PROPOSITIONAL LOGIC

1. Propositional formulas

Propositional logic is a syntactical system for expressing statements (called **formulas**) that are either true or false. We start with some **atomic formulas**

$$P, Q, R, \ldots$$

which we may then combine with connectives $\wedge$ ("and"), $\vee$ ("or"), and $\neg$ ("not") to get more complicated formulas like

$$\neg(P \wedge \neg R) \vee (P \vee (Q \wedge \neg P)).$$

The formal definition is as follows.

An **alphabet** $\mathcal{A}$ is simply an arbitrary set, but whose elements we think of as "symbols". Here are some alphabets:

$$\begin{gathered}\{P, Q, R\} \\ \{0,1,2,3,4,5,6,7,8,9\} \\ \mathbb{N} := \{0,1,2,3,4,5,6,7,8,9,10,11,\ldots\} \quad (\text{note: } 0 \in \mathbb{N}) \\ \{0, 1, e, \pi, +, -, \sin, \textstyle\int\}\end{gathered}$$

Definition 1.1. Let $\mathcal{A}$ be an alphabet. The **propositional $\mathcal{A}$-formulas** are certain syntactic expressions, constructed according to the following rules:

- Every $P \in \mathcal{A}$ is an $\mathcal{A}$-formula.
- If ϕ, ψ are $\mathcal{A}$-formulas, then $\phi \wedge \psi, \phi \vee \psi, \neg\phi$ are $\mathcal{A}$-formulas.
- $\top, \bot$ are $\mathcal{A}$-formulas (thought of as "0-ary" versions of $\wedge, \vee$, which are always true, false respectively).

We denote the set of all $\mathcal{A}$-formulas by $\mathcal{L}(\mathcal{A})$.

1.1. **Formal representations of formulas.**

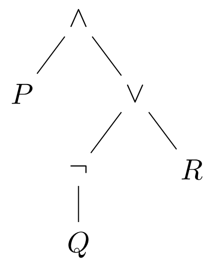
What exactly do we mean by "syntactic expression"? There are several options for formally representing formulas as familiar mathematical objects:

(1) By a formula like $P \wedge (\neg Q \vee R)$, we might mean the finite sequence of symbols

$$(P, \wedge, `(', \neg, Q, \vee, R, `)')$$

(over the extended alphabet $\mathcal{A} \cup \{\wedge, \vee, \neg, \top, \bot, `(', `)'\}$). This has the advantage of being close to what people actually write in practice. The disadvantage is that to be completely rigorous, one has to prove a bunch of technical lemmas explaining why e.g., the above formula has an unambiguous meaning, unlike, say, $P \wedge \neg Q \vee R$. Such issues are the subject of **parsing**, a fascinating area of computer science, which is however largely unrelated to (hence a distraction from) the main ideas of this course.

(2) Alternatively, we might represent $P \wedge (\neg Q \vee R)$ as the "syntax tree"



which is clearly unambiguous, but takes up a lot of space, and also depends on a formal definition of "tree".

(3) One way of formally defining "tree" is as a finite tuple (root, child, child, ...), where each "child" is itself a tree. Then $P \wedge (\neg Q \vee R)$ is represented as the highly nested tuple
$$(\wedge, P, (\vee, (\neg, Q), R)).$$

If we wanted to be fully rigorous, representation (3) is probably the most convenient one, since it captures well the "structure" of a formula. The above definition of $\mathcal{A}$-formula then becomes:

- Every $P \in \mathcal{A}$ is an $\mathcal{A}$-formula.
- If ϕ, ψ are $\mathcal{A}$-formulas, then $(\wedge, \phi, \psi), (\vee, \phi, \psi), (\neg, \phi)$ are $\mathcal{A}$-formulas.
- $\top, \bot$ are $\mathcal{A}$-formulas.

However, it is best to think of the *concept* of a formula as distinct from its concrete *representation* as a mathematical object; the precise representation we choose is largely irrelevant to the kinds of things we want to do in logic.[1][2] For this reason, we will continue to write formulas informally the way we usually do (e.g., $P \wedge (\neg Q \vee R)$), taking for granted that it is trivial to convert to/from these different representations.

Remark 1.2. We stress that formulas are *meaningless, purely syntactic expressions*, so that, e.g.,
$$P \wedge Q \neq Q \wedge P$$
(assuming $P, Q \in \mathcal{A}$ are different symbols). This can be seen from any of the above concrete representations: for example, according to (3), we have
$$(\wedge, P, Q) \neq (\wedge, Q, P).$$
(Of course, we will eventually want to assign these two distinct formulas the same *meaning*; that is, the function taking a formula to its meaning will not be an injection.)

1.2. **Some notational conventions.** We will regard $\neg$ as binding more tightly than $\wedge, \vee$; e.g.,
$$\neg\phi \wedge \psi = (\neg\phi) \wedge \psi \neq \neg(\phi \wedge \psi).$$

For formulas ϕ, ψ, we define the following abbreviations:
$$\begin{aligned} \phi \to \psi &:= \neg\phi \vee \psi, \\ \phi \leftrightarrow \psi &:= (\phi \to \psi) \wedge (\psi \to \phi). \end{aligned}$$

Note that we are *not* treating $\to, \leftrightarrow$ as logical connectives in their own right: in no formula does the symbol $\to$ actually appear. We will regard $\wedge, \vee$ as binding more tightly than $\to, \leftrightarrow$; e.g.,
$$\phi \wedge \psi \to \theta = (\phi \wedge \psi) \to \theta \neq \phi \wedge (\psi \to \theta).$$

[1] For set theorists: this is analogous to how an ordered pair (a, b) can be represented as $\{\{a\}, \{a, b\}\}$, or alternatively as $\{\{1, a\}, \{2, b\}\}$, or ...; which representation we choose is irrelevant to the way we normally use ordered pairs.

[2] For computer scientists: this is analogous to how a number like 1.5 may be represented as different sequences of bits on different CPU architectures; we normally shouldn't have to think about these issues in high-level programming.

Remark 1.3. Why do we choose to treat $\to$ as an abbreviation, but not $\phi \vee \psi := \neg(\neg\phi \wedge \neg\psi)$, say? Mostly for expository purposes: on several occasions, we will use $\vee$ and $\to$ to illustrate different aspects of basic versus derived connectives.

For formulas $\phi_1, \ldots, \phi_n$, we define

$$\phi_1 \wedge \cdots \wedge \phi_n := ((\phi_1 \wedge \phi_2) \wedge \phi_3) \wedge \cdots .$$

When $n = 1$, this is just ϕ_1. When $n = 0$, by convention, we take this to mean $\top$. Similarly for $\vee, \bot$. Note that the choice to associate to the left is completely arbitrary: if we had instead taken

$$\phi_1 \wedge \cdots \wedge \phi_n := \phi_1 \wedge (\phi_2 \wedge (\phi_3 \wedge \cdots)),$$

the resulting formula wouldn't be *equal* to the previous definition (as per Remark 1.2), but it will turn out to have the same meaning.

1.3. **Inductively defined sets.** Definition 1.1 of $\mathcal{A}$-formulas is an example of an **inductively defined set**. We are saying that the set $\mathcal{L}(\mathcal{A})$ of $\mathcal{A}$-formulas is the *smallest* set closed under the three conditions. We will see several such inductive definitions throughout this course.

Compare with the inductive definition of the natural numbers $\mathbb{N}$:

- 0 is a natural number.
- If n is a natural number, then so is the **successor** $n + 1$.[3]

Recall:

Principle of induction for $\mathbb{N}$. To prove a statement $\Phi(n)$ for all $n \in \mathbb{N}$:

- Prove $\Phi(0)$.
- Prove $\Phi(n) \implies \Phi(n + 1)$.

This follows directly from the definition of $\mathbb{N}$ as the *smallest* set containing 0 and closed under successor: we may consider the set $S := \{n \in \mathbb{N} \mid \Phi(n)\}$, which also contains 0 and is closed under successor by the two bullet points above, hence $\mathbb{N} \subseteq S$, i.e., $\Phi(n)$ for all $n \in \mathbb{N}$.

The inductive definition of $\mathcal{A}$-formulas yields an analogous

Principle of induction for propositional formulas. To prove a statement $\Phi(\phi)$ for all $\mathcal{A}$-formulas ϕ:

- Prove $\Phi(P)$ for all $P \in \mathcal{A}$.
- Prove that if $\Phi(\phi)$ and $\Phi(\psi)$, then $\Phi(\phi \wedge \psi), \Phi(\phi \vee \psi), \Phi(\neg\phi)$.
- Prove $\Phi(\top)$ and $\Phi(\bot)$.

Here is a (very simple) example of a proof by induction on formulas:

Proposition 1.4. In every $\mathcal{A}$-formula, only finitely many symbols from $\mathcal{A}$ appear.

Proof. By induction on formulas. For an atomic formula $P \in \mathcal{A}$, P is the only symbol that appears. If only finitely many symbols appear in ϕ, ψ (this is the **induction hypothesis**), then the symbols appearing in $\phi \wedge \psi$ are the combination of these, hence still finite. Similarly for $\phi \vee \psi$ and $\neg\phi$. Finally, no symbols from $\mathcal{A}$ appear in $\top, \bot$. $\square$

We may also *define* things by induction. Recall:

Principle of inductive definition for $\mathbb{N}$. To define an object $f(n)$ depending on $n \in \mathbb{N}$:

- Define $f(0)$.
- Given $f(n)$, define $f(n + 1)$.

