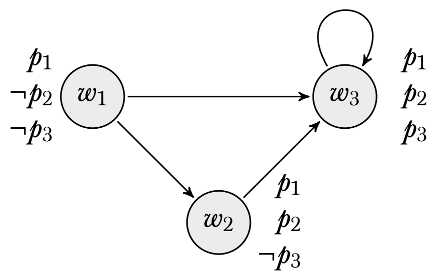
1. $w_1 \in V(p)$ if and only if $w_2 \in V(p)$; and

2. for all $w \in W$: Rw_1w if and only if Rw_2w.

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

Problem 3.4. Let $\mathfrak{M} = \langle M, R, V\rangle$. Show that $\mathfrak{M}, w \Vdash \neg\Diamond A$ if and only if $\mathfrak{M}, w \Vdash \Box\neg A$.

Problem 3.5. Consider the following model $\mathfrak{M}$ 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 $\mathfrak{M}$, i.e., true at every world in $\mathfrak{M}$? 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 3.6. Show that the following are valid:

1. $\vDash \Box p \rightarrow \Box(q \rightarrow p)$;

2. $\vDash \Box\neg\bot$;

3. $\vDash \Box p \rightarrow (\Box q \rightarrow \Box p)$.

Problem 3.7. Show that $A \rightarrow \Box A$ is valid in the class $\mathscr{C}$ of models $\mathfrak{M} = \langle W, R, V \rangle$ where $W = \{w\}$. Similarly, show that $B \rightarrow \Box A$ and $\Diamond A \rightarrow B$ are valid in the class of models $\mathfrak{M} = \langle W, R, V \rangle$ where $R = \emptyset$.

Problem 3.8. Prove Proposition 3.14.

Problem 3.9. Prove the claim in the "only if" part of the proof of Proposition 3.16. (Hint: use induction on A.)

Problem 3.10. Show that none of the following formulas are valid:

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

T: $\Box p \rightarrow p$;

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

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

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

Problem 3.11. Prove that the schemas in the first column of table 3.1 are valid and those in the second column are not valid.

Problem 3.12. Decide whether the following schemas are valid or invalid:

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

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

Problem 3.13. For each of the following schemas find a model M such that every instance of the formula is true in M:

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

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

Problem 3.14. Show that $\Box(A \land B) \vDash \Box A$.

Problem 3.15. Show that $\Box(p \rightarrow q) \nvDash p \rightarrow \Box q$ and $p \rightarrow \Box q \nvDash \Box(p \rightarrow q)$.

CHAPTER 4

Frame Definability

4.1 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 $\mathfrak{M}$ 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 $\mathfrak{M}$).

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

4.2 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 4.1. *Let $M = \langle W, R, V\rangle$ be a model. If R has the property on the left side of table 4.1, 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 all worlds

If R is ...	*then ... is true in $\mathbf{M}$:*
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 4.1: Five correspondence facts.

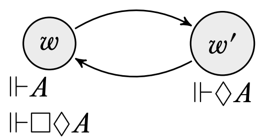


Figure 4.1: The argument from symmetry.

in the model. So let $A \to \Box\Diamond A$ be a given instance of B, 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 4.1). Since $\mathbf{M}, w \Vdash A$, we have $\mathbf{M}, w' \Vdash \Diamond A$. Since w' was an arbitrary world such that Rww', we have $\mathbf{M}, w \Vdash \Box\Diamond A$.

We leave the other cases as exercises. □

Notice that the converse implications of Theorem 4.1 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 $\mathbf{M}, w \Vdash \Box p$ and $\mathbf{M}, w \nVdash p$. But here we have just a single instance of T that fails in $\mathbf{M}$, 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 $\mathbf{M}$ and $\mathbf{M}$ is not reflexive. But there are such models, too:

Proposition 4.2. *Let $\mathbf{M} = \langle W, R, V\rangle$ 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: $\mathbf{M}, u \Vdash A$ if and only if $\mathbf{M}, v \Vdash A$ (use induction on A).*

2. *Every instance of* T *is true in $\mathbf{M}$.*

Since $\mathbf{M}$ is not reflexive (it is, in fact, irreflexive*), the converse of Theorem 4.1 fails in the case of* T *(similar arguments can be given for some—though not all—the other schemas mentioned in Theorem 4.1).*

Although we will focus on the five classical formulas D, T, B, 4, and 5, we record in table 4.2 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: $\forall w \forall u \forall v((Rwu \land Rwv) \to u = v)$	$\Diamond p \to \Box p$
functional: $\forall w \exists v \forall u(Rwu \leftrightarrow u = v)$	$\Diamond p \leftrightarrow \Box p$
weakly dense: $\forall u \forall v(Ruv \to \exists w(Ruw \land Rwv))$	$\Box\Box p \to \Box p$
weakly connected: $\forall w \forall u \forall v((Rwu \land Rwv) \to (Ruv \lor u = v \lor Rvu))$	$\Box((p \land \Box p) \to q) \lor \Box((q \land \Box q) \to p)$ (L)
weakly directed: $\forall w \forall u \forall v((Rwu \land Rwv) \to \exists t(Rut \land Rvt)$	$\Diamond\Box p \to \Box\Diamond p$ (G)

Table 4.2: Five more correspondence facts.

4.3 Frames

Definition 4.3. A *frame* is a pair $F = \langle W, R \rangle$ where W is a non-empty set of worlds and R a binary relation on W. A model M is *based on* a frame $F = \langle W, R \rangle$ if and only if $M = \langle W, R, V \rangle$.

Definition 4.4. If F is a frame, we say that A is *valid in* F, $F \vDash A$, if $M \Vdash A$ for every model M based on F.

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

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

Remark 1. Validity in a class of frames is a special case of the notion of validity in a class of models: $\mathcal{F} \vDash A$ 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.

4.4 Frame Definability

Even though the converse implications of Theorem 4.1 fail, they hold if we replace "model" by "frame": for the properties considered in Theorem 4.1, 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 4.5. If $\mathcal{C}$ is a class of frames, we say A *defines* $\mathcal{C}$ iff $\boldsymbol{F} \vDash A$ for all and only frames $\boldsymbol{F} \in \mathcal{C}$.

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

Theorem 4.6. *If the formula on the right side of table 4.1 is valid in a frame* $\mathbf{F}$*, then* $\mathbf{F}$ *has the property on the left side.*

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

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

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

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

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

5. We proceed contrapositively, assuming that the frame $F = \langle W, R \rangle$ is not euclidean, and show that it falsifies 5, i.e., i.e., $F \nvDash \Diamond p \rightarrow \Box\Diamond p$. Suppose there are worlds $u, v, w \in W$ such that Rwu and Rwv but not Ruv. Define V such that for all worlds z, $z \in V(p)$ if and only if it is *not* the case that Ruz. Let $M = \langle W, R, V \rangle$. 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, $\Diamond p \rightarrow \Box\Diamond p$, fails at w.

□

You'll notice a difference between the proof for D and the other cases: no mention was made of the valuation V. In effect, we proved that if $\mathbf{M} \Vdash \mathrm{D}$ then $\mathbf{M}$ is serial. So D defines the class of serial *models*, not just frames.

Corollary 4.7. *Any model where* D *is true is serial.*

Corollary 4.8. *Each formula on the right side of table 4.1 defines the class of frames which have the property on the left side.*

Proof. In Theorem 4.1, 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 $\mathbf{F}$ has the property on the left, the formula on the left is valid in $\mathbf{F}$. In Theorem 4.6, we proved the converse implications: if a formula on the right is valid in $\mathbf{F}$, $\mathbf{F}$ has the property on the left. □

Theorem 4.6 also shows that the properties can be combined: for instance if both B and 4 are valid in $\mathbf{F}$ 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 $\mathbf{M}, w \Vdash \mathrm{T}$ then $\mathbf{M}, w \Vdash \mathrm{D}$.) We record some such relationships.

Proposition 4.9. *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.*

4.5 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 4.10. 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 four 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 4.11. *The following are equivalent:*

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

Proof. Exercise. □

Proposition 4.11 is the semantic counterpart to Proposition 4.13, in that it gives equivalent characterization of the modal logic of frames over which R is an equivalence (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 4.12. *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 4.13. *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 $M = \langle W, R, V\rangle$ based on a frame $\langle W, R\rangle$, where R is an equivalence on W. So $M, w \nVdash B$. Define a model $M' = \langle W', R', V'\rangle$ as follows:

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

(So the set W' of worlds in M' is represented by the shaded area in Figure 4.2.) It is easy to see that R and R' agree on W'. Then one can show by induction on formulas that for all $w' \in W'$: $M', w' \Vdash A$ if and only if $M, w' \Vdash A$ for each A (this makes sense since $W' \subseteq W$). In particular, $M', w \nVdash B$, and B fails in a model based on a universal frame. □

Problems

Problem 4.1. Complete the proof of Theorem 4.1

Problem 4.2. Prove the claims in Proposition 4.2.

Problem 4.3. Let $M = \langle W, R, V\rangle$ be a model. Show that if R satisfies the left-hand properties of table 4.2, every instance of the corresponding right-hand formula is true in M.

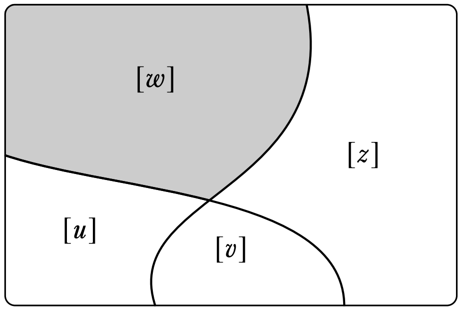


Figure 4.2: A partition of W in equivalence classes.

Problem 4.4. Show that if the formula on the right side of table 4.2 is valid in a frame F, then F has the property on the left side. To do this, consider a frame that does *not* satisfy the property on the left, and define a suitable V such that the formula on the right is false at some world.

Problem 4.5. Prove Proposition 4.9.

Problem 4.6. Prove Proposition 4.11 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 5

Modal Tableaux

5.1 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 consitituës 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

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

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

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

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

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

5.2 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 section 5.2.

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.

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

Table 5.1: Prefixed tableau rules for the propositional connectives

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

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

iff there is a closed tableau for the assumptions

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

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

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

$\dfrac{\sigma\,\mathbb{T}\,\Box A}{\sigma.n\,\mathbb{T}\,A}\,\Box\mathbb{T}$ $\sigma.n$ 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 5.2: The modal rules for K.

The $\mathbb{F}\Box$ rule extends a branch containing $\sigma\,\mathbb{F}\,\Box A$ by $\sigma.n\,\mathbb{F}\,A$. Similarly, the $\mathbb{T}\Diamond$ rule extends a branch containing $\sigma\,\mathbb{T}\,\Diamond A$ by $\sigma.n\,\mathbb{T}\,A$. 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 5.2.

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

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

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

5.3 Tableaux for K

Example 5.1. We give a closed tableau that shows $\vdash (\Box A \wedge \Box B) \rightarrow \Box(A \wedge B)$.

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

Example 5.2. We give a closed tableau that shows $\vdash \Diamond(A \vee B) \rightarrow (\Diamond A \vee \Diamond B)$:

1.	$1\,\mathbb{F}\ \ \Diamond(A \lor B) \to (\Diamond A \lor \Diamond B)$		Assumption
2.	$1\,\mathbb{T}\ \ \Diamond(A \lor B)$		$\to\mathbb{T}\,1$
3.	$1\,\mathbb{F}\ \ \Diamond A \lor \Diamond B$		$\to\mathbb{T}\,1$
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.	$1.1\,\mathbb{F}\ \ A$	$1.1\,\mathbb{F}\ \ B$	$\Diamond\mathbb{F}\,4$; $\Diamond\mathbb{F}\,5$
	$\otimes$	$\otimes$	

5.4 Soundness

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

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

has a closed tableau then $B_1, \dots, B_n \vDash A$. It is easier to prove the contrapositive: if for some M and world w, $M, w \Vdash B_i$ for all $i = 1, \dots, n$ 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 5.3. Let P be some set of prefixes, i.e., $P \subseteq (\mathbb{Z}^+)^* \setminus \{\Lambda\}$ and let M be a model. A function $f : P \to W$ 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 5.4. 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 5.5. *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 5.6 (Soundness). *If Γ has a closed tableau , Γ is unsatisfiable.*

Proof. We call a branch of a tableauxatisfiable iff the set of signed formulas on it is satisfiable, and let's call a tableauxatisfiable 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 $\mathbb{T}\, B \vee C \in \Gamma$: 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 $M, f(\sigma) \nVdash B \rightarrow C$. Then $M, f(\sigma) \Vdash B$ and $M, f(\sigma) \nVdash C$. This means that M, f satisfies both $\sigma\,\mathbb{T}\, B$ and $\sigma\,\mathbb{F}\, C$.

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

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

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

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

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

$M, f(\sigma) \nVdash C$. In the former case, M, f satisfies $\sigma\, \mathbb{F}\, B$, i.e., the left branch is satisfiable. In the latter, M, f satisfies $\sigma\, \mathbb{F}\, C$, i.e., the right branch is satisfiable.

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

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

□

Corollary 5.7. *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 5.6, Δ is unsatisfiable since it has a closed tableau , a contradiction. So we must have $\Gamma \vdash A$ after all. □

Corollary 5.8. *If $\vdash A$ then A is true in all models.*

$\dfrac{\sigma\ \mathbb{T}\ \Box A}{\sigma\ \mathbb{T}\ A}\ \mathrm{T}\Box$	$\dfrac{\sigma\ \mathbb{F}\ \Diamond A}{\sigma\ \mathbb{F}\ A}\ \mathrm{T}\Diamond$
$\dfrac{\sigma\ \mathbb{T}\ \Box A}{\sigma\ \mathbb{T}\ \Diamond A}\ \mathrm{D}\Box$	$\dfrac{\sigma\ \mathbb{F}\ \Diamond A}{\sigma\ \mathbb{F}\ \Box A}\ \mathrm{D}\Diamond$
$\dfrac{\sigma.n\ \mathbb{T}\ \Box A}{\sigma\ \mathbb{T}\ A}\ \mathrm{B}\Box$	$\dfrac{\sigma.n\ \mathbb{F}\ \Diamond A}{\sigma\ \mathbb{F}\ A}\ \mathrm{B}\Diamond$
$\dfrac{\sigma\ \mathbb{T}\ \Box A}{\sigma.n\ \mathbb{T}\ \Box A}\ 4\Box$ $\sigma.n$ 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 5.3: More modal rules.

5.5 Rules for Other Accessibility Relations

In order to deal with logics determined by special accessibility relations, we consider the additional rules in table 5.3.

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

Table 5.4: Tableau rules for various modal logics.

Adding these rules results in systems that are sound and complete for the logics given in table 5.4.

5.6 Tableaux for Other Logics

Example 5.9. 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.	$1.1\,\mathbb{T}\ \ A$	$\Box\mathbb{T}\,2$
	$\otimes$	

5.7 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 5.10. 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 $\mathfrak{M}, f \Vdash \Gamma$, in particular, $\mathfrak{M}, f(\sigma) \Vdash \Box B$. 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$\Diamond$ to $\sigma\,\mathbb{F}\,\Diamond B \in \Gamma$: Exercise. □

Proposition 5.11. D$\Box$ *and* D$\Diamond$ *are sound for serial models.*

Proof. 1. The branch is expanded by applying D$\Box$ to $\sigma\,\mathbb{T}\,\Box B \in \Gamma$: This results in a new signed formula $\sigma\,\mathbb{T}\,\Diamond B$ on the branch. Suppose $\mathfrak{M}, f \Vdash \Gamma$, 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$\Diamond$ to $\sigma\,\mathbb{F}\,\Diamond B \in \Gamma$: Exercise. □

Proposition 5.12. B$\Box$ *and* B$\Diamond$ *are sound for symmetric models.*

Proof.

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

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

Proposition 5.13. 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 $M, f \Vdash \Gamma$, in particular, $M, f(\sigma) \Vdash \Box B$. Since f is an interpretation of prefixes on the branch into 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 $M, f(\sigma) \Vdash \Box B$, $M, w \Vdash B$. Hence, $M, f(\sigma.n) \Vdash \Box B$, and 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 5.14. 4r$\Box$ *and* 4r$\Diamond$ *are sound for euclidean models.*

Proof.

1. The branch is expanded by applying 4r$\Box$ to $\sigma.n\,\mathbb{T}\,\Box B \in \Gamma$: This results in a new signed formula $\sigma\,\mathbb{T}\,\Box B$ on the branch. Suppose $M, f \Vdash \Gamma$, in particular, $M, f(\sigma.n) \Vdash \Box B$. Since f is an interpretation of prefixes on the branch into 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 $M, f(\sigma).n \Vdash \Box B$, $M, w \Vdash B$. Hence, $M, f(\sigma) \Vdash \Box B$, and 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 5.15. *The tableauxystems given in table 5.4 are sound for the respective classes of models.*

5.8 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 5.5.

Example 5.16. 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 5.5: Simplified rules for **S5**.

5.9 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 5.17. 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 5.18. *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 5.19 (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 $\sigma\, \mathbb{T}\, p \in \Delta$ then $\sigma \in V(p)$ (by definition of V) and so $\mathbf{M}(\Delta), \sigma \Vdash p$.

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

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

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

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

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

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

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

6. $A \equiv \Box B$: If $\sigma\,\mathbb{T}\,\Box B \in \Delta$, 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, $M(\Delta), \sigma' \Vdash B$ for every σ' such that $R\sigma\sigma'$. Therefore, $M(\Delta), \sigma \Vdash \Box B$.

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

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

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

Corollary 5.20. *If $\Gamma \vDash A$ then $\Gamma \vdash A$.*

Corollary 5.21. *If A is true in all models, then $\vdash A$.*

5.10 Countermodels from Tableaux

The proof of the completeness theorem doesn't just show that if $\vDash A$ then $\vdash A$, it also gives us a method for constructing countermodels to A if $\nvDash A$. In the case of **K**, this method constitutes a *decision procedure.* For suppose $\nvDash A$. Then the proof of Proposition 5.18 gives a method for constructing a complete tableau . The method in fact always terminates. The propositional rules for **K** only add prefixed formulas of lower complexity, i.e., each propositional rule need only be applied once on a branch for any signed formula $\sigma\, S\, A$. New prefixes are only generated by the $\Box\mathbb{F}$ and $\Diamond\mathbb{T}$ rules, and also only have to be applied once (and produce a single new prefix). $\Box\mathbb{T}$ and $\Diamond\mathbb{F}$ have to be applied potentially multiple times, but only once per prefix, and only finitely many new prefixes are generated. So the construction either results in a closed branch or a complete branch after finitely many stages.

Once a tableau with an open complete branch is constructed, the proof of Theorem 5.19 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 5.22. We know that $\nvdash \Box(p \lor q) \to (\Box p \lor \Box q)$. The construction of a tableau begins with:

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

The tableau is of course not finished yet. In the next step, we consider the only line without a checkmark: the prefixed formula $1\,\mathbb{T}\,\Box(p \lor q)$ 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.	$1\,\mathbb{F}\ \ \Box(p \lor q) \to (\Box p \lor \Box q)$ ✓	Assumption
2.	$1\,\mathbb{T}\ \ \Box(p \lor q)$	$\to\mathbb{F}\,1$
3.	$1\,\mathbb{F}\ \ \Box p \lor \Box q$ ✓	$\to\mathbb{F}\,1$
4.	$1\,\mathbb{F}\ \ \Box p$ ✓	$\lor\mathbb{F}\,3$
5.	$1\,\mathbb{F}\ \ \Box q$ ✓	$\lor\mathbb{F}\,3$
6.	$1.1\,\mathbb{F}\ \ p$ ✓	$\Box\mathbb{F}\,4$
7.	$1.2\,\mathbb{F}\ \ q$ ✓	$\Box\mathbb{F}\,5$
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.		$1\,\mathbb{F}\ \ \Box(p \lor q) \to (\Box p \lor \Box q)$ ✓		Assumption
2.		$1\,\mathbb{T}\ \ \Box(p \lor q)$		$\to\mathbb{F}\,1$
3.		$1\,\mathbb{F}\ \ \Box p \lor \Box q$ ✓		$\to\mathbb{F}\,1$
4.		$1\,\mathbb{F}\ \ \Box p$ ✓		$\lor\mathbb{F}\,3$
5.		$1\,\mathbb{F}\ \ \Box q$ ✓		$\lor\mathbb{F}\,3$
6.		$1.1\,\mathbb{F}\ \ p$ ✓		$\Box\mathbb{F}\,4$
7.		$1.2\,\mathbb{F}\ \ q$ ✓		$\Box\mathbb{F}\,5$
8.		$1.1\,\mathbb{T}\ \ p \lor q$ ✓		$\Box\mathbb{T}\,2$
9.		$1.2\,\mathbb{T}\ \ p \lor q$ ✓		$\Box\mathbb{T}\,2$
10.	$1.1\,\mathbb{T}\ \ p$ ✓ $\otimes$	$1.1\,\mathbb{T}\ \ q$ ✓		$\lor\mathbb{T}\,8$
11.		$1.2\,\mathbb{T}\ \ p$ ✓	$1.2\,\mathbb{T}\ \ q$ ✓ $\otimes$	$\lor\mathbb{T}\,9$

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-

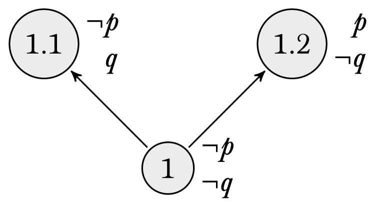


Figure 5.1: A countermodel to $\Box(p \lor q) \to (\Box p \lor \Box q)$.

tains $1.1\, \mathbb{T}\, q$). The model is pictured in Figure 5.1, and you can verify that it is a countermodel to $\Box(p \lor q) \to (\Box p \lor \Box q)$.

Problems

Problem 5.1. 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. $\Diamond p \to \Diamond(p \lor q)$

Problem 5.2. Complete the proof of Theorem 5.6.

Problem 5.3. Give closed tableaux that show the following:

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

Problem 5.4. Complete the proof of Proposition 5.10

Problem 5.5. Complete the proof of Proposition 5.11

Problem 5.6. Complete the proof of Proposition 5.12

Problem 5.7. Complete the proof of Proposition 5.13

Problem 5.8. Complete the proof of Proposition 5.14

Problem 5.9. Complete the proof of Theorem 5.19.

PART III

Conditionals

CHAPTER 6

Introduction

6.1 The Material Conditional

In its simplest form in English, a conditional is a sentence of the form "If …then …," where the … are themselves sentences, such as "If the butler did it, then the gardener is innocent." In introductory logic courses, we 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:

$$\text{If } \Gamma, A \vDash B \text{ then } \Gamma \vDash A \rightarrow B \tag{6.1}$$

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

$$A \rightarrow B \vDash \neg A \lor B \tag{6.2}$$

$$\neg A \lor B \vDash A \rightarrow B \tag{6.3}$$

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

$$B \vDash A \to B \tag{6.4}$$

$$\neg A \vDash A \to B \tag{6.5}$$

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

$$\neg(A \to B) \vDash A \land \neg B \tag{6.6}$$

$$A \land \neg B \vDash \neg(A \to B) \tag{6.7}$$

The material conditional supports modus ponens:

$$A, A \to B \vDash B \tag{6.8}$$

The material conditional agglomerates:

$$A \to B, A \to C \vDash A \to (B \land C) \tag{6.9}$$

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

$$A \to B \vDash (A \land C) \to B \tag{6.10}$$

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

$$A \to B, B \to C \vDash A \to C \tag{6.11}$$

The material conditional is equivalent to its contrapositive:

$$A \to B \vDash \neg B \to \neg A \tag{6.12}$$

$$\neg B \to \neg A \vDash A \to B \tag{6.13}$$

These are all useful and unproblematic inferences in mathematical reasoning. However, the philosophical and linguistic literature is replete with purported counterexamples to the equivalent inferences in non-mathematical contexts. These suggest that the material conditional $\rightarrow$ is not—or at least not always—the appropriate connective to use when symbolizing English "if ... then ..." statements.

6.2 Paradoxes of the Material Conditional

One of the first to criticize the use of $A \rightarrow B$ as a way to symbolize "if ... then ..." statements of English was C. I. Lewis. Lewis was criticizing the use of the material condition in Whitehead and Russell's *Principia Mathematica*, who pronounced $\rightarrow$ as "implies." Lewis rightly complained that if $\rightarrow$ meant "implies," then any false proposition p implies that p implies q, since $p \rightarrow (p \rightarrow q)$ is true if p is false, and that any true proposition q implies that p implies q, since $q \rightarrow (p \rightarrow q)$ is true if q is true.

Logicians of course know that *implication*, i.e., logical entailment, is not a connective but a relation between formulas or statements. So we should just not read $\rightarrow$ as "implies" to avoid confusion.[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 $(A \rightarrow B) \lor (B \rightarrow A)$. This would suggest that if you take two indicative sentences S and T from the newspaper at random, the sentence "If S then T, or if T then S" should be true.

6.3 The Strict Conditional

Lewis introduced the *strict conditional* $\strictif$ 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, $(A \strictif B) \lor (B \strictif A)$ is not valid. The strict conditional $A \strictif B$ is also not equivalent to $\lnot A \lor B$, so it is not truth functional.

We have:

$$A \strictif B \vDash \lnot A \lor B \text{ but:} \tag{6.14}$$

$$\lnot A \lor B \nvDash A \strictif B \tag{6.15}$$

$$B \nvDash A \strictif B \tag{6.16}$$

$$\lnot A \nvDash A \strictif B \tag{6.17}$$

$$\lnot(A \rightarrow B) \nvDash A \land \lnot B \text{ but:} \tag{6.18}$$

$$A \land \lnot B \vDash \lnot(A \strictif B) \tag{6.19}$$

However, the strict conditional still supports modus ponens:

$$A, A \multimap B \vDash B \tag{6.20}$$

The strict conditional agglomerates:

$$A \multimap B, A \multimap C \vDash A \multimap (B \wedge C) \tag{6.21}$$

Antecedent strengthening holds for the strict conditional:

$$A \multimap B \vDash (A \wedge C) \multimap B \tag{6.22}$$

The strict conditional is also transitive:

$$A \multimap B, B \multimap C \vDash A \multimap C \tag{6.23}$$

Finally, the strict conditional is equivalent to its contrapositive:

$$A \multimap B \vDash \neg B \multimap \neg A \tag{6.24}$$

$$\neg B \multimap \neg A \vDash A \multimap B \tag{6.25}$$

However, the strict conditional still has its own "paradoxes." Just as a material conditional with a false antecedent or a true consequent is true, a strict conditional with a *necessarily* false antecedent or a necessarily true consequent is true. Moreover, any true strict conditional is necessarily true, and any false strict conditional is necessarily false. In other words, we have

$$\Box A \vDash A \multimap B \tag{6.26}$$

$$\Box \neg B \vDash A \multimap B \tag{6.27}$$

$$A \multimap B \vDash \Box(A \multimap B) \tag{6.28}$$

$$\neg(A \multimap B) \vDash \Box\neg(A \multimap B) \tag{6.29}$$

These are not problems if you think of $\multimap$ 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 $\multimap$ as a logical connective that is supposed to capture "if ... then ...," especially the last two. For surely there are "if ... then ..." statements that are contingently true or contingently false—in fact, they generally are neither necessary nor impossible.