[3] Formally, the successor is defined before the number 1 (which is defined as the successor of 0). For the set-theoretically inclined: recall that in formal set theory, 0 is represented as $\varnothing$, 1 as $\{0\} = \{\varnothing\}$, 2 as $\{0, 1\} = \{\varnothing, \{\varnothing\}\}$, etc. Thus, the successor of n is $n \cup \{n\}$, and $\mathbb{N}$ is the smallest set such that $\varnothing \in \mathbb{N}$, and if $n \in \mathbb{N}$, then $n \cup \{n\} \in \mathbb{N}$.

Example 1.5. Factorials are defined inductively via

$$0! := 1,$$
$$(n+1)! := (n+1) \cdot n!.$$

Analogously,

Principle of inductive definition for propositional formulas. To define an object $f(\phi)$ depending on an $\mathcal{A}$-formula ϕ:

- Define $f(P)$ for each atomic formula $P \in \mathcal{A}$.
- Given $f(\phi), f(\psi)$, define $f(\phi \wedge \psi), f(\phi \vee \psi), f(\neg\phi)$.
- Define $f(\top), f(\bot)$.

The following makes precise the notion of "symbols appearing in ϕ" used in Proposition 1.4:

Example 1.6. For each $\mathcal{A}$-formula ϕ, we define the set $\mathrm{AT}(\phi)$ of atomic formulas appearing in ϕ inductively as follows:

$$\begin{aligned}
\mathrm{AT}(P) &:= \{P\} \quad \text{for } P \in \mathcal{A},\\
\mathrm{AT}(\phi \wedge \psi) &:= \mathrm{AT}(\phi) \cup \mathrm{AT}(\psi),\\
\mathrm{AT}(\phi \vee \psi) &:= \mathrm{AT}(\phi) \cup \mathrm{AT}(\psi),\\
\mathrm{AT}(\neg\phi) &:= \mathrm{AT}(\phi),\\
\mathrm{AT}(\top) &:= \mathrm{AT}(\bot) := \varnothing.
\end{aligned}$$

Example 1.7. For each $\mathcal{A}$-formula ϕ, we define the set $\mathrm{SF}(\phi)$ of **subformulas** of ϕ inductively:

$$\begin{aligned}
\mathrm{SF}(P) &:= \{P\} \quad \text{for } P \in \mathcal{A},\\
\mathrm{SF}(\phi \wedge \psi) &:= \mathrm{SF}(\phi) \cup \mathrm{SF}(\psi) \cup \{\phi \wedge \psi\},\\
\mathrm{SF}(\phi \vee \psi) &:= \mathrm{SF}(\phi) \cup \mathrm{SF}(\psi) \cup \{\phi \vee \psi\},\\
\mathrm{SF}(\neg\phi) &:= \mathrm{SF}(\phi) \cup \{\neg\phi\},\\
\mathrm{SF}(\top) &:= \{\top\},\\
\mathrm{SF}(\bot) &:= \{\bot\}.
\end{aligned}$$

For example, according to this definition, the subformulas of $P \wedge (\neg Q \vee R)$ are

$$\begin{aligned}
\mathrm{SF}(P \wedge (\neg Q \vee R)) &= \mathrm{SF}(P) \cup \mathrm{SF}(\neg Q \vee R) \cup \{P \wedge (\neg Q \vee R)\}\\
&= \{P\} \cup (\mathrm{SF}(\neg Q) \cup \mathrm{SF}(R) \cup \{\neg Q \vee R\}) \cup \{P \wedge (\neg Q \vee R)\}\\
&= \{P\} \cup ((\mathrm{SF}(Q) \cup \{\neg Q\}) \cup \{R\} \cup \{\neg Q \vee R\}) \cup \{P \wedge (\neg Q \vee R)\}\\
&= \{P\} \cup ((\{Q\} \cup \{\neg Q\}) \cup \{R\} \cup \{\neg Q \vee R\}) \cup \{P \wedge (\neg Q \vee R)\}\\
&= \{P, Q, \neg Q, R, \neg Q \vee R, P \wedge (\neg Q \vee R)\}.
\end{aligned}$$

In order to prove something about an inductively defined concept, we usually have to use induction as well:

Proposition 1.8. For any $\mathcal{A}$-formula ϕ, $\mathrm{AT}(\phi) \subseteq \mathrm{SF}(\phi)$.

Proof. By induction on ϕ. For an atomic formula $P \in \mathcal{A}$, we have $\mathrm{AT}(P) = \{P\} = \mathrm{SF}(P)$. If $\mathrm{AT}(\phi) \subseteq \mathrm{SF}(\phi)$ and $\mathrm{AT}(\psi) \subseteq \mathrm{SF}(\psi)$, then

$$\begin{aligned}
\mathrm{AT}(\phi \wedge \psi) &= \mathrm{AT}(\phi) \cup \mathrm{AT}(\psi)\\
&\subseteq \mathrm{SF}(\phi) \cup \mathrm{SF}(\psi) \quad \text{by the IH}\\
&\subseteq \mathrm{SF}(\phi) \cup \mathrm{SF}(\psi) \cup \{\phi \wedge \psi\} = \mathrm{SF}(\phi \wedge \psi),
\end{aligned}$$

and similarly for $\vee$ and $\neg$. Finally, $\mathrm{AT}(\top) = \varnothing \subseteq \{\top\} = \mathrm{SF}(\top)$, and similarly for $\bot$. $\square$

2. Propositional semantics

Let $\mathcal{A}$ be an alphabet. An **$\mathcal{A}$-truth assignment** (also known as **valuation**) is a function $m : \mathcal{A} \to \{0, 1\}$. We think of 1 as "true", 0 as "false". For $P \in \mathcal{A}$, we refer to $m(P)$ as the **interpretation of P according to m.**

Given a truth assignment $m : \mathcal{A} \to \{0, 1\}$, we extend it inductively to $m : \mathcal{L}(A) \to \{0, 1\}$:

$$\begin{aligned} m(\phi \wedge \psi) &:= \min(m(\phi), m(\psi)), \\ m(\phi \vee \psi) &:= \max(m(\phi), m(\psi)), \\ m(\neg\phi) &:= 1 - m(\phi), \\ m(\top) &:= 1, \\ m(\bot) &:= 0. \end{aligned}$$

As before, we call $m(\phi)$ the **interpretation of ϕ according to m.**

Some alternative notations are occasionally useful. We also write

$$\begin{aligned} \phi^m &:= m(\phi), \\ m \models \phi \;&:\Longleftrightarrow\; \phi^m = 1, \end{aligned}$$

pronounced "m **satisfies** ϕ". Note that ϕ^m is a number (0 or 1), a type of mathematical *object*, whereas "$m \models \phi$" is a *statement* or *assertion*, much like e.g., "$3 < \pi$". Using this notation, the above inductive definition of m becomes

$$\begin{aligned} m \models \phi \wedge \psi \;&\Longleftrightarrow\; m \models \phi \text{ and } m \models \psi, \\ m \models \phi \vee \psi \;&\Longleftrightarrow\; m \models \phi \text{ or } m \models \psi, \\ m \models \neg\phi \;&\Longleftrightarrow\; m \not\models \phi, \\ m \models \top &\quad \text{always}, \\ m \models \bot &\quad \text{never}. \end{aligned}$$

Note, again, the distinction between the formal syntactic expressions on the LHS and the "meta" and's and or's on the RHS: $\wedge$ operates on *objects* (formulas), while "and" combines *statements*.

Note also that according to our definitions of $\to$ and $\leftrightarrow$ as abbreviations (see Section 1.2),

$$\begin{aligned} m \models \phi \to \psi \;&\Longleftrightarrow\; m \models \neg\phi \vee \psi \\ &\Longleftrightarrow\; m \not\models \phi \text{ or } m \models \psi \\ &\Longleftrightarrow\; (m \models \phi \implies m \models \psi), \\ m \models \phi \leftrightarrow \psi \;&\Longleftrightarrow\; m \models \phi \to \psi \text{ and } m \models \psi \to \phi \\ &\Longleftrightarrow\; (m \models \phi \implies m \models \psi) \text{ and } (m \models \psi \implies m \models \phi). \end{aligned}$$

Let $\phi \in \mathcal{L}(\mathcal{A})$ be a formula. It is sometimes convenient to display the interpretation of ϕ according to *all* truth assignments (assuming there only finitely many), via a **truth table**.

Example 2.1. Here is a truth table for the formula $\phi = P \wedge \neg Q$ over the alphabet $\mathcal{A} = \{P, Q\}$:

P	Q	$P \wedge \neg Q$
0	0	0
0	1	0
1	0	1
1	1	0

There is one row of the table for each truth assignment $m : \mathcal{A} \to \{0, 1\}$, consisting of the entries $P^m, Q^m, (P \wedge \neg Q)^m$.

Example 2.2. Here is a truth table showing three $\mathcal{A} = \{P, Q, R\}$-formulas simultaneously:

P	Q	R	$(P \wedge Q) \vee R$	$P \wedge (Q \vee R)$	$P \wedge (Q \vee R) \to (P \wedge Q) \vee R$
0	0	0	0	0	1
0	0	1	1	0	1
0	1	0	0	0	1
0	1	1	1	0	1
1	0	0	0	0	1
1	0	1	1	1	1
1	1	0	1	1	1
1	1	1	1	1	1

We say that a formula ϕ is a **semantic tautology**, written

$$\models \phi,$$

if it is satisfied by all truth assignments; this means that its column in the truth table is all 1's.

If $\models \phi \to \psi$, i.e., every m satisfying ϕ also satisfies ψ, then we say ϕ **semantically implies** ψ. For example, Example 2.2 shows that $P \wedge (Q \vee R)$ semantically implies $(P \wedge Q) \vee R$.

If $\models \phi \leftrightarrow \psi$, i.e., ϕ, ψ semantically imply each other, then we call them **semantically equivalent**. For example, $P \wedge Q$ and $Q \wedge P$ are semantically equivalent, but not equal (see Remark 1.2).