6.4 Counterfactuals

A very common and important form of "if ... then ..." constructions in English are built using the past subjunctive form of *to be*: "if it were the case that ... then it would be the case that ..." Because usually the antecedent of such a conditional is false, i.e., counter to fact, they are called *counterfactual conditionals* (and because they use the subjunctive form of *to be*, also *subjunctive conditionals*. They are distinguished from *indicative* conditionals which take the form of "if it is the case that ... then it is the case that ..." Counterfactual and indicative conditionals differ in truth conditions. Consider Adams's famous example:

> If Oswald didn't kill Kennedy, someone else did.
>
> If Oswald hadn't killed Kennedy, someone else would have.

The first is indicative, the second counterfactual. The first is clearly true: we know JFK was killed by someone, and if that someone wasn't (contrary to the Warren Report) Lee Harvey Oswald, then someone else killed JFK. The second one says something different. It claims that if Oswald hadn't killed Kennedy, i.e., if the Dallas shooting had been avoided or had been unsuccessful, history would have subsequently unfolded in such a way that another assassination would have been successful. In order for it to be true, it would have to be the case that powerful forces had conspired to ensure JFK's death (as many JFK conspiracy theorists believe).

It is a live debate whether the *indicative* conditional is correctly captured by the material conditional, in particular, whether the paradoxes of the material conditional can be "explained" in a way that is compatible with it giving the truth conditions for English indicative conditionals. By contrast, it is uncontroversial that counterfactual conditionals cannot be symbolized correctly by the material conditionals. That is clear because, even though generally the antecedents of counterfactuals are false, not all counterfactuals with false antecedents are true—for instance, if you believe the Warren Report, and there was no conspiracy to assassinate JFK, then Adams's counterfactual conditional is an example.

Counterfactual conditionals play an important role in causal reasoning: a prime example of the use of counterfactuals is to express causal relationships. E.g., striking a match causes it to light, and you can express this by saying "if this match were struck, it would light." Material, and generally indicative conditionals, cannot be used to express this: "the match is struck $\rightarrow$ the match lights" is true if the match is never struck, regardless of what would happen if it were. Even worse, "the match is struck $\rightarrow$ the match turns into a bouquet of flowers" is also true if it is never struck, but the match would certainly not turn into a bouquet of flowers if it were struck.

It is still debated What exactly the correct logic of counterfactuals is. An influential analysis of counterfactuals was given by Stalnaker and Lewis. According to them, a counterfactual "if it were the case that S then it would be the case that T" is true iff T is true in the counterfactual situation ("possible world") that is closest to the way the actual world is and where S is true. This is called an "ontic" analysis, since it makes reference to an ontology of possible worlds. Other analyses make use of conditional probabilities or theories of belief revision. There is a proliferation of different proposed logics of counterfactuals. There isn't even a single Lewis-Stalnaker logic of counterfactuals: even though Stalnaker and Lewis proposed accounts along similar lines with reference to closest possible worlds, the assumptions they made result in different valid inferences.

Problems

Problem 6.1. 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. $\neg\Box(p \to q) \nvDash p \wedge \neg q$

4. $\nvDash \Box(p \to q) \vee \Box(q \to p)$

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

1. $\Box(A \to B) \vDash \neg A \vee B$

2. $A \wedge \neg B \vDash \neg\Box(A \to B)$

3. $A, \Box(A \to B) \vDash B$

4. $\Box(A \to B), \Box(A \to C) \vDash \Box(A \to (B \wedge C))$

5. $\Box(A \to B) \vDash \Box((A \wedge C) \to B)$

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

1. $\Box\neg B \vDash A \prec B$

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 7

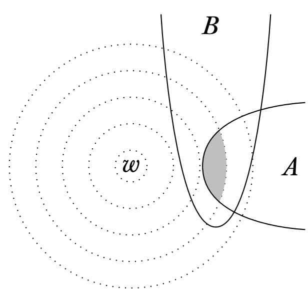
Minimal Change Semantics

7.1 Introduction

Stalnaker and Lewis proposed accounts of counterfactual conditionals such as "If the match were struck, it would light." Their accounts were proposals for how to properly understand the truth conditions for such sentences. The idea behind both proposals is this: to evaluate whether a counterfactual conditional is true, we have to consider those possible worlds which are minimally different from the way the world actually is to make the antecedent true. If the consequent is true in these possible worlds, then the counterfactual is true. For instance, suppose I hold a match and a matchbook in my hand. In the actual world I only look at them and ponder what would happen if I were to strike the match. The minimal change from the actual world where I strike the match is that where I decide to act and strike the match. It is minimal in that nothing else changes: I don't also jump in the air, striking the match doesn't also light my hair on fire, I don't suddenly lose all strength in my fingers, I am not simultaneously doused with water in a SuperSoaker ambush, etc. In that alternative possibility, the match lights. Hence, it's true that if I were to strike the match, it would light.

This intuitive account can be paired with formal semantics for logics of counterfactuals. Lewis introduced the symbol "$\mathrel{\Box\!\!\to}$" for the counterfactual while Stalnaker used the symbol "$>$". We'll use $\mathrel{\Box\!\!\to}$, and add it as a binary connective to propositional logic. So, we have, in addition to formulas of the form $A \to B$ also formulas of the form $A \mathrel{\Box\!\!\to} B$. 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 $M, w \Vdash A \mathrel{\Box\!\!\to} B$ 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.

7.2 Sphere Models

One way of providing a formal semantics for counterfactuals is to turn Lewis's informal account into a mathematical structure. The spheres around a world w then are sets of worlds. Since the spheres are nested, the sets of worlds around w have to be linearly ordered by the subset relation.

Definition 7.1. A *sphere model* is a triple $M = \langle W, O, V \rangle$ where W is a non-empty set of worlds, $V : \mathrm{At}_0 \to \wp(W)$ is a valuation, and $O : W \to \wp(\wp(W))$ 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 7.1 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 next larger sphere is $S_2 = \{w, w_1, w_2, w_3\}$. The worlds further out

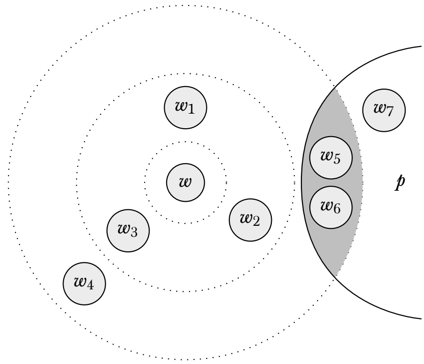


Figure 7.1: Diagram of a sphere model

are w_4, w_5, w_6, so the outermost sphere is $S_3 = \{w, w_1, \dots, w_6\}$. The system of spheres around w is $O_w = \{S_1, S_2, S_3\}$. The world w_7 is not in any sphere around w. The closest worlds in which p is true are w_5 and w_6, and so the smallest p-admitting sphere is S_3.

To define satisfaction of a formula A at world w in a sphere model M, $M, w \Vdash A$, we expand the definition for modal formulas to include a clause for $B \mathrel{\Box\!\!\rightarrow} C$:

Definition 7.2. $M, w \Vdash B \mathrel{\Box\!\!\rightarrow} C$ iff either

1. For all $u \in \bigcup O_w$, $M, u \nVdash C$, or
2. For some $S \in O_w$,
   a) $M, u \Vdash B$ for some $u \in S$, and
   b) for all $v \in S$, either $M, v \nVdash B$ or $M, v \Vdash C$.

According to this definition, $M, w \Vdash B \mathrel{\Box\!\!\rightarrow} C$ 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 that we didn't require in the definition that S is the *innermost* B-admitting sphere, contrary to what one might expect from the

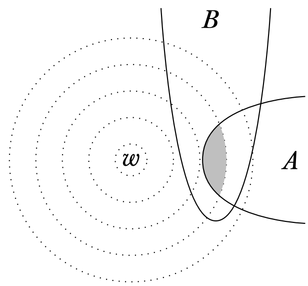


Figure 7.2: Non-vacuously true counterfactual

intuitive explanation. But if the condition in (2) is satisfied for some sphere S, then it is also satisfied for all spheres S contains, and hence in particular for the innermost sphere.

Note also that the definition of sphere models does not require that there *is* an innermost B-admitting sphere: we may have an infinite sequence $S_1 \supsetneq S_2 \supsetneq \cdots \supsetneq \{w\}$ of B-admitting spheres, and hence no innermost B-admitting spheres. In that case, $M, w \Vdash B \boxright C$ iff $B \rightarrow C$ holds throughout the spheres S_i, S_{i+1}, …, for some i.

7.3 Truth and Falsity of Counterfactuals

A counterfactual $A \boxright B$ is (non-vacuously) true if the closest A-worlds are all B-worlds, as depicted in Figure 7.2. A counterfactual is also true at w if the system of spheres around w has no A-admitting spheres at all. In that case it is *vacuously* true (see Figure 7.3).

It can be false in two ways. One way is if the closest A-worlds are not all B-worlds, but some of them are. In this case, $A \boxright \neg B$ is also false (see Figure 7.4). If the closest A-worlds do not overlap with the B-worlds at all, then $A \boxright B$. But, in this case all the closest A-worlds are $\neg B$-worlds, and so $A \boxright \neg B$ is true (see Figure 7.5).

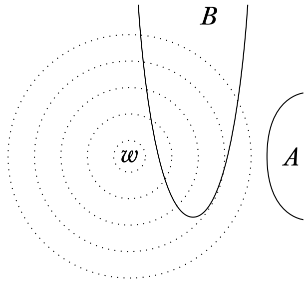


Figure 7.3: Vacuously true counterfactual

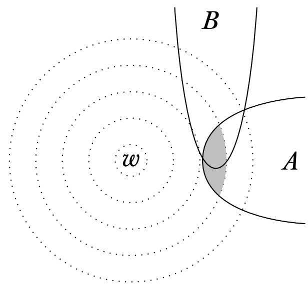


Figure 7.4: False counterfactual, false opposite

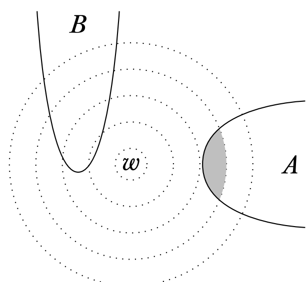


Figure 7.5: False counterfactual, true opposite

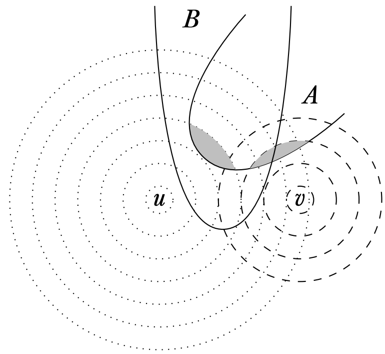


Figure 7.6: Contingent counterfactual

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

7.4 Antecedent Strengthenng

"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 for counterfactuals. Suppose it is true that if I were to strike this match, it would light. (That means, there is nothing wrong with the match or the matchbook surface, I will not break the match, etc.) But it is not true that if I were to light this match in outer space, it would light. So the following inference is invalid:

> I the match were struck, it would light.
>
> Therefore, if the match were struck in outer space, it would light.

The Lewis-Stalnaker account of conditionals explains this: the closest world where I light the match and I do so in outer

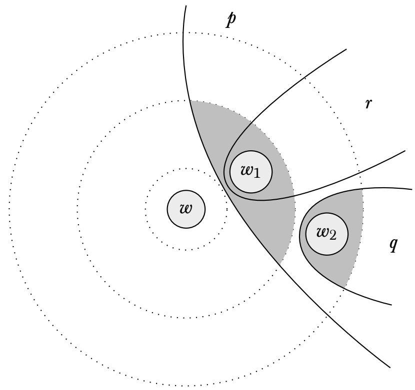


Figure 7.7: Counterexample to antecedent strengthening

space is much further removed from the actual world than the closest world where I light the match is. So although it's true that the match lights in the latter, it is not in the former. And that is as it schould be.

Example 7.3. The sphere semantics invalidates the inference, i.e., we have $p \mathrel{\Box\!\!\rightarrow} r \nvDash (p \wedge q) \mathrel{\Box\!\!\rightarrow} r$. Consider the model $\boldsymbol{M} = \langle W, O, V \rangle$ 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 $\boldsymbol{M}, w \Vdash p \mathrel{\Box\!\!\rightarrow} r$. There is also a $(p \wedge q)$-admitting sphere $S' = \{w, w_1, w_2\}$ but $\boldsymbol{M}, w_2 \nVdash (p \wedge q) \rightarrow r$, so $\boldsymbol{M}, w \nVdash (p \wedge q) \mathrel{\Box\!\!\rightarrow} r$ (see Figure 7.7).

7.5 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 7.4. The sphere semantics invalidates the inference, i.e., we have $p \mathrel{\Box\!\!\to} q, q \mathrel{\Box\!\!\to} r \nvDash p \mathrel{\Box\!\!\to} r$. Consider the model $\mathfrak{M} = \langle W, O, V\rangle$ where $W = \{w, w_1, w_2\}$, $O_w = \{\{w\}, \{w, w_1\}, \{w, w_1, w_2\}\}$, $V(p) = \{w_2\}$, $V(q) = \{w_1, w_2\}$, and $V(r) = \{w_1\}$. There is a p-admitting sphere $S = \{w, w_1, w_2\}$ and $q \to q$ is true at all worlds in it, so $\mathfrak{M}, w \Vdash p \mathrel{\Box\!\!\to} q$. There is also a q-admitting sphere

[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.

$S' = \{w, w_1\}$ and $M \nVdash q \to r$ is true at all worlds in it, so $M, w \Vdash q \,\Box\!\!\to r$. However, the p-admitting sphere $\{w, w_1, w_2\}$ contains a world, namely w_2, where $M, w_2 \nVdash p \to r$.

7.6 Contraposition

Material and strict conditionals are equivalent to their contrapositives. Counterfactuals are not. Here is an example due to Kratzer:

> If Goethe hadn't died in 1832, he would (still) be dead now.
>
> If Goethe weren't dead now, he would have died in 1832.

The first sentence is true: humans don't live hundreds of years. The second is clearly false: if Goethe weren't dead now, he would be still alive, and so couldn't have died in 1832.

Example 7.5. The sphere semantics invalidates contraposition, i.e., we have $p \,\Box\!\!\to q \nvDash \neg q \,\Box\!\!\to \neg p$. Think of p as "Goethe didn't die in 1832" and q as "Goethe is dead now." We can capture this in a model $M_1 = \langle W, O, V \rangle$ with $W = \{w, w_1, w_2\}$, $O = \{\{w\}, \{w, w_1\}, \{w, w_1, w_2\}\}$, $V(p) = \{w_1, w_2\}$ and $V(q) = \{w, w_1\}$. So w is the actual world where Goethe died in 1832 and is still dead; w_1 is the (close) world where Goethe died in, say, 1833, and is still dead; and w_2 is a (remote) world where Goethe is still alive. There is a p-admitting sphere $S = \{w, w_1\}$ and $p \to q$ is true at all worlds in it, so $M, w \Vdash p \,\Box\!\!\to q$. However, the $\neg q$-admitting sphere $\{w, w_1, w_2\}$ contains a world, namely w_2, where q is false and p is true, so $M, w_2 \nVdash \neg q \to \neg p$.

Problems

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

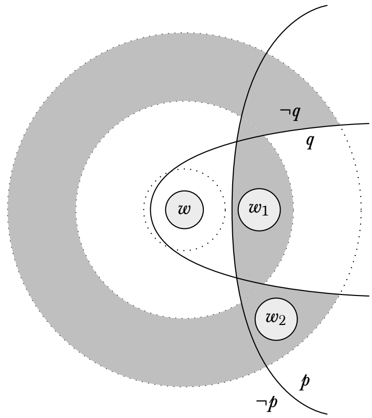


Figure 7.8: Counterexample to contraposition

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

Problem 7.3. In Example 7.4, world w_2 is where Hoover is born in Russia, is a communist, and not a traitor, and w_1 is the world where Hoover is born in the US, is a communist, and a traitor. In this model, w_1 is closer to w than w_2 is. Is this necessary? Can you give a counterexample that does not assume that Hoover's being born in Russia is a more remote possibility than him being a Communist?

PART IV

Appendices

APPENDIX A

Sets

A.1 Basics

Sets are the most fundamental building blocks of mathematical objects. In fact, almost every mathematical object can be seen as a set of some kind. In logic, as in other parts of mathematics, sets and set-theoretical talk is ubiquitous. So it will be important to discuss what sets are, and introduce the notations necessary to talk about sets and operations on sets in a standard way.

Definition A.1 (Set). A *set* is a collection of objects, considered independently of the way it is specified, of the order of the objects in the set, or of their multiplicity. The objects making up the set are called *elements* or *members* of the set. If a is an element of a set X, we write $a \in X$ (otherwise, $a \notin X$). The set which has no elements is called the *empty* set and denoted by the symbol $\emptyset$.

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 $S = \{\text{Ruth}\}$. In general, when we have some objects a_1, …, a_n, then the set consisting of exactly those objects is written $\{a_1, \dots, a_n\}$. Frequently we'll specify a set by some property that its elements share—as we just did, for instance, by specifying S as the set of Richard's siblings. We'll use the following shorthand notation for that: $\{x : \ldots x \ldots\}$, where the $\ldots x \ldots$ stands for the property that x has to have in order to be counted among the elements of the set. In our example, we could have specified S also as

$$S = \{x : x \text{ is a sibling of Richard}\}.$$

When we say that sets are independent of the way they are specified, we mean that the elements of a set are all that matters. For instance, it so happens that

$$\{\text{Nicole}, \text{Jacob}\},$$
$$\{x : \text{is a niece or nephew of Richard}\}, \text{ and}$$
$$\{x : \text{is a child of Ruth}\}$$

are three ways of specifying one and the same set.

Saying that sets are considered independently of the order of their elements and their multiplicity is a fancy way of saying that

$$\{\text{Nicole}, \text{Jacob}\} \text{ and}$$
$$\{\text{Jacob}, \text{Nicole}\}$$

are two ways of specifying the same set; and that

$$\{\text{Nicole}, \text{Jacob}\} \text{ and}$$
$$\{\text{Jacob}, \text{Nicole}, \text{Nicole}\}$$

are also two ways of specifying the same set. In other words, all that matters is which elements a set has. The elements of a set are not ordered and each element occurs only once. When we *specify* or *describe* a set, elements may occur multiple times and in different orders, but any descriptions that only differ in the order of elements or in how many times elements are listed describes the same set.

Definition A.3 (Extensionality). If X and Y are sets, then X and Y are *identical*, $X = Y$, iff every element of X is also an element of Y, and vice versa.

Extensionality gives us a way for showing that sets are identical: to show that $X = Y$, show that whenever $x \in X$ then also $x \in Y$, and whenever $y \in Y$ then also $y \in X$.

A.2 Some Important Sets

Example A.4. Mostly we'll be dealing with sets that have mathematical objects as members. You will remember the various sets of numbers: $\mathbb{N}$ is the set of *natural* numbers $\{0, 1, 2, 3, \dots\}$; $\mathbb{Z}$ the set of *integers*,

$$\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\};$$

$\mathbb{Q}$ the set of *rational* numbers ($\mathbb{Q} = \{z/n : z \in \mathbb{Z}, n \in \mathbb{N}, n \neq 0\}$); and $\mathbb{R}$ the set of *real* numbers. These are all *infinite* sets, that is, they each have infinitely many elements. As it turns out, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ have the same number of elements, while $\mathbb{R}$ has a whole bunch more—$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ are "countable and infinite" whereas $\mathbb{R}$ is "uncountable".

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.5 (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,

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

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 $\mathrm{len}(x) = n$.

Example A.6 (Infinite sequences). For any set A we may also consider the set A^ω of infinite sequences of elements of A. An infinite sequence $a_1 a_2 a_3 a_4 \ldots$ consists of a one-way infinite list of objects, each one of which is an element of A.

A.3 Subsets

Sets are made up of their elements, and every element of a set is a part of that set. But there is also a sense that some of the elements of a set *taken together* are a "part of" that set. For instance, the number 2 is part of the set of integers, but the set of even numbers is also a part of the set of integers. It's important to keep those two senses of being part of a set separate.

Definition A.7 (Subset). If every element of a set X is also an element of Y, then we say that X is a *subset* of Y, and write $X \subseteq Y$.

Example A.8. First of all, 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 $\{a, b, c\}$.

Note that a set may contain other sets, not just as subsets but as elements! In particular, a set may happen to *both* be an element and a subset of another, e.g., $\{0\} \in \{0, \{0\}\}$ and also $\{0\} \subseteq \{0, \{0\}\}$.

Extensionality gives a criterion of identity for sets: $X = Y$ iff every element of X is also an element of Y and vice versa. The definition of "subset" defines $X \subseteq Y$ precisely as the first half of this criterion: every element of X is also an element of Y. Of course the definition also applies if we switch X and Y: $Y \subseteq X$ iff every element of Y is also an element of X. And that, in turn, is exactly the "vice versa" part of extensionality. In other words, extensionality amounts to: $X = Y$ iff $X \subseteq Y$ and $Y \subseteq X$.

Definition A.9 (Power Set). The set consisting of all subsets of a set X is called the *power set of* X, written $\wp(X)$.

$$\wp(X) = \{Y : Y \subseteq X\}$$

Example A.10. 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.4 Unions and Intersections

We can define new sets by abstraction, and the property used to define the new set can mention sets we've already defined. So for instance, if X and Y are sets, the set $\{x : x \in X \vee x \in Y\}$ defines a set which consists of all those objects which are elements of either X or Y, i.e., it's the set that combines the elements of X and Y. This operation on sets—combining them—is very useful and common, and so we give it a name and a symbol.

Definition A.11 (Union). The *union* of two sets X and Y, written $X \cup Y$, is the set of all things which are elements of X, Y, or both.

$$X \cup Y = \{x : x \in X \vee x \in Y\}$$

Example A.12. Since the multiplicity of elements doesn't matter, the union of two sets which have an element in common contains that element only once, e.g., $\{a, b, c\} \cup \{a, 0, 1\} = \{a, b, c, 0, 1\}$.

The union of a set and one of its subsets is just the bigger set: $\{a, b, c\} \cup \{a\} = \{a, b, c\}$.

The union of a set with the empty set is identical to the set: $\{a, b, c\} \cup \emptyset = \{a, b, c\}$.

The operation that forms the set of all elements that X and Y have in common is called their *intersection*.

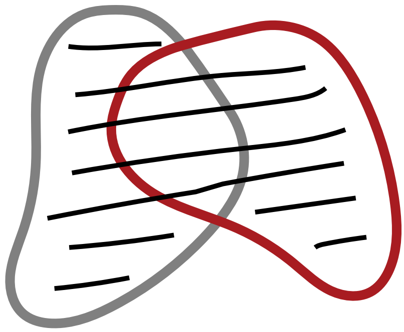


Figure A.1: The union $X \cup Y$ of two sets is set of elements of X together with those of Y.

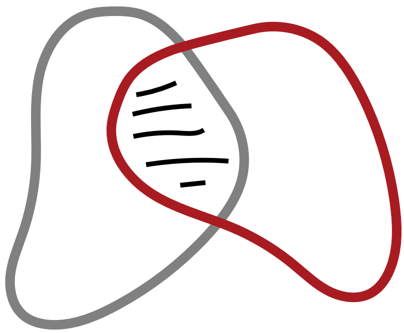


Figure A.2: The intersection $X \cap Y$ of two sets is the set of elements they have in common.

Definition A.13 (Intersection). The *intersection* of two sets X and Y, written $X \cap Y$, is the set of all things which are elements of both X and Y.

$$X \cap Y = \{x : x \in X \land x \in Y\}$$

Two sets are called *disjoint* if their intersection is empty. This means they have no elements in common.

Example A.14. 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.15. If Z is a set of sets, then $\bigcup Z$ is the set of elements of elements of Z:

$$\bigcup Z = \{x : x \text{ belongs to an element of } Z\}, \text{ i.e.,}$$
$$\bigcup Z = \{x : \text{there is a } Y \in Z \text{ so that } x \in Y\}$$

Definition A.16. If Z is a set of sets, then $\bigcap Z$ is the set of objects which all elements of Z have in common:

$$\bigcap Z = \{x : x \text{ belongs to every element of } Z\}, \text{ i.e.,}$$
$$\bigcap Z = \{x : \text{for all } Y \in Z, x \in Y\}$$

Example A.17. Suppose $Z = \{\{a, b\}, \{a, d, e\}, \{a, d\}\}$. Then $\bigcup Z = \{a, b, d, e\}$ and $\bigcap Z = \{a\}$.

We could also do the same for a sequence of sets X_1, X_2, ...

$$\bigcup_i X_i = \{x : x \text{ belongs to one of the } X_i\}$$

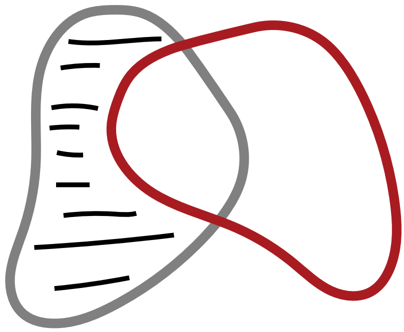


Figure A.3: The difference $X \setminus Y$ of two sets is the set of those elements of X which are not also elements of Y.

$$\bigcap_i X_i = \{x : x \text{ belongs to every } X_i\}.$$

Definition A.18 (Difference). The *difference* $X \setminus Y$ is the set of all elements of X which are not also elements of Y, i.e.,

$$X \setminus Y = \{x : x \in X \text{ and } x \notin Y\}.$$

A.5 Pairs, Tuples, Cartesian Products

Sets have no order to their elements. We just think of them as an unordered collection. 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 $\langle x, y\rangle \neq \langle y, x\rangle$.