Exercise 2.3 (Quiz 1)**.** Verify, using a truth table, that $(\phi \to \psi) \wedge (\psi \to \phi)$ and $(\phi \wedge \psi) \vee (\neg\phi \wedge \neg\psi)$ are semantically equivalent. (Recall that we took the former as the official definition of $\phi \leftrightarrow \psi$.)

Exercise 2.4 (de Morgan's law)**.** Verify that $\neg(\phi \wedge \psi)$ and $\neg\phi \vee \neg\psi$ are semantically equivalent.

Finally, we say that ϕ is **satisfiable** if it is satisfied by *some* truth assignment, and **unsatisfiable** otherwise. Thus ϕ is unsatisfiable iff $\neg\phi$ is a semantic tautology.

2.1. **Theories.** An **$\mathcal{A}$-theory** $\mathcal{T}$ is a set of $\mathcal{A}$-formulas. A formula $\phi \in \mathcal{T}$ is also called an **axiom** of $\mathcal{T}$. Informally, a theory is a set of statements which are assumed to be true. (Note that this does not quite correspond to some informal uses of the word "theory" in math and science.)

A truth assignment $m : \mathcal{A} \to \{0, 1\}$ is a **model** of $\mathcal{T}$, written

$$m \models \mathcal{T},$$

if it satisfies every axiom in $\mathcal{T}$, i.e.,

$$m \models \mathcal{T} \;:\Longleftrightarrow\; \forall \phi \in \mathcal{T}\,(m \models \phi).$$

We can think of a theory as a syntactic description of a set of "possible worlds"

Example 2.5. Every truth assignment is a model of $\mathcal{T} = \varnothing$ (vacuously).

Example 2.6. Let $\mathcal{A} = \{P, Q\}$, where we think of P, Q as representing the statements

$P =$ "it's raining", $\qquad Q =$ "the ground is wet",

and let $\mathcal{T} := \{P \to Q\}$. Then the truth assignment in the third row of the truth table

P	Q	$P \to Q$
0	0	1
0	1	1
1	0	0
1	1	1

(which represents that it's raining, but the ground is not wet) is not a model of $\mathcal{T}$.

Example 2.7. Suppose we have a sequence $x_0, x_1, x_2, \dots$ (of numbers, or of whatever other kind of mathematical objects), and we are interested in which terms x_i, x_j are the same. We can describe this situation via propositional logic as follows. Let

$$\mathcal{A} := \{P_{i,j} \mid i, j \in \mathbb{N}\} = \{P_{0,0}, P_{0,1}, P_{0,2}, \dots, P_{1,0}, P_{1,1}, \dots\},$$

where we think of $P_{i,j}$ as representing "$x_i = x_j$". Given an actual sequence $x_0, x_1, \dots$, we can define a truth assignment $m : \mathcal{A} \to \{0, 1\}$ by

$$(*) \qquad m \models P_{i,j} \;:\Longleftrightarrow\; x_i = x_j.$$

What constraints on m ensure that m actually describes equality of terms of a sequence in this way? That is, we would like to find an $\mathcal{A}$-theory $\mathcal{T}$ such that for all $m : \mathcal{A} \to \{0, 1\}$, $m \models \mathcal{T}$ iff there exists a sequence $x_0, x_1, \dots$ such that m is defined as in $(*)$. Here are some obvious constraints that such m must satisfy:

- Clearly $x_i = x_i$, so m must satisfy
$$P_{i,i}.$$
- If $m \models P_{i,j}$, i.e., $x_i = x_j$, then obviously $x_j = x_i$, i.e., $m \models P_{j,i}$. So m must satisfy
$$P_{i,j} \to P_{j,i}.$$
- Finally, we cannot have $m \models P_{i,j}$ and $m \models P_{j,k}$ but $m \not\models P_{i,k}$, since that would mean $x_i = x_j$ and $x_j = x_k$ but $x_i \neq x_k$. In other words, m must satisfy
$$P_{i,j} \wedge P_{j,k} \to P_{i,k}.$$

Let $\mathcal{T} \subseteq \mathcal{L}(A)$ be the (infinite) set of all of these axioms:

$$\mathcal{T} := \{P_{i,i} \mid i \in \mathbb{N}\} \cup \{P_{i,j} \to P_{j,i} \mid i, j \in \mathbb{N}\} \cup \{P_{i,j} \wedge P_{j,k} \to P_{i,k} \mid i, j, k \in \mathbb{N}\}.$$

We just argued above that *if* m comes from a sequence $x_0, x_1, \dots$ via $(*)$, then $m \models \mathcal{T}$.

Conversely, suppose $m \models \mathcal{T}$; we claim that we can somehow build a sequence $x_0, x_1, \dots$ such that $(*)$ holds. Define the following binary relation $\sim$ on $\mathbb{N}$:

$$i \sim j \;:\Longleftrightarrow\; m \models P_{i,j}.$$

Then the axioms in $\mathcal{T}$ ensure precisely that $\sim$ is an equivalence relation on $\mathbb{N}$:

- Reflexivity: since $m \models P_{i,i}$, we have $i \sim i$ for all $i \in \mathbb{N}$.
- Symmetry: if $i \sim j$, i.e., $m \models P_{i,j}$, then from $m \models P_{i,j} \to P_{j,i}$, we get $m \models P_{j,i}$, i.e., $j \sim i$.
- Transitivity: if $i \sim j \sim k$, i.e., $m \models P_{i,j}$ and $m \models P_{j,k}$, then from $m \models P_{i,j} \wedge P_{j,k} \to P_{i,k}$, we get $m \models P_{i,k}$, i.e., $i \sim k$,

Let $[i] := \{j \mid j \sim i\}$ be the equivalence class of i. Recall the key property of equivalence classes:

$$\begin{aligned} [i] = [j] &\iff i \sim j \\ &\iff m \models P_{i,j}. \end{aligned}$$

Thus if we let $x_i := [i]$, then $(*)$ will hold, as desired.

Example 2.8. Consider the following simple-sounding situation: we have a set, say $\mathbb{N}$, and we pick a single element $x \in \mathbb{N}$. We can attempt to describe this as follows. Let

$$\mathcal{A} := \{P_i \mid i \in \mathbb{N}\} = \{P_0, P_1, \dots\},$$

where we think of P_i as representing "$x = i$". A truth assignment $m : \mathcal{A} \to \{0, 1\}$ will thus be a correct description of such a choice of $x \in \mathbb{N}$ iff there is a *unique* i for which $m \models P_i$. We can axiomatize uniqueness via the axioms

$$\{\neg(P_i \wedge P_j) \mid i \neq j \in \mathbb{N}\},$$

but what about existence? We would like to say

$$P_0 \vee P_1 \vee P_2 \vee \cdots ;$$

but we can't, since formulas must be *finite* expressions. (Such a formula would work if we replaced $\mathbb{N}$ with all the numbers from 0 to 30, say.)

In fact, we will eventually be able to prove (see Exercise 4.20 below) that there is *no* theory whose models are precisely those $m : \mathcal{A} \to \{0, 1\}$ which satisfy a unique P_i. This is a first instance of one of the main themes of logic: the limited expressive power of mathematical language.

Recall the semantic truth notions for formulas defined shortly before the start of this subsection. We extend these notions to theories as follows.

We say that a formula ϕ is a **semantic consequence** of $\mathcal{T}$, or **semantically implied** by $\mathcal{T}$, written

$$\mathcal{T} \models \phi,$$

if it is satisfied by every model of $\mathcal{T}$. Note that when $\mathcal{T}$ is itself a single formula, we recover the previously defined notion of semantic implication:

$$\{\phi\} \models \psi \iff \phi \text{ semantically implies } \psi.$$

Recall also the notion of semantic tautology from before. Since every truth assignment is a model of the empty theory, a semantic tautology is the same thing as a semantic consequence of $\varnothing$:

$$\models \phi \iff \varnothing \models \phi.$$

We say that a theory $\mathcal{T}$ is **satisfiable** if it has a model, and **unsatisfiable** otherwise. A formula ϕ is satisfiable (as defined before) iff the singleton theory $\{\phi\}$ is satisfiable. Note that

$$\mathcal{T} \text{ unsatisfiable} \iff \mathcal{T} \models \bot;$$

indeed, no truth assignment can satisfy $\bot$, so $\mathcal{T} \models \bot$ means exactly that $\mathcal{T}$ has no models.

3. Deductive systems and proofs

We now define the syntactic notion of a *formal proof* of a formula from a theory. Formal proofs will be a type of syntactic expression, defined inductively just as formulas were, but with a much more complicated structure, intended to capture the structure of proofs that mathematicians write in practice. For example, here is an informal proof, which we've written out in a verbose fashion meant to hint at a formal "tree-like" structure; can you come up with an inductive definition which actually captures this structure?

Example 3.1. We claim that every perfect square leaves a remainder of 0 or 1 when divided by 4.

Proof. Let n be an integer; we must show n^2 leaves a remainder of 0 or 1 when divided by 1.

> By a known fact (the Euclidean division theorem), either n is even or n is odd.
>
> > First, suppose n is even, i.e., $n = 2k$ for some integer k.
> >
> > > Then $n^2 = 4k^2$ leaves a remainder of 0 when divided by 4.
> >
> > Now suppose n is odd, i.e., $n = 2k + 1$ for some integer k.
> >
> > > Then $n^2 = 4k^2 + 4k + 1$ leaves a remainder of 1 when divided by 4.
>
> So in both cases, n^2 leaves a remainder of 0 or 1 when divided by 4.