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. In fact, we can think of triples as special ordered pairs, where the first element is itself an ordered pair: $\langle x, y, z\rangle$ is short for $\langle\langle x, y\rangle, z\rangle$. The same is true for quadruples: $\langle x, y, z, u\rangle$ is short for $\langle\langle\langle x, y\rangle, z\rangle, u\rangle$, and so on. In general, we talk of *ordered n-tuples* $\langle x_1, \dots, x_n\rangle$.

Definition A.19 (Cartesian product). Given sets X and Y, their *Cartesian product* $X \times Y$ is $\{\langle x, y\rangle : x \in X \text{ and } y \in Y\}$.

Example A.20. If $X = \{0, 1\}$, and $Y = \{1, a, b\}$, then their product is

$$X \times Y = \{\langle 0, 1\rangle, \langle 0, a\rangle, \langle 0, b\rangle, \langle 1, 1\rangle, \langle 1, a\rangle, \langle 1, b\rangle\}.$$

Example A.21. If X is a set, the product of X with itself, $X \times X$, is also written X^2. It is the set of *all* pairs $\langle x, y\rangle$ with $x, y \in X$. The set of all triples $\langle x, y, z\rangle$ is X^3, and so on. We can give an inductive definition:

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

Proposition A.22. *If X has n elements and Y has m elements, then $X \times Y$ has $n \cdot m$ elements.*

Proof. For every element x in X, there are m elements of the form $\langle x, y\rangle \in X \times Y$. Let $Y_x = \{\langle x, y\rangle : y \in Y\}$. Since whenever $x_1 \neq x_2$, $\langle x_1, y\rangle \neq \langle x_2, y\rangle$, $Y_{x_1} \cap Y_{x_2} = \emptyset$. But if $X = \{x_1, \dots, x_n\}$, then $X \times Y = Y_{x_1} \cup \dots \cup Y_{x_n}$, and so has $n \cdot m$ elements.

To visualize this, arrange the elements of $X \times Y$ in a grid:

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

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.23. If X is a set, a *word* over X is any sequence of elements of X. A sequence can be thought of as an n-tuple of elements of X. For instance, if $X = \{a, b, c\}$, 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, of course. By convention, we count elements of X as sequences of length 1, and $\emptyset$ as the sequence of length 0. The set of *all* words over X then is

$$X^* = \{\emptyset\} \cup X \cup X^2 \cup X^3 \cup \dots$$

A.6 Russell's Paradox

We said that one can define sets by specifying a property that its elements share, e.g., defining the set of Richard's siblings as

$$S = \{x : x \text{ is a sibling of Richard}\}.$$

In the very general context of mathematics one must be careful, however: not every property lends itself to *comprehension*. Some properties do not define sets. If they did, we would run into outright contradictions. One example of such a case is Russell's Paradox.

Sets may be elements of other sets—for instance, the power set of a set X is made up of sets. And so it makes sense, of course, 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, surely *all* sets form a collection of objects, so we should be able to collect them 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 . The set of all sets does not have this property, but all sets we have encountered so far have it. $\mathbb{N}$ is not an element of $\mathbb{N}$, since it is a set, not a natural number. $\wp(X)$ is generally not an element of $\wp(X)$; e.g., $\wp(\mathbb{R}) \notin \wp(\mathbb{R})$ since it is a set of sets of real numbers, not a set of real numbers. 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?

If R exists, it makes sense to ask if $R \in R$ or not—it must be either $\in R$ or $\notin R$. Suppose the former is true, i.e., $R \in R$. R was defined as the set of all sets that are not elements of themselves, and so if $R \in R$, then R does not have this defining property of R. But only sets that have this property are in R, hence, R cannot be an element of R, i.e., $R \notin R$. But R can't both be and not be an element of R, so we have a contradiction.

Since the assumption that $R \in R$ leads to a contradiction, we have $R \notin R$. But this also leads to a contradiction! For if $R \notin R$, it does have the defining property of R, and so would be an element of R just like all the other non-self-containing sets. And again, it can't both not be and be an element of R.

Problems

Problem A.1. Show that there is only one empty set, i.e., show that if X and Y are sets without members, then $X = Y$.

Problem A.2. List all subsets of $\{a, b, c, d\}$.

Problem A.3. Show that if X has n elements, then $\wp(X)$ has 2^n elements.

Problem A.4. Prove rigorously that if $X \subseteq Y$, then $X \cup Y = Y$.

Problem A.5. Prove rigorously that if $X \subseteq Y$, then $X \cap Y = X$.

Problem A.6. List all elements of $\{1, 2, 3\}^3$.

Problem A.7. Show, by induction on k, that for all $k \geq 1$, if X has n elements, then X^k has n^k elements.

APPENDIX B

Relations

B.1 Relations as Sets

You will no doubt remember some interesting relations between objects of some of the sets we've mentioned. For instance, numbers come with an *order relation* $<$ and from the theory of whole numbers the relation of *divisibility without remainder* (usually written $n \mid m$) may be familar. There is also the relation *is identical with* that every object bears to itself and to no other thing. But there are many more interesting relations that we'll encounter, and even more possible relations. Before we review them, we'll just point out that we can look at relations as a special sort of set. For this, first recall what a *pair* is: if a and b are two objects, we can combine them into the *ordered pair* $\langle a, b\rangle$. Note that for ordered pairs the order *does* matter, e.g, $\langle a, b\rangle \neq \langle b, a\rangle$, in contrast to unordered pairs, i.e., 2-element sets, where $\{a, b\} = \{b, a\}$.

If X and Y are sets, then the *Cartesian product* $X \times Y$ of X and Y is the set of all pairs $\langle a, b\rangle$ with $a \in X$ and $b \in Y$. In particular, $X^2 = X \times X$ is the set of all pairs from X.

Now consider a relation on a set, e.g., the $<$-relation on the set $\mathbb{N}$ of natural numbers, and consider the set of all pairs of numbers $\langle n, m\rangle$ where $n < m$, i.e.,

$$R = \{\langle n, m\rangle : n, m \in \mathbb{N} \text{ and } n < m\}.$$

Then there is a close connection between the number n being less than a number m and the corresponding pair $\langle n, m\rangle$ being a member of R, namely, $n < m$ if and only if $\langle n, m\rangle \in R$. In a sense we can consider the set R to *be* the $<$-relation on the set $\mathbb{N}$. 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 X is a subset of X^2. If $R \subseteq X^2$ is a binary relation on X and $x, y \in X$, we 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:

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

The subset consisting of the pairs lying on the diagonal, i.e.,

$$\{\langle 0, 0\rangle, \langle 1, 1\rangle, \langle 2, 2\rangle, \ldots\},$$

is the *identity relation on* $\mathbb{N}$. (Since the identity relation is popular, let's define $\mathrm{Id}_X = \{\langle x, x\rangle : x \in X\}$ for any set X.) The subset of all pairs lying above the diagonal, i.e.,

$$L = \{\langle 0, 1\rangle, \langle 0, 2\rangle, \ldots, \langle 1, 2\rangle, \langle 1, 3\rangle, \ldots, \langle 2, 3\rangle, \langle 2, 4\rangle, \ldots\},$$

is the *less than* relation, i.e., Lnm iff $n < m$. The subset of pairs below the diagonal, i.e.,

$$G = \{\langle 1, 0\rangle, \langle 2, 0\rangle, \langle 2, 1\rangle, \langle 3, 0\rangle, \langle 3, 1\rangle, \langle 3, 2\rangle, \ldots\},$$

is the *greater than* relation, i.e., Gnm iff $n > m$. The union of L with I, $K = L \cup I$, 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*. 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 X^2 is a relation on X, 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 X^2 itself is a relation on X 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 $\wp(X)$ 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 $R \subseteq X^2$ is *reflexive* iff, for every $x \in X$, Rxx.

Definition B.4 (Transitivity). A relation $R \subseteq X^2$ is *transitive* iff, whenever Rxy and Ryz, then also Rxz.

Definition B.5 (Symmetry). A relation $R \subseteq X^2$ is *symmetric* iff, whenever Rxy, then also Ryx.

Definition B.6 (Anti-symmetry). A relation $R \subseteq X^2$ 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 $R \subseteq X^2$ is *connected* if for all $x, y \in X$, if $x \neq y$, then either Rxy or Ryx.

Definition B.8 (Partial order). A relation $R \subseteq X^2$ that is reflexive, transitive, and anti-symmetric is called a *partial order*.

Definition B.9 (Linear order). A partial order that is also connected is called a *linear order.*

Definition B.10 (Equivalence relation). A relation $R \subseteq X^2$ that is reflexive, symmetric, and transitive is called an *equivalence relation.*

B.3 Orders

Very often we are interested in comparisons between objects, where one object may be less or equal or greater than another in a certain respect. Size is the most obvious example of such a comparative relation, or *order.* But not all such relations are alike in all their properties. For instance, some comparative relations require any two objects to be comparable, others don't. (If they do, we call them *linear* or *total.*) Some include identity (like $\leq$) and some exclude it (like $<$). Let's get some order into all this.

Definition B.11 (Preorder). A relation which is both reflexive and transitive is called a *preorder.*

Definition B.12 (Partial order). A preorder which is also antisymmetric is called a *partial order.*

Definition B.13 (Linear order). A partial order which is also connected is called a *total order* or *linear order.*

Example B.14. 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 X is a preorder, since it is reflexive and transitive. But, if X has more than one element , the universal relation is not anti-symmetric, and so not a partial order. For a somewhat less silly example, consider the *no longer than* relation $\preccurlyeq$ on $\mathbb{B}^*$: $x \preccurlyeq y$ iff $\mathrm{len}(x) \leq \mathrm{len}(y)$. 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 $01 \neq 10$.

The relation of *divisibility without remainder* gives us an example of a partial order which isn't a linear order: for integers n, m, we say n (evenly) divides m, in symbols: $n \mid m$, if there is some 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 anti-symmetry fails: $1 \mid -1$ and $-1 \mid 1$ but $1 \neq -1$. Another important partial order is the relation $\subseteq$ on a set of sets.

Notice that the examples L and G from Example B.2, although we said there that they were called "strict orders," are not linear orders even though they are connected (they are not reflexive). But there is a close connection, as we will see momentarily.

Definition B.15 (Irreflexivity). A relation R on X is called *irreflexive* if, for all $x \in X$, $\neg Rxx$.

Definition B.16 (Asymmetry). A relation R on X is called *asymmetric* if for no pair $x, y \in X$ we have Rxy and Ryx.

Definition B.17 (Strict order). A *strict order* is a relation which is irreflexive, asymmetric, and transitive.

Definition B.18 (Strict linear order). A strict order which is also connected is called a *strict linear order.*

A strict order on X can be turned into a partial order by adding the diagonal Id_X, 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_X.

Proposition B.19.

1. *If R is a strict (linear) order on X, then $R^+ = R \cup \mathrm{Id}_X$ is a partial order (linear order).*
2. *If R is a partial order (linear order) on X, then $R^- = R \setminus \mathrm{Id}_X$ is a strict (linear) order.*

Proof.

1. Suppose R is a strict order, i.e., $R \subseteq X^2$ and R is irreflexive, asymmetric, and transitive. Let $R^+ = R \cup \mathrm{Id}_X$. We have to show that R^+ is reflexive, antisymmetric, and transitive.

   R^+ is clearly reflexive, since for all $x \in X$, $\langle x, x\rangle \in \mathrm{Id}_X \subseteq R^+$.

   To show R^+ is antisymmetric, suppose R^+xy and R^+yx, i.e., $\langle x, y\rangle$ and $\langle y, x\rangle \in R^+$, and $x \neq y$. Since $\langle x, y\rangle \in R \cup \mathrm{Id}_X$, but $\langle x, y\rangle \notin \mathrm{Id}_X$, we must have $\langle x, y\rangle \in R$, i.e., Rxy. Similarly we get that Ryx. But this contradicts the assumption that R is asymmetric.

   Now suppose that R^+xy and R^+yz. If both $\langle x, y\rangle \in R$ and $\langle y, z\rangle \in R$, it follows that $\langle x, z\rangle \in R$ since R is transitive.

Otherwise, either $\langle x, y\rangle \in \mathrm{Id}_X$, i.e., $x = y$, or $\langle y, z\rangle \in \mathrm{Id}_X$, i.e., $y = z$. In the first case, we have that R^+yz by assumption, $x = y$, hence R^+xz. Similarly in the second case. In either case, R^+xz, thus, R^+ is also transitive.

If R is connected, then 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.

2. Exercise.

□

Example B.20. $\leq$ is the linear order corresponding to the strict linear order $<$. $\subseteq$ is the partial order corresponding to the strict order $\subsetneq$.

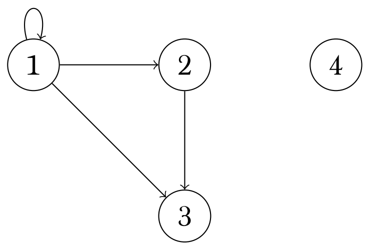
B.4 Graphs

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.

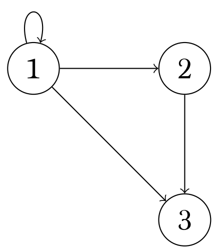
Definition B.21 (Directed graph). A *directed graph* $G = \langle V, E\rangle$ is a set of *vertices* V and a set of *edges* $E \subseteq V^2$.

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.22. 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.5 Operations on Relations

It is often useful to modify or combine relations. We've already used the union of relations above (which is just the union of two relations considered as sets of pairs). Here are some other ways:

Definition B.23. Let R, $S \subseteq X^2$ be relations and Y a set.

1. The *inverse* R^{-1} of R is $R^{-1} = \{\langle y, x\rangle : \langle x, y\rangle \in R\}$.
2. The *relative product* $R \mid S$ of R and S is

$$(R \mid S) = \{\langle x, z\rangle : \text{for some } y, Rxy \text{ and } Syz\}$$

3. The *restriction* $R \restriction Y$ of R to Y is $R \cap Y^2$
4. The *application* $R[Y]$ of R to Y is

$$R[Y] = \{y : \text{for some } x \in Y, Rxy\}$$

Example B.24. Let $S \subseteq \mathbb{Z}^2$ be the successor relation on $\mathbb{Z}$, i.e., the set of pairs $\langle x, y\rangle$ where $x + 1 = y$, for $x, y \in \mathbb{Z}$. Sxy holds iff y is the successor of x.

1. The inverse S^{-1} of S is the predecessor relation, i.e., $S^{-1}xy$ iff $x - 1 = y$.
2. The relative product $S \mid S$ is the relation x bears to y if $x + 2 = y$.
3. The restriction of S to $\mathbb{N}$ is the successor relation on $\mathbb{N}$.
4. The application of S to a set, e.g., $S[\{1, 2, 3\}]$ is $\{2, 3, 4\}$.

Definition B.25 (Transitive closure). The *transitive closure* R^+ of a relation $R \subseteq X^2$ is $R^+ = \bigcup_{i=1}^{\infty} R^i$ where $R^1 = R$ and $R^{i+1} = R^i \mid R$.

The *reflexive transitive closure* of R is $R^* = R^+ \cup \mathrm{Id}_X$.

Example B.26. Take the successor relation $S \subseteq \mathbb{Z}^2$. S^2xy iff $x + 2 = y$, S^3xy iff $x + 3 = y$, etc. So R^*xy iff for some $i \geq 1$, $x + i = y$. In other words, S^+xy iff $x < y$ (and R^*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. Complete the proof of Proposition B.19, i.e., prove that if R is a partial order on X, then $R^{-} = R \setminus \mathrm{Id}_X$ is a strict order.

Problem B.4. 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.5. 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 proof system—probably a natural deduction or Fitch style proof 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 proof 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 proof 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 to 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 some assumption that clarifies about which kinds of things we're proving something. 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 about, 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 x, write "assume that x," or "suppose that x"). 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 X and Y, $X \cup Y = Y \cup X$.*

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 X and Y are *identical*, $X = Y$, iff every element of X is an element of Y, and vice versa.

This definition uses X and Y as placeholders for arbitrary sets. What it defines—the *definiendum*—is the expression "$X = Y$" by giving the condition under which $X = Y$ is true. This condition—"every element of X is an element of Y, and vice versa"—is the *definiens.*[1] The definition specifies that $X = Y$ is true if, and only if (we abbreviate this to "iff") the condition holds.

When you apply the definition, you have to match the X and Y in the definition to the case you're dealing with. In our case, it means that in order for $X \cup Y = Y \cup X$ to be true, each $z \in X \cup Y$ must also be in $Y \cup X$, and vice versa. The expression $X \cup Y$ in the proposition plays the role of X in the definition, and $Y \cup X$ that of Y. Since X and Y 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

[1] In this particular case—and very confusingly!—when $X = Y$, the sets X and Y 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.

prove the proposition by showing that every element of X is an element of Y, and vice versa—that would show that $X = Y$, not that $X \cup Y = Y \cup X$. (Also, since X and Y may be any two sets, you won't get very far, because if nothing is assumed about X and Y 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, $X \cup Y$ is defined as $\{z : z \in X \text{ or } z \in Y\}$. So if you want to prove that $x \in X \cup Y$, unpacking the definition of $\cup$ tells you that you have to prove $x \in \{z : z \in X \text{ or } z \in Y\}$. Now you also have to remember that $x \in \{z : \ldots z \ldots\}$ iff $\ldots x \ldots$. So, further unpacking the definition of the $\{z : \ldots z \ldots\}$ notation, what you have to show is: $x \in X$ or $x \in Y$. So, "every element of $X \cup Y$ is also an element of $Y \cup X$" really means: "for every x, if $x \in X$ or $x \in Y$, then $x \in Y$ or $x \in X$." 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 X and Y: (a) for every x, if $x \in X$ or $x \in Y$, then $x \in Y$ or $x \in X$, and (b) for every x, if $x \in Y$ or $x \in X$, then $x \in X$ or $x \in Y$.*

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 $U = V$ (a). Then we may apply the definition of = for sets and infer: "Thus, by definition from (a), every element of U is an element of V 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 $U = V$, but we want to. If $U = V$ is the conclusion we aim for, then we can restate this aim also by applying the definition: to prove $U = V$ we have to prove that every element of U is an element of V and vice versa. So our proof will have the form: (a) prove that every element of U is an element of V; (b) every element of V is an element of U; (c) therefore, from (a) and (b) by definition of =, $U = V$. But we would usually not write it this way. Instead we might write something like,

> We want to show $U = V$. By definition of =, this amounts to showing that every element of U is an element of V and vice versa.
>
> (a) ... (a proof that every element of U is an element of V) ...
>
> (b) ... (a proof that every element of V is an element of U) ...

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 $U = V$ we've established that every element of U is an element of V and vice versa. From this we can conclude that every element of V is an element of U (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 $U = V$ you will find that, after unpacking the definition of =, you have to prove: every element of U is an element of V *and* every element of V is an element of U).

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 $X \subseteq Y$, that means (unpacking the definition of $\subseteq$), that, for every $x \in X$, $x \in Y$. Thus, if you already know that $z \in X$, you can conclude $z \in Y$.

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, X 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 all every P has the property Q. In short, x is a stand-in for *all* things with property P.

Proposition C.4. *For all sets X and Y, $X \subseteq X \cup Y$.*

Proof. Let X and Y be arbitrary sets. We want to show that $X \subseteq X \cup Y$. By definition of $\subseteq$, this amounts to: for every x, if $x \in X$ then $x \in X \cup Y$. So let $x \in X$ be an arbitrary element of X. We have to show that $x \in X \cup Y$. Since $x \in X$, $x \in X$ or $x \in Y$. Thus, $x \in \{x : x \in X \lor x \in Y\}$. But that, by definition of $\cup$, means $x \in X \cup Y$. □

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 $x \in Y \cup Z$. $Y \cup Z$ is defined as $\{x : x \in Y \text{ or } x \in Z\}$. In other words, by definition, $x \in Y$ or $x \in Z$. We would prove that $x \in X$ from this by first assuming that $x \in Y$, and proving $x \in X$ from this assumption, and then assume $x \in Z$, and again prove $x \in X$ from this. You would write "We distinguish cases" under the assumption, then "Case (1): $x \in Y$" underneath, and "Case (2): $x \in Z$ 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* $Y \subseteq U$ *and* $Z \subseteq V$. *Then* $Y \cup Z \subseteq U \cup V$.

Proof. Assume (a) that $Y \subseteq U$ and (b) $Z \subseteq V$. By definition, any $x \in Y$ is also $\in U$ (c) and any $x \in Z$ is also $\in V$ (d). To show that $Y \cup Z \subseteq U \cup V$, we have to show that if $x \in Y \cup Z$ then $x \in U \cup V$ (by definition of $\subseteq$). $x \in Y \cup Z$ iff $x \in Y$ or $x \in Z$ (by definition of $\cup$). Similarly, $x \in U \cup V$ iff $x \in U$ or $x \in V$. So, we have to show: for any x, if $x \in Y$ or $x \in Z$, then $x \in U$ or $x \in V$.

> 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 $x \in Y$ or $x \in Z$ and show that $x \in U$ or $x \in V$.[2]

Suppose that $x \in Y$ or $x \in Z$. We have to show that $x \in U$ or $x \in V$. We distinguish cases.

Case 1: $x \in Y$. By (c), $x \in U$. Thus, $x \in U$ or $x \in V$. (Here we've made the inference discussed in the preceding subsection!)

Case 2: $x \in Z$. By (d), $x \in V$. Thus, $x \in U$ or $x \in V$. □

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

[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.

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 $x \in Y$. Then there is an X such that $X \subseteq Y$ and $X \neq \emptyset$.*

Proof. Assume $x \in Y$. Let $X = \{x\}$.

> Here we've defined the set X 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 X here. However, we still have to show that X has the properties required by the proposition. The proof isn't complete without that!

Since $x \in X$, $X \neq \emptyset$.

> This relies on the definition of X as $\{x\}$ and the obvious facts that $x \in \{x\}$ and $x \notin \emptyset$.

Since x is the only element of $\{x\}$, and $x \in Y$, every element of X is also an element of Y. By definition of $\subseteq$, $X \subseteq Y$. □

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, $x \in X$" or "there is an $x \in X$." 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 X 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 $\in X$). 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, $x \in X$" to "Let $a \in X$." 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 $a \in X$, and you've shown that it follows just from the assumption "for some x, $x \in X$."

Proposition C.7. *If $X \neq \emptyset$, then $X \cup Y \neq \emptyset$.*

Proof. Suppose $X \neq \emptyset$. So for some x, $x \in X$.

> Here we first just restated the hypothesis of the proposition. This hypothesis, i.e., $X \neq \emptyset$, hides an existential claim, which you get to only by unpacking a few definitions. The definition of $=$ tells us that $X = \emptyset$ iff every $x \in X$ is also $\in \emptyset$ and every $x \in \emptyset$ is also $\in X$. Negating both sides, we get: $X \neq \emptyset$ iff either some $x \in X$ is $\notin \emptyset$ or some $x \in \emptyset$ is $\notin X$. Since nothing is $\in \emptyset$, the second disjunct can never be true, and "$x \in X$ and $x \notin \emptyset$" reduces to just $x \in X$. So $x \neq \emptyset$ iff for some x, $x \in X$. That's an existence claim. Now we use that existence claim by introducing a name for one of the elements of X:

Let $a \in X$.

> Now we've introduced a name for one of the things $\in X$. We'll continue to argue about a, but we'll be careful to only assume that $a \in X$ and nothing else:

Since $a \in X$, $a \in X \cup Y$, by definition of $\cup$. So for some x, $x \in X \cup Y$, i.e., $X \cup Y \neq \emptyset$.

> In that last step, we went from "$a \in X \cup Y$" to "for some x, $x \in X \cup Y$." That doesn't mention a anymore, so we know that "for some x, $x \in X \cup Y$" follows from "for some x, $x \in X$ alone." But that means that $X \cup Y \neq \emptyset$.

□

It's maybe good practice to keep bound variables like "x" separate from hypothtical names like a, like we did. In practice, however, we often don't and just use x, like so:

> Suppose $X \neq \emptyset$, i.e., there is an $x \in X$. By definition of $\cup$, $x \in X \cup Y$. So $X \cup Y \neq \emptyset$.

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 $X \neq \emptyset$ and $Y \neq \emptyset$ then $X \cap Y \neq \emptyset$" (which is not true).

> Suppose $X \neq \emptyset$ and $Y \neq \emptyset$. So for some x, $x \in X$ and also for some x, $x \in Y$. Since $x \in X$ and $x \in Y$, $x \in X \cap Y$, by definition of $\cap$. So $X \cap Y \neq \emptyset$.

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 X, Y, and Z, $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$*

Let's prove it!

Proof. We want to show that for any sets X, Y, and Z, $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$

> 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 $X \cup (Y \cap Z)$ are also elements of $(X \cup Y) \cap (X \cup Z)$, and vice versa. The "vice versa" means that also every element of $(X \cup Y) \cap (X \cup Z)$ must be an element of $X \cup (Y \cap Z)$. So in unpacking the definition, we see that we have to prove a conjunction. Let's record this:

By definition, $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$ iff every element of $X \cup (Y \cap Z)$ is also an element of $(X \cup Y) \cap (X \cup Z)$, and every element of $(X \cup Y) \cap (X \cup Z)$ is an element of $X \cup (Y \cap Z)$.

> Since this is a conjunction, we must prove each conjunct separately. Lets start with the first: let's prove that every element of $X \cup (Y \cap Z)$ is also an element of $(X \cup Y) \cap (X \cup Z)$.
>
> This is a universal claim, and so we consider an arbitrary element of $X \cup (Y \cap Z)$ and show that it must also be an element of $(X \cup Y) \cap (X \cup Z)$. We'll pick a variable to call this arbitrary element by, say, z. Our proof continues:

First, we prove that every element of $X \cup (Y \cap Z)$ is also an element of $(X \cup Y) \cap (X \cup Z)$. Let $z \in X \cup (Y \cap Z)$. We have to show that $z \in (X \cup Y) \cap (X \cup Z)$.

> Now it is time to unpack the definition of $\cup$ and $\cap$. For instance, the definition of $\cup$ is: $X \cup Y = \{z : z \in X \text{ or } z \in Y\}$. When we apply the definition to "$X \cup (Y \cap Z)$," the role of the "Y" in the definition is now played by "$Y \cap Z$," so $X \cup (Y \cap Z) = \{z : z \in X \text{ or } z \in Y \cap Z\}$. So our assumption that $z \in X \cup (Y \cap Z)$ amounts to: $z \in \{z : z \in X \text{ or } z \in Y \cap Z\}$. And $z \in \{z : \ldots z \ldots\}$ iff $\ldots z \ldots$, i.e., in this case, $z \in X$ or $z \in Y \cap Z$.