Since n was arbitrary, every perfect square leaves a remainder of 0 or 1 when divided by 4. $\square$

3.1. **Deductive systems.** Let $\mathcal{S}$ be a set (whose elements will usually be some kind of syntactic expressions, the "statements" we're deducing). A **deductive system over** $\mathcal{S}$ is a set of expressions of the form

$$(I)\,\frac{A_1 \quad A_2 \quad \cdots \quad A_n}{B}$$

where $A_1, \ldots, A_n, B \in \mathcal{S}$ (possibly with $n = 0$). Such an expression is called an **inference rule**, and is informally thought of as meaning

"from the hypotheses $A_1, \ldots, A_n$, we may deduce B".

The symbol I is a "label" which uniquely identifies the inference rule (it need not come from $\mathcal{S}$).

A **deduction** (over this deductive system) is an expression, constructed inductively as follows:

- Every $A \in \mathcal{S}$ is a deduction.
- If

  $$(I)\,\frac{A_1 \quad A_2 \quad \cdots \quad A_n}{B}$$

  is an inference rule, and

  $$\mathcal{D}_1 = \frac{\vdots}{A_1}, \qquad \mathcal{D}_2 = \frac{\vdots}{A_2}, \qquad \cdots, \qquad \mathcal{D}_n = \frac{\vdots}{A_n}$$

  are deductions ending in $A_1, \ldots, A_n$ respectively, then

  $$(I)\,\frac{\mathcal{D}_1 \quad \mathcal{D}_2 \quad \cdots \quad \mathcal{D}_n}{B}$$

  is a deduction.

We call the expression appearing at the bottom of a deduction its **conclusion**, and those appearing at the top (not below any inference rule) its **hypotheses**. Formally, these are defined inductively:

- For $A \in \mathcal{S}$, its conclusion is A and its set of hypotheses is $\{A\}$.
- For a deduction ending in an inference rule

  $$(I)\,\frac{\mathcal{D}_1 \quad \mathcal{D}_2 \quad \cdots \quad \mathcal{D}_n}{B}$$

  where $\mathcal{D}_1, \ldots, \mathcal{D}_n$ are deductions, the conclusion is B, and the set of hypotheses is the union of the hypotheses of $\mathcal{D}_1, \ldots, \mathcal{D}_n$.

We also say that a deduction is **of** its conclusion **from** its hypotheses.

Example 3.2. Let $\mathcal{S} = \{P, Q, R\}$, and consider the following inference rules:

$$(\mathrm{I})\,\frac{}{P} \qquad\qquad (\mathrm{J})\,\frac{P}{Q} \qquad\qquad (\mathrm{K})\,\frac{P \quad Q}{R}$$

Here is a deduction of R from P:

$$(\mathrm{K})\,\frac{P \qquad (\mathrm{J})\,\dfrac{P}{Q}}{R}$$

If we add a rule (I) above each P in this deduction, we get a deduction of R from no hypotheses.

As with any kind of inductively defined expression, we can formally represent deductions as nested tuples, as in Section 1.1, although we rarely bother to do so. For example, the above deduction may be represented as

$$(\mathrm{K}, P, (\mathrm{J}, P, Q), R),$$

while the one after adding rule (I) is

$$(\mathrm{K}, (\mathrm{I}, P), (\mathrm{J}, (\mathrm{I}, P), Q), R).$$

Example 3.3. Here is a rather meaningless example of a deductive system. The set $\mathcal{S}$ is $\mathbb{N}$, and the inference rules are

$$(\textsc{Add}_{m,n})\ \frac{m \qquad n}{m+n} \qquad\qquad (\textsc{Mult}_{m,n})\ \frac{m \qquad n}{mn}$$

for all m, n. In other words, each of these expressions actually denotes an infinite family of inference rules, rather than a single rule; we call them **rule schemas**. Here is a deduction:

$$(\textsc{Mult})\ \cfrac{(\textsc{Add})\ \cfrac{3 \qquad 4}{7} \qquad (\textsc{Add})\ \cfrac{(\textsc{Mult})\ \cfrac{2 \qquad 2}{4} \qquad 1}{5}}{35}$$

When it is clear which rule is being used, we abbreviate or omit altogether the label.

Since deductions are inductively defined, we may prove things about them by induction. Here is a (silly) example:

Proposition 3.4. In the above deductive system, if there is a deduction $\mathcal{D}$ of $n \in \mathbb{N}$ from even numbers, then n is even.

Proof. By induction on $\mathcal{D}$. If $\mathcal{D}$ is a single hypothesis n, then n is even by assumption. If $\mathcal{D}$ ends in either of

$$(\textsc{Add}_{m,n})\ \frac{m \qquad n}{m+n}, \qquad\qquad (\textsc{Mult}_{m,n})\ \frac{m \qquad n}{mn},$$

then by the induction hypothesis applied to the deductions of m, n, both are even, hence $m + n, mn$ are also even. $\square$

Example 3.5. Let $\mathcal{A} = \{P, Q, R, \dots\}$ be an alphabet. We define a deductive system over $\mathcal{L}(\mathcal{A})$ with the following rule schemas:

$$\frac{\phi \qquad \psi}{\phi \wedge \psi} \qquad \frac{\phi \wedge \psi}{\phi} \qquad \frac{\phi \wedge \psi}{\psi} \qquad \frac{\phi}{\phi \vee \psi} \qquad \frac{\psi}{\phi \vee \psi}$$

for arbitrary formulas $\phi, \psi \in \mathcal{L}(\mathcal{A})$. Here is a deduction:

$$\cfrac{\cfrac{P \qquad Q}{P \wedge Q}}{(P \wedge Q) \vee R}$$

This last example is beginning to resemble a formal proof system for propositional logic. However, note that the system is far from complete: there are many semantically true statements which it cannot prove. For example, note that we do not have a rule for how to *use* a disjunction to prove something else. In informal proofs, such as Example 3.1, in order to use a disjunction, we have to do case analysis; there is no way to capture this in the simple inductive structure of deductions over formulas as in Example 3.5, since in the subproofs for each case, we have to temporarily modify not only the conclusion we're proving, but also the assumptions we're working from (e.g., that n is even in Example 3.1). Thus, in order to incorporate proof structures like case analysis, we will consider deductions over more complicated expressions which record a formula *as well as the background assumptions at a given point in the proof.*

3.2. **Natural deduction.** Let $\mathcal{A}$ be an alphabet. An **$\mathcal{A}$-sequent** is an expression of the form

$$\mathcal{T} \vdash \phi,$$

read "$\mathcal{T}$ proves ϕ", where $\mathcal{T}$ is an $\mathcal{A}$-theory (note: may be infinite) and ϕ is an $\mathcal{A}$-formula. Informally, this denotes the formula ϕ under the background assumptions $\mathcal{T}$.

We now define a deductive system over the set of $\mathcal{A}$-sequents (rather than $\mathcal{A}$-formulas), called the **natural deduction system for propositional logic**.[4] All of the inference rules in the system are in fact rule schemas, although we just refer to them as rules, for short; also, we will omit subscripts (for $\mathcal{T}, \phi$) on the labels.

- We have the **assumption rule (schema)**

$$(\mathrm{A})\ \frac{}{\mathcal{T} \vdash \phi} \quad \text{whenever } \phi \in \mathcal{T}.$$

Next, for each logical connective, we have an **introduction rule(s)**, which allows us to prove a formula containing the connective, and an **elimination rule(s)**, which allows us to use a formula containing the connective to prove something else.

- For $\wedge$, there is one introduction rule ("to prove $\phi \wedge \psi$, prove ϕ and then prove ψ") and two elimination rules ("from $\phi \wedge \psi$, we may deduce ϕ, as well as deduce ψ"):

$$(\wedge\mathrm{I})\ \frac{\mathcal{T} \vdash \phi \qquad \mathcal{T} \vdash \psi}{\mathcal{T} \vdash \phi \wedge \psi} \qquad (\wedge\mathrm{E1})\ \frac{\mathcal{T} \vdash \phi \wedge \psi}{\mathcal{T} \vdash \phi} \qquad (\wedge\mathrm{E2})\ \frac{\mathcal{T} \vdash \phi \wedge \psi}{\mathcal{T} \vdash \psi}$$

- For $\vee$, there are two introduction rules and one elimination rule (which captures the informal proof structure of case analysis):

$$(\vee\mathrm{I1})\ \frac{\mathcal{T} \vdash \phi}{\mathcal{T} \vdash \phi \vee \psi} \quad (\vee\mathrm{I2})\ \frac{\mathcal{T} \vdash \psi}{\mathcal{T} \vdash \phi \vee \psi} \quad (\vee\mathrm{E})\ \frac{\mathcal{T} \vdash \phi \vee \psi \qquad \mathcal{T} \cup \{\phi\} \vdash \theta \qquad \mathcal{T} \cup \{\psi\} \vdash \theta}{\mathcal{T} \vdash \theta}$$

- For $\top, \bot$, we have the "0-ary" versions of the rules for $\wedge, \vee$:

$$(\top\mathrm{I})\ \frac{}{\mathcal{T} \vdash \top} \qquad\qquad (\bot\mathrm{E})\ \frac{\mathcal{T} \vdash \bot}{\mathcal{T} \vdash \theta}$$

  Note that there is no elimination rule for $\top$ (knowing $\top$ gives no information), and no introduction rule for $\bot$ (although see ($\neg$E) below).
- For $\neg$, we have ("to prove $\neg\phi$, assume ϕ and derive a contradiction", and "from ϕ and $\neg\phi$, we get a contradiction"):

$$(\neg\mathrm{I})\ \frac{\mathcal{T} \cup \{\phi\} \vdash \bot}{\mathcal{T} \vdash \neg\phi} \qquad\qquad (\neg\mathrm{E})\ \frac{\mathcal{T} \vdash \phi \qquad \mathcal{T} \vdash \neg\phi}{\mathcal{T} \vdash \bot}$$

[4]There are three commonly used styles of proof system: natural deduction systems try to capture as closely as possible the structure of informal proofs that people write in practice, e.g., Example 3.1. **Hilbert systems** are deductive systems over formulas which encode all background assumptions as implications ($\to$); they are thus simpler to define than natural deduction systems, but much harder to use in practice, involving bizzare-looking formulas like $(\phi \to (\psi \to \theta)) \to ((\phi \to \psi) \to (\phi \to \theta))$. **Gentzen sequent calculi** are like natural deduction systems that have been turned "inside out", progressing from simple to increasingly complicated formulas, rather than "forward" towards the conclusion; they are no simpler to define than natural deduction, nor so easy to use in practice, but more convenient for analyzing the structure of proofs themselves. We will not consider these other types of proof systems in this course.

Finally, we have a rule which is neither an introduction nor an elimination rule.

- The **contradiction rule** ("to prove ϕ, assume $\neg\phi$ and derive a contradiction"):

$$\text{(C)}\ \frac{\mathcal{T} \cup \{\neg\phi\} \vdash \bot}{\mathcal{T} \vdash \phi}$$

By abuse of notation, we write

$$\mathcal{T} \vdash \phi$$

as a *statement*, and say $\mathcal{T}$ **proves** ϕ, if there is a deduction of the *sequent* $\mathcal{T} \vdash \phi$ from no hypotheses; this is an abuse of notation, because it means $\mathcal{T} \vdash \phi$ can denote either an *object* (the sequent) or a *statement*. [5] We also say in this case that ϕ is a **provable consequence** of $\mathcal{T}$. We write

$$\vdash \phi \ :\Longleftrightarrow\ \varnothing \vdash \phi$$

and say that ϕ is a **provable tautology** in this case.

Example 3.6. In Example 2.2, we showed that $\{P \wedge (Q \vee R)\} \models (P \wedge Q) \vee R$. We now show that the corresponding *provable* implication also holds: $\{P \wedge (Q \vee R)\} \vdash (P \wedge Q) \vee R$. Informally, this means that assuming $P \wedge (Q \vee R)$, we can prove $(P \wedge Q) \vee R$. Here is an informal proof:

(∧E2) Since $P \wedge (Q \vee R)$, we know $Q \vee R$.
(∨E) So there are two cases: either Q, or R.
Case 1: Q holds.
(∧E1) Then since $P \wedge (Q \vee R)$, we know P.
(∧I) So $P \wedge Q$ holds.
(∨I1) So $(P \wedge Q) \vee R$ holds.
Case 2: R holds.
(∨I2) Then $(P \wedge Q) \vee R$ holds.
Thus in both cases, $(P \wedge Q) \vee R$ holds.

Note how carefully we wrote out each step of the proof, even trivial steps that we normally wouldn't bother to write out, like going from "P and ..." to P. Here is the formal deduction corresponding to the above informal proof (we've labelled above where each inference rule except for (A) is used):

$$\text{(∨E)}\ \frac{\text{(∧E1)}\ \dfrac{\text{(A)}\ \dfrac{}{\{P \wedge (Q \vee R)\} \vdash P \wedge (Q \vee R)}}{\{P \wedge (Q \vee R)\} \vdash Q \vee R} \quad \text{(∨I1)}\ \dfrac{\text{(∧I)}\ \dfrac{\text{(∧E1)}\ \dfrac{\text{(A)}\ \dfrac{}{\{P \wedge (Q \vee R), Q\} \vdash P \wedge (Q \vee R)}}{\{P \wedge (Q \vee R), Q\} \vdash P} \quad \text{(A)}\ \dfrac{}{\{P \wedge (Q \vee R), Q\} \vdash Q}}{\{P \wedge (Q \vee R), Q\} \vdash P \wedge Q}}{\{P \wedge (Q \vee R), Q\} \vdash (P \wedge Q) \vee R} \quad \text{(∨I2)}\ \dfrac{\text{(A)}\ \dfrac{}{\{P \wedge (Q \vee R), R\} \vdash R}}{\{P \wedge (Q \vee R), R\} \vdash (P \wedge Q) \vee R}}{\{P \wedge (Q \vee R)\} \vdash (P \wedge Q) \vee R}$$

Example 3.7. For any formulas ϕ, ψ, we have $\{\phi \wedge \psi\} \vdash \psi \wedge \phi$:

$$\text{(∧I)}\ \frac{\text{(∧E2)}\ \dfrac{\text{(A)}\ \dfrac{}{\{\phi \wedge \psi\} \vdash \phi \wedge \psi}}{\{\phi \wedge \psi\} \vdash \psi} \quad \text{(∧E1)}\ \dfrac{\text{(A)}\ \dfrac{}{\{\phi \wedge \psi\} \vdash \phi \wedge \psi}}{\{\phi \wedge \psi\} \vdash \phi}}{\{\phi \wedge \psi\} \vdash \psi \wedge \phi}$$

Since ϕ, ψ are arbitrary, we may swap them to get $\{\psi \wedge \phi\} \vdash \phi \wedge \psi$. Thus, these two formulas are *provably* equivalent (they are obviously also semantically equivalent).

[5] This is somewhat analogous to how $A \cong B$ can denote a bijection (or isomorphism) between two sets, as well as the statement that such a bijection exists.

Example 3.8. For any formula ϕ, we have $\{\phi\} \vdash \neg\neg\phi$:

$$\cfrac{(\neg\mathrm{E})\ \cfrac{(\mathrm{A})\ \cfrac{}{\{\phi, \neg\phi\} \vdash \phi} \qquad (\mathrm{A})\ \cfrac{}{\{\phi, \neg\phi\} \vdash \neg\phi}}{(\neg\mathrm{I})\ \{\phi, \neg\phi\} \vdash \bot}}{\{\phi\} \vdash \neg\neg\phi}$$

as well as $\{\neg\neg\phi\} \vdash \phi$:

$$\cfrac{(\neg\mathrm{E})\ \cfrac{(\mathrm{A})\ \cfrac{}{\{\neg\neg\phi, \neg\phi\} \vdash \neg\neg\phi} \qquad (\mathrm{A})\ \cfrac{}{\{\neg\neg\phi, \neg\phi\} \vdash \neg\phi}}{(\mathrm{C})\ \{\neg\neg\phi, \neg\phi\} \vdash \bot}}{\{\neg\neg\phi\} \vdash \phi}$$

So $\phi, \neg\neg\phi$ are provably equivalent.

Remark 3.9. Recall from Section 2 that semantic implication can be defined as either $\{\phi\} \models \psi$, or $\models \phi \to \psi$. It turns out that the corresponding notions of syntactic implication, $\{\phi\} \vdash \psi$ and $\vdash \phi \to \psi$, are also equivalent, although this is not obvious (see Example 3.18). Nonetheless, these are two distinct notions, representing different aspects of informal mathematical language:

- $\phi \to \psi$ represents the explicit statement "if ϕ, then ψ";
- $\{\phi\} \vdash \psi$ represents the statement ψ asserted in a context with background assumption ϕ.

For example, a chapter in a calculus textbook may begin with the sentence "We assume throughout that $f : \mathbb{R} \to \mathbb{R}$ is a continuous function on the real number line." Then later in the chapter, one may see the following

Theorem. If f is bounded and increasing, then $\lim_{x\to\infty} f(x)$ exists.

This would be represented by the sequent

$$\{\text{“}f : \mathbb{R} \to \mathbb{R} \text{ is cts”}, \dots\} \vdash \text{“}f \text{ bdd \& incr.”} \to \text{“}\lim_{x\to\infty} f(x) \text{ exists”}.$$

(Of course, propositional logic isn't nearly expressive enough to formalize any of the quoted statements here; at least first-order logic would be needed. The theory on the LHS would also contain many other background assumptions that have been made earlier, e.g., basic properties of $\mathbb{R}$.)

3.3. **Derivable and admissible rules.** It is often convenient to use inference rules which are not in the basic list above, but which can be built out of those. We say that an inference rule (schema)

$$\frac{S_1 \quad S_2 \quad \cdots \quad S_n}{T}$$

(where S_i, T are sequents) is:

- **derivable** if there is a deduction with hypotheses $S_1, \dots, S_n$ and conclusion T;
- **admissible** if whenever there are deductions of $S_1, \dots, S_n$ from *no* hypotheses, then there is a deduction of T from *no* hypotheses.

Thus derivable rules are admissible, but not vice-versa: admissibility means there is *some* way to turning deductions of $S_1, \dots, S_n$ into a deduction of T, while derivability means this can be done simply by gluing a fixed deduction below those of $S_1, \dots, S_n$.

Example 3.10. To say that $\mathcal{T} \vdash \phi$ is the same as saying that

$$\frac{}{\mathcal{T} \vdash \phi}$$

is derivable (or admissible).

Example 3.11. Recalling our convention that $\phi \wedge \psi \wedge \theta := (\phi \wedge \psi) \wedge \theta$ (Section 1.2), the following ternary version of ($\wedge$I)

$$\frac{\mathcal{T} \vdash \phi \qquad \mathcal{T} \vdash \psi \qquad \mathcal{T} \vdash \theta}{\mathcal{T} \vdash \phi \wedge \psi \wedge \theta}$$

is derivable:

$$(\wedge\mathrm{I})\,\frac{(\wedge\mathrm{I})\,\dfrac{\mathcal{T} \vdash \phi \qquad \mathcal{T} \vdash \psi}{\mathcal{T} \vdash \phi \wedge \psi} \qquad \mathcal{T} \vdash \theta}{\mathcal{T} \vdash (\phi \wedge \psi) \wedge \theta}$$

Proposition 3.12 (weakening)**.** The following rule is admissible, for theories $\mathcal{T} \subseteq \mathcal{T}'$:

$$(\mathrm{W})\,\frac{\mathcal{T} \vdash \phi}{\mathcal{T}' \vdash \phi}$$

Informally, this means that if you can prove ϕ under assumptions $\mathcal{T}$, then you should be able to prove it under even more assumptions.