By the definition of $\cup$, either $z \in X$ or $z \in Y \cap Z$.

> 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, "$z \in (X \cup Y) \cap (X \cup Z)$") obtains.

Case 1: Suppose that $z \in X$.

> 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 $z \in (X \cup Y) \cap (X \cup Z)$. Based on the definition of $\cap$, if we want to show that $z \in (X \cup Y) \cap (X \cup Z)$, we have to show that it's in both $(X \cup Y)$ and $(X \cup Z)$. But $z \in X \cup Y$ iff $z \in X$ or $z \in Y$, and we already have (as the assumption of case 1) that $z \in X$. By the same reasoning—switching Z for Y—$z \in X \cup Z$. This argument went in the reverse direction, so let's record our reasoning in the direction needed in our proof.

Since $z \in X$, $z \in X$ or $z \in Y$, and hence, by definition of $\cup$, $z \in X \cup Y$. Similarly, $z \in X \cup Z$. But this means that $z \in (X \cup Y) \cap (X \cup Z)$, 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 $z \in Y \cap Z$.

Case 2: Suppose that $z \in Y \cap Z$.

> Again, we are working with the intersection of two sets. Let's apply the definition of $\cap$:

Since $z \in Y \cap Z$, z must be an element of both Y and Z, by definition of $\cap$.

> It's time to look at our conclusion again. We have to show that z is in both $(X \cup Y)$ and $(X \cup Z)$. And again, the solution is immediate.

Since $z \in Y$, $z \in (X \cup Y)$. Since $z \in Z$, also $z \in (X \cup Z)$. So, $z \in (X \cup Y) \cap (X \cup Z)$.

> 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 $z \in X \cup (Y \cap Z)$ then $z \in (X \cup Y) \cap (X \cup Z)$.

> Now we just want to show the other direction, that every element of $(X \cup Y) \cap (X \cup Z)$ is an element of $X \cup (Y \cap Z)$. 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 $z \in (X \cup Y) \cap (X \cup Z)$. We want to show that $z \in X \cup (Y \cap Z)$.

> We are now working from the hypothesis that $z \in (X \cup Y) \cap (X \cup Z)$. 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 $X \cup Y$ and $X \cup Z$, by definition of $\cap$. And by the definition of $\cup$, we can further unpack this to: either $z \in X$ or $z \in Y$, and also either $z \in X$ or $z \in Z$. 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 X, Y or Z. But that would be a mistake. We have to be careful, so let's consider each disjunction in turn.

By definition of $\cap$, $z \in X \cup Y$ and $z \in X \cup Z$. By definition of $\cup$, $z \in X$ or $z \in Y$. We distinguish cases.

> Since we're focusing on the first disjunction, we haven't gotten our second disjunction (from unpacking $X \cup Z$) yet. In fact, we don't need it yet. The first case is $z \in X$, 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 $z \in X$. It follows that $z \in X \cup (Y \cap Z)$.

> Now for the second case, $z \in Y$. Here we'll unpack the second $\cup$ and do another proof-by-cases:

Case 2: Suppose that $z \in Y$. Since $z \in X \cup Z$, either $z \in X$ or $z \in Z$. We distinguish cases further:

Case 2a: $z \in X$. Then, again, $z \in X \cup (Y \cap Z)$.

> Ok, this was a bit weird. We didn't actually need the assumption that $z \in Y$ for this case, but that's ok.

Case 2b: $z \in Z$. Then $z \in Y$ and $z \in Z$, so $z \in Y \cap Z$, and consequently, $z \in X \cup (Y \cap Z)$.

> This concludes both proofs-by-cases and so we're done with the second half.

So, if $z \in (X \cup Y) \cap (X \cup Z)$ then $z \in X \cup (Y \cap Z)$. □

C.6 Another Example

Proposition C.9. *If $X \subseteq Z$, then $X \cup (Z \setminus X) = Z$.*

Proof. Suppose that $X \subseteq Z$. We want to show that $X \cup (Z \setminus X) = Z$.

> We begin by observing that this is a conditional statement. It is tacitly universally quantified: the proposition holds for all sets X and Z. So X and Z 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 $X \subseteq Z$. Let's unpack the definition of $\subseteq$: the assumption means that all elements of X are also elements of Z. Let's write this down—it's an important fact that we'll use throughout the proof.

By the definition of $\subseteq$, since $X \subseteq Z$, for all z, if $z \in X$, then $z \in Z$.

> 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 $X \cup (Z \setminus X) = Z$, and so we set up a proof similarly to the last example: we show that every element of $X \cup (Z \setminus X)$ is also an element of Z and, conversely, every element of Z is an element of $X \cup (Z \setminus X)$. We can shorten this to: $X \cup (Z \setminus X) \subseteq Z$ and $Z \subseteq X \cup (Z \setminus X)$. (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 $X \cup (Z \setminus X)$ is also an element of Z, we assume that $z \in X \cup (Z \setminus X)$ for an arbitrary z and show that $z \in Z$. By the definition of $\cup$, we can conclude that $z \in X$ or $z \in Z \setminus X$ from $z \in X \cup (Z \setminus X)$. You should now be getting the hang of this.

$X \cup (Z \setminus X) = Z$ iff $X \cup (Z \setminus X) \subseteq Z$ and $Z \subseteq (X \cup (Z \setminus X)$. First we prove that $X \cup (Z \setminus X) \subseteq Z$. Let $z \in X \cup (Z \setminus X)$. So, either $z \in X$ or $z \in (Z \setminus X)$.

> We've arrived at a disjunction, and from it we want to prove that $z \in Z$. We do this using proof by cases.

Case 1: $z \in X$. Since for all z, if $z \in X$, $z \in Z$, we have that $z \in Z$.

> Here we've used the fact recorded earlier which followed from the hypothesis of the proposition that $X \subseteq Z$. The first case is complete, and we turn to the second case, $z \in (Z \setminus X)$. Recall that $Z \setminus X$ denotes the *difference* of the two sets, i.e., the set of all elements of Z which are not elements of X. But any element of Z not in X is in particular an element of Z.

Case 2: $z \in (Z \setminus X)$. This means that $z \in Z$ and $z \notin X$. So, in particular, $z \in Z$.

> Great, we've proved the first direction. Now for the second direction. Here we prove that $Z \subseteq X \cup (Z \setminus X)$. So we assume that $z \in Z$ and prove that $z \in X \cup (Z \setminus X)$.

Now let $z \in Z$. We want to show that $z \in X$ or $z \in Z \setminus X$.

> Since all elements of X are also elements of Z, and $Z \setminus X$ is the set of all things that are elements of Z but not X, it follows that z is either in X or in $Z \setminus X$. 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: $z \in X$ or $z \notin X$. 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 $z \in X$.) Since this is a disjunction, we can again use proof-by-cases.

Either $z \in X$ or $z \notin X$. In the former case, $z \in X \cup (Z \setminus X)$. In the latter case, $z \in Z$ and $z \notin X$, so $z \in Z \setminus X$. But then $z \in X \cup (Z \setminus X)$.

> Our proof is complete: we have shown that $X \cup (Z \setminus X) = Z$. □

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 $X \subseteq Y$ and $Y = \emptyset$, then X has no elements.*

Proof. Suppose $X \subseteq Y$ and $Y = \emptyset$. We want to show that X has no elements.

> Since this is a conditional claim, we assume the antecedent and want to prove the consequent. The consequent is: X has no elements. We can make that a bit more explicit: it's not the case that there is an $x \in X$.

X has no elements iff it's not the case that there is an x such that $x \in X$.

> 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 $x \in X$. To use proof by contradiction, we have to assume the corresponding positive claim p, i.e., there is an $x \in X$, 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 $x \in X$.

> This is now the new assumption we'll use to obtain a contradiction. We have two more assumptions: that $X \subseteq Y$ and that $Y = \emptyset$. The first gives us that $x \in Y$:

Since $X \subseteq Y$, $x \in Y$.

> But since $Y = \emptyset$, every element of Y (e.g., x) must also be an element of $\emptyset$.

Since $Y = \emptyset$, $x \in \emptyset$. 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 ($X \subseteq Y$ and $Y = \emptyset$) are not contested, it must be the last assumption introduced (there is an $x \in X$) that must be false. But if we want to be thorough, we can spell this out.

Thus, our assumption that there is an $x \in X$ must be false, hence, X 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 X has no elements, and so the assumption we made for the purpose of proof by contradiction (i.e., that there is an $x \in X$) was a positive claim. It gave us something to work with, namely the hypothetical $x \in X$ 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 "$X = \emptyset$" instead of the negative consequent "X has no elements." (By definition of $=$, "$X = \emptyset$" is a general claim, since it unpacks to "every element of X is an element of $\emptyset$ and vice versa".) But it is easily seen to be equivalent to the negative claim "not: there is an $x \in X$."

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 example), 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. $X \subseteq X \cup Y$.

Proof. We want to show that $X \subseteq X \cup Y$.

> On the face of it, this is a positive claim: every $x \in X$ is also in $X \cup Y$. The negation of that is: some $x \in X$ is $\notin X \cup Y$. So we can prove the claim indirectly by assuming this negated claim, and showing that it leads to a contradiction.

Suppose not, i.e., $X \nsubseteq X \cup Y$.

> We have a definition of $X \subseteq X \cup Y$: every $x \in X$ is also $\in X \cup Y$. To understand what $X \nsubseteq X \cup Y$ means, we have to use some elementary logical manipulation on the unpacked definition: it's false that every $x \in X$ is also $\in X \cup Y$ iff there is *some* $x \in X$ that is $\notin Z$. (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, $X \nsubseteq X \cup Y$ iff there is an x such that $x \in X$ and $x \notin X \cup Y$. From then on, it's easy.

So, there is an $x \in X$ such that $x \notin X \cup Y$. By definition of $\cup$, $x \in X \cup Y$ iff $x \in X$ or $x \in Y$. Since $x \in X$, we have $x \in X \cup Y$. This contradicts the assumption that $x \notin X \cup Y$. □

Proposition C.12. *If $X \subseteq Y$ and $Y \subseteq Z$ then $X \subseteq Z$.*

Proof. Suppose $X \subseteq Y$ and $Y \subseteq Z$. We want to show $X \subseteq Z$.

> Let's proceed indirectly: we assume the negation of what we want to etablish.

Suppose not, i.e., $X \nsubseteq Z$.

> As before, we reason that $X \nsubseteq Z$ iff not every $x \in X$ is also $\in Z$, i.e., some $x \in X$ is $\notin Z$. 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 $x \in X$ and $x \notin Z$.

> 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 $X \subseteq Y$, $x \in Y$. Since $Y \subseteq Z$, $x \in Z$. But this contradicts $x \notin Z$. □

Proposition C.13. *If $X \cup Y = X \cap Y$ then $X = Y$.*

Proof. Suppose $X \cup Y = X \cap Y$. We want to show that $X = Y$.

> The beginning is now routine:

Assume, by way of contradiction, that $X \neq Y$.

> Our assumption for the proof by contradiction is that $X \neq Y$. Since $X = Y$ iff $X \subseteq Y$ an $Y \subseteq X$, we get that $X \neq Y$ iff $X \nsubseteq Y$ *or* $Y \nsubseteq X$. (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.

$X \neq Y$ iff $X \nsubseteq Y$ or $Y \nsubseteq X$. We distinguish cases.

> In the first case, we assume $X \nsubseteq Y$, i.e., for some x, $x \in X$ but $\notin Y$. $X \cap Y$ is defined as those elements that X and Y have in common, so if something isn't in one of them, it's not in the intersection. $X \cup Y$ is X together with Y, so anything in either is also in the union. This tells us that $x \in X \cup Y$ but $x \notin X \cap Y$, and hence that $X \cap Y \neq Y \cap X$.

Case 1: $X \nsubseteq Y$. Then for some x, $x \in X$ but $x \notin Y$. Since $x \notin Y$, then $x \notin X \cap Y$. Since $x \in X$, $x \in X \cup Y$. So, $X \cap Y \neq Y \cap X$, contradicting the assumption that $X \cap Y = X \cup Y$.

Case 2: $Y \nsubseteq X$. Then for some y, $y \in Y$ but $y \notin X$. As before, we have $y \in X \cup Y$ but $y \notin X \cap Y$, and so $X \cap Y \neq X \cup Y$, again contradicting $X \cap Y = X \cup Y$. □

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 ofen 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 X, Y,*

$$X \cap (X \cup Y) = X$$

Proof. If $z \in X \cap (X \cup Y)$, then $z \in X$, so $X \cap (X \cup Y) \subseteq X$. Now suppose $z \in X$. Then also $z \in X \cup Y$, and therefore also $z \in X \cap (X \cup Y)$. □

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 X and Y, and when the proof mentions X or Y, 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 $z \in X \cap (X \cup Y)$, then $z \in X$, so $X \cap (X \cup Y) \subseteq X$."

This is the first half of the proof of the identity: it estabishes that if an arbitrary z is an element of the left side, it is also an element of the right, i.e., $X \cap (X \cup Y) \subseteq X$. Assume that $z \in X \cap (X \cup Y)$. Since z is an element of the intersection of two sets iff it is an element of both sets, we can conclude that $z \in X$ and also $z \in X \cup Y$. In particular, $z \in X$, 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 $X \subseteq X \cap (X \cup Y)$.