Proof. Assume there is a deduction $\mathcal{D}$ of $\mathcal{T} \vdash \phi$ (from no hypotheses); we must show that there is a deduction $\mathcal{D}'$ of $\mathcal{T}' \vdash \phi$ (from no hypotheses). We use induction on $\mathcal{D}$.

- If $\mathcal{D} = (\mathrm{A})\,\frac{}{\mathcal{T} \vdash \phi}$, then $\phi \in \mathcal{T} \subseteq \mathcal{T}'$, so $\mathcal{D}' := (\mathrm{A})\,\frac{}{\mathcal{T}' \vdash \phi}$ works.
- If $\mathcal{D}$ ends with

$$\mathcal{D} = (\wedge\mathrm{I})\,\frac{\mathcal{D}_1 = \dfrac{\vdots}{\mathcal{T} \vdash \phi} \qquad \mathcal{D}_2 = \dfrac{\vdots}{\mathcal{T} \vdash \psi}}{\mathcal{T} \vdash \phi \wedge \psi},$$

then by the induction hypothesis applied to the sub-deductions $\mathcal{D}_1, \mathcal{D}_2$, we have deductions of $\mathcal{T}' \vdash \phi$ and $\mathcal{T}' \vdash \psi$, whence applying ($\wedge$I), we get a deduction of $\mathcal{T}' \vdash \phi \wedge \psi$.
- If $\mathcal{D}$ ends with

$$\mathcal{D} = (\vee\mathrm{E})\,\frac{\mathcal{D}_1 = \dfrac{\vdots}{\mathcal{T} \vdash \phi \vee \psi} \qquad \mathcal{D}_2 = \dfrac{\vdots}{\mathcal{T} \cup \{\phi\} \vdash \theta} \qquad \mathcal{D}_3 = \dfrac{\vdots}{\mathcal{T} \cup \{\psi\} \vdash \theta}}{\mathcal{T} \vdash \theta},$$

then by the induction hypothesis applied to $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$, we have deductions of

$$\mathcal{T}' \vdash \phi \vee \psi, \qquad \mathcal{T}' \cup \{\phi\} \vdash \theta, \qquad \mathcal{T}' \cup \{\psi\} \vdash \theta,$$

whence applying ($\vee$E), we get a deduction of $\mathcal{T}' \vdash \theta$.

The rest of the cases are similarly straightforward. □

Remark 3.13. (W) is *not* derivable. To see this, note that in each of the basic inference rules, every sequent that appears in the hypothesis has a theory on the LHS which contains the theory $\mathcal{T}$ in the conclusion. Thus, if $\mathcal{T} \subsetneq \mathcal{T}'$ in (W), the only way it could possibly be derivable is via a deduction which does not use the hypothesis $\mathcal{T} \vdash \phi$ at all, i.e., if in fact $\mathcal{T}' \vdash \phi$. But of course it is possible that $\mathcal{T}' \not\vdash \phi$, e.g., $\mathcal{T} := \varnothing$, $\mathcal{T}' := \{Q\}$, and $\phi := P$ (by soundness, Proposition 3.22).

Once we have shown that a rule is derivable or admissible, we may use it in deductions to show that other rules are derivable or admissible.

Example 3.14. There was no need to include the ($\bot$E) rule in our deductive system, since it is admissible given the other rules, using the admissible (W) rule:[6]

$$\dfrac{\dfrac{\mathcal{T} \vdash \bot}{\mathcal{T} \cup \{\neg\phi\} \vdash \bot}\,(\mathrm{W})}{\mathcal{T} \vdash \phi}\,(\mathrm{C})$$

Example 3.15. Analogously to Example 3.11, the following ternary version of ($\vee$E)

$$\dfrac{\mathcal{T} \vdash \phi \vee \psi \vee \theta \qquad \mathcal{T} \cup \{\phi\} \vdash \rho \qquad \mathcal{T} \cup \{\psi\} \vdash \rho \qquad \mathcal{T} \cup \{\theta\} \vdash \rho}{\mathcal{T} \vdash \rho}$$

is admissible:

$$\dfrac{\mathcal{T} \vdash (\phi \vee \psi) \vee \theta \qquad \dfrac{\dfrac{}{\mathcal{T} \cup \{\phi \vee \psi\} \vdash \phi \vee \psi}\,(\mathrm{A}) \qquad \dfrac{\mathcal{T} \cup \{\phi\} \vdash \rho}{\mathcal{T} \cup \{\phi \vee \psi\} \cup \{\phi\} \vdash \rho}\,(\mathrm{W}) \qquad \dfrac{\mathcal{T} \cup \{\psi\} \vdash \rho}{\mathcal{T} \cup \{\phi \vee \psi\} \cup \{\psi\} \vdash \rho}\,(\mathrm{W})}{\mathcal{T} \cup \{\phi \vee \psi\} \vdash \rho}\,(\vee\mathrm{E}) \qquad \mathcal{T} \cup \{\theta\} \vdash \rho}{\mathcal{T} \vdash \rho}\,(\vee\mathrm{E})$$

Note that the use of the admissible (W) rule makes this admissible rather than derivable.

Exercise 3.16 (HW3)**.** The following rule (called the **law of excluded middle**) is derivable:

$$\dfrac{}{\mathcal{T} \vdash \phi \vee \neg\phi}\,(\mathrm{LEM})$$

This rule is often immediately followed by ($\vee$E), which allows us to do casework depending on whether an arbitrary formula is true or false.

Example 3.17. Recall our abbreviation $\phi \to \psi := \neg\phi \vee \psi$. The following admissible introduction and elimination rules for $\to$ capture the informal proof structures "to prove $\phi \to \psi$, assume ϕ and then prove ψ" and "to use $\phi \to \psi$, prove ϕ and conclude ψ":

$$(\to\mathrm{I})\,\dfrac{\mathcal{T} \cup \{\phi\} \vdash \psi}{\mathcal{T} \vdash \phi \to \psi} \qquad\qquad (\to\mathrm{E})\,\dfrac{\mathcal{T} \vdash \phi \to \psi \qquad \mathcal{T} \vdash \phi}{\mathcal{T} \vdash \psi}$$

The derivation of ($\to$I) uses (LEM) from Exercise 3.16:

$$\dfrac{\dfrac{}{\mathcal{T} \vdash \phi \vee \neg\phi}\,(\mathrm{LEM}) \qquad \dfrac{\mathcal{T} \cup \{\phi\} \vdash \psi}{\mathcal{T} \cup \{\phi\} \vdash \neg\phi \vee \psi}\,(\vee\mathrm{I2}) \qquad \dfrac{\dfrac{}{\mathcal{T} \cup \{\neg\phi\} \vdash \neg\phi}\,(\mathrm{A})}{\mathcal{T} \cup \{\neg\phi\} \vdash \neg\phi \vee \psi}\,(\vee\mathrm{I1})}{\mathcal{T} \vdash \underbrace{\neg\phi \vee \psi}_{\phi \to \psi}}\,(\vee\mathrm{E})$$

The derivation of ($\to$E) again uses weakening:

$$\dfrac{\mathcal{T} \vdash \overbrace{\neg\phi \vee \psi}^{\phi \to \psi} \qquad \dfrac{\dfrac{\dfrac{\mathcal{T} \vdash \phi}{\mathcal{T} \cup \{\neg\phi\} \vdash \phi}\,(\mathrm{W}) \qquad \dfrac{}{\mathcal{T} \cup \{\neg\phi\} \vdash \neg\phi}\,(\mathrm{A})}{\mathcal{T} \cup \{\neg\phi\} \vdash \bot}\,(\neg\mathrm{E})}{\mathcal{T} \cup \{\neg\phi\} \vdash \psi}\,(\bot\mathrm{E}) \qquad \dfrac{}{\mathcal{T} \cup \{\psi\} \vdash \psi}\,(\mathrm{A})}{\mathcal{T} \vdash \psi}\,(\vee\mathrm{E})$$

(So ($\to$I) is derivable, while ($\to$E) is merely admissible.)

[6]We are using here that the above proof of Proposition 3.12 continues to work when we remove the ($\bot$E) rule, since the inductive case for each rule only appeals to that same rule.

Example 3.18. Recalling Examples 3.7 and 3.8, we have

$$\dfrac{(\wedge\mathrm{I})\ \dfrac{(\to\mathrm{I})\ \dfrac{\dfrac{\vdots}{\{\phi\wedge\psi\}\vdash\psi\wedge\phi}}{\vdash\phi\wedge\psi\to\psi\wedge\phi}\qquad(\to\mathrm{I})\ \dfrac{\dfrac{\vdots}{\{\psi\wedge\phi\}\vdash\phi\wedge\psi}}{\vdash\psi\wedge\phi\to\phi\wedge\psi}}{\vdash\phi\wedge\psi\leftrightarrow\psi\wedge\phi}}{}\!,$$

i.e., $\phi\wedge\psi\leftrightarrow\psi\wedge\phi$ is a provable tautology. Similarly, $\phi\leftrightarrow\neg\neg\phi$ is a provable tautology.

Example 3.19. The formula $(\phi\to(\psi\to\theta))\to((\phi\to\psi)\to(\phi\to\theta))$ from Footnote 4 is a provable tautology:

$$(\to\mathrm{I})\ \dfrac{(\to\mathrm{I})\ \dfrac{(\to\mathrm{I})\ \dfrac{(\to\mathrm{E})\ \dfrac{(\to\mathrm{E})\ \dfrac{(\mathrm{A})\ \dfrac{}{\mathcal{T}\vdash\phi\to(\psi\to\theta)}\quad(\mathrm{A})\ \dfrac{}{\mathcal{T}\vdash\phi}}{\mathcal{T}\vdash\psi\to\theta}\qquad(\to\mathrm{E})\ \dfrac{(\mathrm{A})\ \dfrac{}{\mathcal{T}\vdash\phi\to\psi}\quad(\mathrm{A})\ \dfrac{}{\mathcal{T}\vdash\phi}}{\mathcal{T}\vdash\psi}}{\mathcal{T}:=\{\phi\to(\psi\to\theta),\phi\to\psi,\phi\}\vdash\theta}}{\{\phi\to(\psi\to\theta),\phi\to\psi\}\vdash\phi\to\theta}}{\{\phi\to(\psi\to\theta)\}\vdash(\phi\to\psi)\to(\phi\to\theta)}}{\vdash(\phi\to(\psi\to\theta))\to((\phi\to\psi)\to(\phi\to\theta))}$$

Exercise 3.20 (HW3)**.** The following **cut rule** is admissible:

$$(\textsc{Cut})\ \dfrac{\mathcal{T}\vdash\phi\qquad\mathcal{T}\cup\{\phi\}\vdash\psi}{\mathcal{T}\vdash\psi}$$

More generally, for two theories $\mathcal{T},\mathcal{T}'$, if $\mathcal{T}$ proves every formula in $\mathcal{T}'$, and $\mathcal{T}\cup\mathcal{T}'\vdash\psi$, then $\mathcal{T}\vdash\psi$ (the cut rule is the case $\mathcal{T}'=\{\phi\}$).

Of course, the converse of weakening is false: we may not freely remove assumptions from proofs. However, an important part of the converse is true: we may remove all but finitely many of the assumptions, essentially because proofs are finite expressions.

Proposition 3.21 (syntactic compactness)**.** If $\mathcal{T}\vdash\phi$, then there is a finite $\mathcal{T}'\subseteq\mathcal{T}$ such that $\mathcal{T}'\vdash\phi$.

Proof. By induction on the deduction of $\mathcal{T}\vdash\phi$.

- If the deduction ends with (A), then $\mathcal{T}':=\{\phi\}$ works.
- If the deduction ends with

  $$(\vee\mathrm{E})\ \dfrac{\mathcal{T}\vdash\phi\vee\psi\qquad\mathcal{T}\cup\{\phi\}\vdash\theta\qquad\mathcal{T}\cup\{\psi\}\vdash\theta}{\mathcal{T}\vdash\theta},$$

  then by the induction hypothesis (applied to the sub-deductions of $\mathcal{T}\vdash\phi\vee\psi$, $\mathcal{T}\cup\{\phi\}\vdash\theta$, and $\mathcal{T}\cup\{\psi\}\vdash\theta$), there are finite $\mathcal{T}_1\subseteq\mathcal{T}$, $\mathcal{T}_2\subseteq\mathcal{T}\cup\{\phi\}$, and $\mathcal{T}_3\subseteq\mathcal{T}\cup\{\psi\}$ such that

  $$\mathcal{T}_1\vdash\phi\vee\psi,\qquad\mathcal{T}_2\vdash\theta,\qquad\mathcal{T}_3\vdash\theta.$$

  Let $\mathcal{T}':=\mathcal{T}_1\cup(\mathcal{T}_2\cap\mathcal{T})\cup(\mathcal{T}_3\cap\mathcal{T})$. Then $\mathcal{T}_1\subseteq\mathcal{T}'$, $\mathcal{T}_2\subseteq\mathcal{T}'\cup\{\phi\}$ (because $\mathcal{T}_2\subseteq\mathcal{T}\cup\{\phi\}$), and $\mathcal{T}_3\subseteq\mathcal{T}'\cup\{\phi\}$, so by (W),

  $$\mathcal{T}'\vdash\phi\vee\psi,\qquad\mathcal{T}'\cup\{\phi\}\vdash\theta,\qquad\mathcal{T}'\cup\{\psi\}\vdash\theta.$$

  So by $(\vee\mathrm{E})$, $\mathcal{T}'\vdash\theta$. (Note that $\mathcal{T}':=\mathcal{T}_1\cup\mathcal{T}_2\cup\mathcal{T}_3$ would not satisfy $\mathcal{T}'\subseteq\mathcal{T}$.)

The rest of the cases are similar. ∎

3.4. **Soundness.**

Proposition 3.22 (soundness)**.** If $\mathcal{T} \vdash \phi$, then $\mathcal{T} \models \phi$.

Proof. We assume that there is a deduction $\mathcal{D}$ of $\mathcal{T} \vdash \phi$, and we must show that for every model $m \models \mathcal{T}$, we have $m \models \phi$. We use induction on $\mathcal{D}$.

- If $\mathcal{D}$ ends with (A), then $\phi \in \mathcal{T}$, so since $m \models \mathcal{T}$, we have $m \models \phi$.
- If $\mathcal{D}$ ends with
$$(\wedge\mathrm{I})\ \frac{\mathcal{T} \vdash \phi \qquad \mathcal{T} \vdash \psi}{\mathcal{T} \vdash \phi \wedge \psi},$$
then by the induction hypothesis applied to the sub-deductions of $\mathcal{T} \vdash \phi$ and $\mathcal{T} \vdash \psi$, we have $m \models \phi$ and $m \models \psi$, whence $m \models \phi \wedge \psi$.
- If $\mathcal{D}$ ends with
$$(\wedge\mathrm{E1})\ \frac{\mathcal{T} \vdash \phi \wedge \psi}{\mathcal{T} \vdash \phi},$$
then by the IH, we have $m \models \phi \wedge \psi$, whence $m \models \phi$. Similarly if $\mathcal{D}$ ends with ($\wedge$E2).
- If $\mathcal{D}$ ends with
$$(\vee\mathrm{I1})\ \frac{\mathcal{T} \vdash \phi}{\mathcal{T} \vdash \phi \vee \psi},$$
then by the IH, we have $m \models \phi$, whence $m \models \phi \vee \psi$. Similarly if $\mathcal{D}$ ends with ($\vee$I2).
- If $\mathcal{D}$ ends with
$$(\vee\mathrm{E})\ \frac{\mathcal{T} \vdash \phi \vee \psi \qquad \mathcal{T} \cup \{\phi\} \vdash \theta \qquad \mathcal{T} \cup \{\psi\} \vdash \theta}{\mathcal{T} \vdash \theta},$$
then by the IH applied to $\mathcal{T} \vdash \phi \vee \psi$, we have $m \models \phi \vee \psi$, i.e., either $m \models \phi$ or $m \models \psi$. In the former case, since $m \models \mathcal{T}$ and $m \models \phi$, we have $m \models \mathcal{T} \cup \{\phi\}$, whence by the IH applied to $\mathcal{T} \cup \{\phi\} \vdash \theta$, we get $m \models \theta$. The latter case is similar.
- If $\mathcal{D}$ ends with ($\top$I), then $\phi = \top$, which m always satisfies.
- If $\mathcal{D}$ ends with
$$(\bot\mathrm{E})\ \frac{\mathcal{T} \vdash \bot}{\mathcal{T} \vdash \theta},$$
then by the IH, every $m \models \mathcal{T}$ satisfies $\bot$, which is impossible; thus there are no models of $\mathcal{T}$, and so every $m \models \mathcal{T}$ (vacuously) satisfies θ.
- If $\mathcal{D}$ ends with
$$(\neg\mathrm{I})\ \frac{\mathcal{T} \cup \{\phi\} \vdash \bot}{\mathcal{T} \vdash \neg\phi},$$
then we cannot have $m \models \phi$, or else together with $m \models \mathcal{T}$ we would get $m \models \mathcal{T} \cup \{\phi\}$, whence by the IH, we get $m \models \bot$ which is impossible; thus $m \models \neg\phi$.
- If $\mathcal{D}$ ends with
$$(\neg\mathrm{E})\ \frac{\mathcal{T} \vdash \phi \qquad \mathcal{T} \vdash \neg\phi,}{\mathcal{T} \vdash \bot}$$
then by the IH, every model of $\mathcal{T}$ satisfies both ϕ and $\neg\phi$, which is impossible; thus again there are no models of $\mathcal{T}$.
- Finally, the (C) case is similar to the ($\neg$I) case. $\square$

4. Completeness

Let $\mathcal{A}$ be an alphabet, $\mathcal{T}$ be an $\mathcal{A}$-theory, ϕ be an $\mathcal{A}$-formula. Recall that soundness (Proposition 3.22) says that if $\mathcal{T} \vdash \phi$, then $\mathcal{T} \models \phi$, i.e., "provable statements are true in all models".

Theorem 4.1 (completeness)**.** If $\mathcal{T} \models \phi$, then $\mathcal{T} \vdash \phi$.

Our proof strategy will be as follows. We prove the contrapositive: that is, we suppose there is no deduction of $\mathcal{T} \vdash \phi$, and we must show that $\mathcal{T} \not\models \phi$, i.e., that there is a model $m \models \mathcal{T}$ which does not satisfy ϕ. To construct m, which is just a function $m : \mathcal{A} \to \{0, 1\}$, we need to define $m(P)$, i.e., specify whether or not $m \models P$, for every atomic formula $P \in \mathcal{A}$. We would like to take

$$m \models P \;:\Longleftrightarrow\; \mathcal{T} \vdash P.$$

In other words, m interprets P as true iff it has to, by soundness (since we want $m \models \mathcal{T}$). Is the resulting truth assignment m a model of $\mathcal{T}$, and does it fail to satisfy ϕ? By the above definition, m will satisfy every *atomic* $P \in \mathcal{T}$ (since $\mathcal{T} \vdash P$ by the (A) rule), as well as fail to satisfy any atomic P such that $\mathcal{T} \not\vdash P$. Thus, we would like to know also

$$m \models \phi \iff \mathcal{T} \vdash \phi$$

for non-atomic ϕ. Can we prove this by induction on ϕ?