> "Now suppose $z \in X$. Then also $z \in X \cup Y$, and therefore also $z \in X \cap (X \cup Y)$."

We start by assuming that $z \in X$, since we are showing that, for any z, if $z \in X$ then $z \in X \cap (X \cup Y)$. To show that $z \in X \cap (X \cup Y)$, we have to show (by definition of "$\cap$") that (i) $z \in X$ and also (ii) $z \in X \cup Y$. 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 $z \in X$, and $X \cup Y$ is the union of X and Y, this is the case here. So $z \in X \cup Y$. We've shown both (i) $z \in X$ and (ii) $z \in X \cup Y$, hence, by definition of "$\cap$," $z \in X \cap (X \cup Y)$. 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 $z \in X$ that $z \in X \cup Y$, and from $z \in X$ and $z \in X \cup Y$ that $z \in X \cap (X \cup Y)$.

Here's another version of the proof above, with everything made explicit:

Proof. [By definition of = for sets, $X \cap (X \cup Y) = X$ we have to show (a) $X \cap (X \cup Y) \subseteq X$ and (b) $X \cap (X \cup Y) \subseteq X$. (a): By definition of $\subseteq$, we have to show that if $z \in X \cap (X \cup Y)$, then $z \in X$.] If $z \in X \cap (X \cup Y)$, then $z \in X$ [since by definition of $\cap$, $z \in X \cap (X \cup Y)$ iff $z \in X$ and $z \in X \cup Y$], so $X \cap (X \cup Y) \subseteq X$. [(b): By definition of $\subseteq$, we have to show that if $z \in X$, then $z \in X \cap (X \cup Y)$.] Now suppose [(1)] $z \in X$. Then also [(2)] $z \in X \cup Y$ [since by (1) $z \in X$ or $z \in Y$, which by definition of $\cup$ means $z \in X \cup Y$], and therefore also $z \in X \cap (X \cup Y)$ [since the definition of $\cap$ requires that $z \in X$, i.e., (1), and $z \in X \cup Y$), 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 the 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* by Daniel Solow and *How to Prove It: A Structured Approach* by Daniel Velleman in particular. The *Book of Proof* by Richard Hammack and *Mathematical Reasoning* by Ted Sundstrom are books on proof that are freely available. Philosophers might find *More Precisely: The Math you need to do Philosophy* by Eric Steinhart 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" by Michael Hutchings and "How to write proofs" by Eugenia Chang.

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 $X \cap Y \neq \emptyset$. Unpack all the definitions occuring here, i.e., restate this in a way that does not mention "$\cap$", "=", or "$\emptyset$.

Problem C.2. Prove *indirectly* that $X \cap Y \subseteq X$.

Problem C.3. Expand the following proof of $X \cup (X \cap Y) = X$, 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 $z \in X \cup (X \cap Y)$ then $z \in X$ or $z \in X \cap Y$. If $z \in X \cap Y$, $z \in X$. Any $z \in X$ is also $\in X \cup (X \cap Y)$. $\square$

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 particlar 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 that 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 n has the property, so does $n + 1$. Induction on natural numbers can then also often be used to prove general 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 suchh 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 $P(n)$, 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 n, if $P(n)$ then $P(n + 1)$.

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 n, if $P(n)$ then $P(n + 1)$" by putting 0 for n. 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 k, after doing this k steps, we eventually arrive at $P(k)$. So (1) and (2) together establish $P(k)$ for any $k \in \mathbb{N}$.

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 5 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 $s_n = n(n+1)/2$ then $s_{n+1} = (n+1)(n+2)/2$. (1) is obvious. To prove (2), we assume the inductive hypothesis: $s_n = n(n+1)/2$. Using it, we have to show that $s_{n+1} = (n+1)(n+2)/2$.

What is s_{n+1}? By the definition, $s_{n+1} = s_n + (n+1)$. By inductive hypothesis, $s_n = n(n+1)/2$. We can substitute this into the previous equation, and then just need a bit of arithmetic of fractions:

$$\begin{aligned} s_{n+1} &= \frac{n(n+1)}{2} + (n+1) = \\ &= \frac{n(n+1)}{2} + \frac{2(n+1)}{2} = \\ &= \frac{n(n+1) + 2(n+1)}{2} = \\ &= \frac{(n+2)(n+1)}{2}. \end{aligned}$$

□

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 $n+1$ to the case for n.

D.3 Strong Induction

In the principle of induction discussed above, we prove $P(0)$ and also if $P(n)$, then $P(n+1)$. In the second part, we assume that $P(n)$ is true and use this assumption to prove $P(n+1)$. Equivalently, of course, we could assume $P(n-1)$ and use it to prove $P(n)$—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 $n - 1$ of n, but for all numbers smaller than n, and use this assumption to establish the claim for n. This also gives us the claim $P(k)$ for all $k \in \mathbb{N}$. 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 $l < n$ then $P(n)$, we know this in particular for $n = 1$. So we can conclude $P(2)$. With this we have proved $P(0)$, $P(1)$, $P(2)$, i.e., $P(l)$ for all $l < 3$, and since we have also the conditional, if $P(l)$ for all $l < 3$, then $P(3)$, we can conclude $P(3)$, and so on.

In fact, if we can establish the general conditional "for all n, if $P(l)$ for all $l < n$, then $P(n)$," we do not have to establish $P(0)$ anymore, since it follows from it. For remember that a general claim like "for all $l < n$, $P(l)$" is true if there are no $l < n$. 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 "$\forall l\,(l < n \rightarrow P(l))$"—and that is true if there are no $l < n$. And if $n = 0$ 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 $l < n$, then $P(n)$" thus automatically establishes $P(0)$.

This variant is useful if establishing the claim for n can't be made to just rely on the claim for $n - 1$ but may require the assumption that it is true for one or more $l < n$.

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 ∘, and brackets [and], such as "[[c ∘ d][", "[a[]∘]", "a" or "[[a ∘ b] ∘ 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 "∘" 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 ∘ we can find "nice" expressions on either side, surrounded by a pair of parenteses.

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 ∘ 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 and s' are nice terms, then so is $[s \circ s']$.
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 [a ∘ [a ∘ a]]—where t is a from step 1 and s is [a ∘ a] from step 2—and [[b ∘ c] ∘ [d ∘ b]] constructed out of the two terms [b ∘ c] and [d ∘ 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 ∘ 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 $[r \circ s]$ for some other pair of nice terms r and s. 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 $k < n$, any parexpression of length k has $k/2$ ['s, then any parexpression of length n has $n/2$ ['s.

To show this conditional, assume that its antecedent is true, i.e., assume that for any $k < n$, parexpressions of length k contain $< k/2$ ['s. We call this assumption the inductive hypothesis. We want to show the same is true for parexpressions of length n.

So suppose t is a nice term of length n. Because parexpressions are inductively defined, we have three two cases: (1) t is a letter by itself, or t is $[r \circ s]$ for some nice terms r and s.

1. t is a letter. Then $n = 1$, and the number of [in t is 0. Since $0 < 1/2$, the claim holds.

2. t is $[s \circ s']$ for some nice terms s and s'. Let's let k be the length of s and k' be the length of s'. Then the length n of t is $k + k' + 3$ (the lengths of s and s' plus three symbols [, $\circ$,]). Since $k + k' + 3$ is always greater than k, $k < n$. Similarly, $k' < n$. That means that the induction hypothesis applies to the terms s and s': the number m of [in s is $< k/2$, and the number of [in s' is $< k'/2$.

   The number of [in t is the number of [in s, plus the number of [in s', plus 1, i.e., it is $m + m' + 1$. Since $m < k/2$ and $m' < k'/2$ we have:

$$m + m' + 1 < \frac{k}{2} + \frac{k'}{2} + 1 = \frac{k + k' + 2}{2} < \frac{k + k' + 3}{2} = n/2.$$

In each case, we've shown that the number of [in t is $< n/2$ (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

$$o(s, s') = [s \circ s']$$

You can even think of the natural numbers $\mathbb{N}$ themselves as being given be 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 $[s \circ s']$ provided it is true of s and s' 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 operations 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 equals the number of], and the same is true for s'. The number of [in $o(s, s')$, i.e., in $[s \circ s']$, is the sum of the number of [in s and s'. The number of] in $o(s, s')$ is the sum of the number of] in s and s'. Thus, the number of [in $o(s, s')$ equals the number of] in $o(s, s')$.

□

Let's give another proof by structural induction: a proper initial segment of a string of symbols t is any string t' 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. $t = [s \circ s']$ for some nice terms s and s'. 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 $[r'$ where r' is a proper initial segment of s: Since s is a nice term, by induction hypothesis, r' has more [than] and the same is true for $[r'$.

   c) r is $[s$ or $[s \circ$: By the previous result, the number of [and] in s is equal; so the number of [in $[s$ or $[s \circ$ is one more than the number of].

   d) r is $[s \circ r'$ where r' is a proper initial segment of s': By induction hypothesis, r' contains more [than]. By the previous result, the number of [and of] in s is equal. So the number of [in $[s \circ r'$ is greater than the number of].

   e) r is $[s \circ s'$: By the previous result, the number of [and] in s is equal, and the same for s'. So there is one more [in $[s \circ s'$ 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 is a subterm of a nice term t' if it occurs as a part of it. Let's use a symbol for it: $t \sqsubseteq t'$. 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 being a subterm of t', $t \sqsubseteq t'$, is defined by induction on s' as follows:

1. If t' is a letter, then $t \sqsubseteq t'$ iff $t = t'$.
2. If t' is $[s \circ s']$, then $t \sqsubseteq t'$ iff $t = t'$, $t \sqsubseteq s$, or $t \sqsubseteq s'$.

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 and s' such that $t = [s \circ s']$. "Uniquely determined" here means that if $t = [s \circ s']$ it isn't *also* $= [r \circ r']$ with $s \neq r$ or $s' \neq r'$. If this were the case, then clause (2) may come in conflict with itself: reading t' as $[s \circ s']$ we might get $t \sqsubseteq t'$, but if we read t' as $[r \circ r']$ we might get not $t \sqsubseteq t'$. 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 bracketles term.
2. If s and s' are bracketless terms, then $s \circ s'$ is a bracketless term.
3. Nothing else is a bracketless term.

Bracketless terms are, e.g., a, b ∘ d, b ∘ a ∘ b. Now if we defined "subterm" for bracketless terms the way we did above, the second clause would read

$$\text{If } t' = s \circ s', \text{ then } t \sqsubseteq t' \text{ iff } t = t',\ t \sqsubseteq s, \text{ or } t \sqsubseteq s'.$$

Now $\mathrm{b} \circ \mathrm{a} \circ \mathrm{b}$ is of the form $s \circ s'$ with $s = \mathrm{b}$ and $s' = \mathrm{a} \circ \mathrm{b}$. It is also of the form $r \circ r'$ with $r = \mathrm{b} \circ \mathrm{a}$ and $r' = \mathrm{b}$. 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, s' such that $t = [s \circ s']$.*

Proof. If t is a letter by itself, the condition issatisfied. So assume t isn't a letter by itself. We can tell from the inductive definition that then t must be of the form $[s \circ s']$ for some nice terms s and s'. It remains to show that these are uniquely determined, i.e., if $t = [r \circ r']$, then $s = r$ and $s' = r'$.

So suppose $t = [s \circ s']$ and $t = [r \circ r']$ for nice terms s, s', r, r'. We have to show that $s = r$ and $s' = r'$. First, s and r must be identical, for otherwise one is a proper initial segment of the other. But by Proposition D.6, that is impossible if s and r are both nice terms. But if $s = r$, then clearly also $s' = r'$. □

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:

$$
\begin{aligned}
f(s) &= 0 \text{ if } s \text{ is a letter}\\
f([s \circ s'] &= \max(f(s), f(s')) + 1
\end{aligned}
$$

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 an s' are nice terms, and make use of the fact that every nice term is either a letter or of the form $[s \circ s']$. 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:

$$
\begin{aligned}
g(s) &= 0 \text{ if } s \text{ is a letter}\\
g(s \circ s') &= \max(g(s), g(s')) + 1
\end{aligned}
$$

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 $s \circ s'$ with $s = \mathrm{a}$ and $s' = \mathrm{b} \circ \mathrm{c} \circ \mathrm{d}$, or as $r \circ r'$ with $r = \mathrm{a} \circ b$ and $r' = \mathrm{c} \circ \mathrm{d}$. Calculating g according to the first way of reading it would give

$$
\begin{aligned}
g(s \circ s') &= \max(g(\mathrm{a}), g(\mathrm{b} \circ \mathrm{c} \circ \mathrm{d})) + 1 =\\
&= \max(0, 2) + 1 = 3
\end{aligned}
$$

while according to the other reading we get

$$
g(r \circ r') = \max(g(\mathrm{a} \circ \mathrm{b}), g(\mathrm{c} \circ \mathrm{d})) + 1 =
$$

$$= \max(1, 1) + 1 = 2$$

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 t and s are supernice terms, then so is $[t \circ s]$.
4. Nothing else is a supernice term.

Show that the number of [in a supernice term s 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 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).

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.

The *Open Logic Text* is a collaborative project and is under active development. 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 redistribute our work, as long as they give appropriate credit.

Please see the Open Logic Project website at openlogicproject.org for additional information.