- Suppose ($m \models \phi \iff \mathcal{T} \vdash \phi$) and ($m \models \psi \iff \mathcal{T} \vdash \psi$); we want to show that $m \models \phi \wedge \psi \iff \mathcal{T} \vdash \phi \wedge \psi$. Indeed, we have

$$\begin{aligned} m \models \phi \wedge \psi &\iff m \models \phi \text{ and } m \models \psi && \text{by definition} \\ &\iff \mathcal{T} \vdash \phi \text{ and } \mathcal{T} \vdash \psi && \text{by IH} \\ &\iff \mathcal{T} \vdash \phi \wedge \psi && \text{by the } (\wedge\text{I}) \text{ and } (\wedge\text{E}) \text{ rules.} \end{aligned}$$

- Suppose ($m \models \phi \iff \mathcal{T} \vdash \phi$) and ($m \models \psi \iff \mathcal{T} \vdash \psi$); we want to show that $m \models \phi \vee \psi \iff \mathcal{T} \vdash \phi \vee \psi$. Imitating the above, we have

$$\begin{aligned} m \models \phi \vee \psi &\iff m \models \phi \text{ or } m \models \psi && \text{by definition} \\ &\iff \mathcal{T} \vdash \phi \text{ or } \mathcal{T} \vdash \psi && \text{by IH} \\ &\implies \mathcal{T} \vdash \phi \vee \psi && \text{by the } (\vee\text{I}) \text{ rules.} \end{aligned}$$

  However, the converse of the last implication fails; e.g., $\mathcal{T} \vdash \phi \vee \neg\phi$ always by (LEM), though $\mathcal{T}$ may not prove either ϕ or $\neg\phi$.
- We have $m \models \top$ always, and also $\mathcal{T} \vdash \top$ always by ($\top$I).
- We have $m \not\models \bot$ always, and also $\mathcal{T} \not\vdash \bot$, or else by ($\bot$E) we would have $\mathcal{T} \vdash \phi$ for all ϕ; but we are assuming $\mathcal{T} \not\vdash \phi$ for some ϕ.
- Finally, suppose $m \models \phi \iff \mathcal{T} \vdash \phi$; we want to show $m \models \neg\phi \iff \mathcal{T} \vdash \neg\phi$. We have

$$\begin{aligned} m \models \neg\phi &\iff m \not\models \phi \\ &\iff \mathcal{T} \not\vdash \phi && \text{by IH} \\ &\impliedby \mathcal{T} \vdash \neg\phi \end{aligned}$$

  since if $\mathcal{T} \vdash \neg\phi$ and also $\mathcal{T} \vdash \phi$, then by ($\neg$E), $\mathcal{T} \vdash \bot$, which as we noted above is impossible. However, the converse fails, again because $\mathcal{T}$ may prove neither of $\phi, \neg\phi$.

Based on this discussion, we isolate the two properties we need: we define a theory $\mathcal{T}$ to be

- **consistent** if $\mathcal{T} \not\vdash \bot$, or equivalently (by the ($\bot$E) rule) $\mathcal{T} \not\vdash \phi$ for at least one formula ϕ;
- **complete**[7] if for all formulas ϕ, either $\mathcal{T} \vdash \phi$ or $\mathcal{T} \vdash \neg\phi$.

[7] This unfortunate terminology is distinct from the usage in the "completeness theorem".

Lemma 4.2. Let $\mathcal{T}$ be a theory, and let $m : \mathcal{A} \to \{0,1\}$ be defined by

$$m \models P \;:\iff\; \mathcal{T} \vdash P.$$

Then $\mathcal{T}$ is consistent and complete iff for all formulas ϕ,

$$m \models \phi \iff \mathcal{T} \vdash \phi.$$

Proof. ($\Longleftarrow$) Since $m \not\models \bot$, $\mathcal{T} \nvdash \bot$. For any formula ϕ, either $m \models \phi$ or $m \not\models \phi$, i.e., $m \models \neg\phi$; thus either $\mathcal{T} \vdash \phi$ or $\mathcal{T} \vdash \neg\phi$.

($\Longrightarrow$) By induction on ϕ. Most of the cases are handled by the above discussion. The only remaining case is to show that if $\mathcal{T} \vdash \phi \vee \psi$, then $\mathcal{T} \vdash \phi$ or $\mathcal{T} \vdash \psi$. Suppose $\mathcal{T} \nvdash \phi$. Then since $\mathcal{T}$ is complete, $\mathcal{T} \vdash \neg\phi$. Now combine $\mathcal{T} \vdash \phi \vee \psi$ and $\mathcal{T} \vdash \neg\phi$ with

$$\dfrac{\mathcal{T} \vdash \phi \vee \psi \qquad \dfrac{\dfrac{\dfrac{}{\mathcal{T} \cup \{\phi\} \vdash \phi}\,(\mathrm{A}) \qquad \dfrac{\mathcal{T} \vdash \neg\phi}{\mathcal{T} \cup \{\phi\} \vdash \neg\phi}\,(\mathrm{W})}{\mathcal{T} \cup \{\phi\} \vdash \bot}\,(\neg\mathrm{E})}{\mathcal{T} \cup \{\phi\} \vdash \psi}\,(\bot\mathrm{E}) \qquad \dfrac{}{\mathcal{T} \cup \{\psi\} \vdash \psi}\,(\mathrm{A})}{\mathcal{T} \vdash \psi}\,(\vee\mathrm{E}).$$

□

Most of the remaining work in proving the completeness theorem is in the following

Lemma 4.3. Let $\mathcal{T}$ be a theory, ϕ be a formula such that $\mathcal{T} \nvdash \phi$. Then there is a complete theory $\mathcal{T}' \supseteq \mathcal{T}$ such that $\mathcal{T}' \nvdash \phi$.

Proof of completeness given preceding lemmas. Suppose $\mathcal{T} \nvdash \phi$. Then by Lemma 4.3, there is a complete theory $\mathcal{T}' \supseteq \mathcal{T}$ such that $\mathcal{T}' \nvdash \phi$; thus $\mathcal{T}'$ is also consistent. By Lemma 4.2, we get a truth assignment $m : \mathcal{A} \to \{0,1\}$ such that

$$m \models \psi \iff \mathcal{T}' \vdash \psi$$

for all formulas ψ. In particular, for all $\psi \in \mathcal{T} \subseteq \mathcal{T}'$, we have $\mathcal{T}' \vdash \psi$ (by the (A) rule) so $m \models \psi$, i.e., $m \models \mathcal{T}$; and since $\mathcal{T}' \nvdash \phi$, $m \nvdash \phi$. So m witnesses that $\mathcal{T} \not\models \phi$. □

The proof of Lemma 4.3 is via the most brute-force idea imaginable: we keep adding axioms to $\mathcal{T}$ until it becomes complete. In order for this to work, we need to know that (1) we can add a single axiom, and (2) we can repeat until we run out of things to add.

Lemma 4.4. Let $\mathcal{T} \nvdash \phi$, and let ψ be another formula. Then either $\mathcal{T} \cup \{\psi\} \nvdash \phi$ or $\mathcal{T} \cup \{\neg\psi\} \nvdash \phi$.

Proof. Suppose that $\mathcal{T} \cup \{\psi\} \vdash \phi$ and $\mathcal{T} \cup \{\neg\psi\} \vdash \phi$. Then we have

$$\dfrac{\dfrac{}{\mathcal{T} \vdash \psi \vee \neg\psi}\,(\mathrm{LEM}) \qquad \mathcal{T} \cup \{\psi\} \vdash \phi \qquad \mathcal{T} \cup \{\neg\psi\} \vdash \phi}{\mathcal{T} \vdash \phi}\,(\vee\mathrm{E}).$$

□

Lemma 4.5. Let $\mathcal{T}_0 \subseteq \mathcal{T}_1 \subseteq \cdots$ be an increasing sequence of theories such that $\mathcal{T}_n \nvdash \phi$ for each n. Then $\bigcup_n \mathcal{T}_n \nvdash \phi$.

Proof. Suppose $\bigcup_n \mathcal{T}_n \vdash \phi$. By syntactic compactness (Proposition 3.21), there are finitely many formulas $\psi_1, \dots, \psi_k \in \bigcup_n \mathcal{T}_n$ such that $\{\psi_1, \dots, \psi_k\} \vdash \phi$. For each $1 \le i \le k$, there is some n_i such that $\psi_i \in \mathcal{T}_{n_i}$. Letting $n := \max_i n_i$, we have $\psi_1, \dots, \psi_k \in \mathcal{T}_n$, so by weakening, $\mathcal{T}_n \vdash \phi$, a contradiction. □

Proof of Lemma 4.3. First, assume that $\mathcal{A}$ is countable, i.e., we can enumerate it as $\mathcal{A} = \{P_0, P_1, \dots\}$ where the indices are (finite) natural numbers. Then the set $\mathcal{L}(\mathcal{A})$ of all $\mathcal{A}$-formulas is also countable. Intuitively, this is because a formula can be represented as a finite string, and there are only countably many finite strings of each length n from a countable alphabet (since a finite Cartesian product of countable sets is countable), hence countably many finite strings in total (since a countable union of countable sets is countable). To give a rigorous proof without having to deal with finite strings, note that we can define $\mathcal{L}(\mathcal{A})$ as follows:

$$\begin{aligned}
\mathcal{L}_0(\mathcal{A}) &:= \mathcal{A}, \\
\mathcal{L}_{n+1}(\mathcal{A}) &:= \mathcal{L}_n(\mathcal{A}) \cup \{\psi \wedge \theta \mid \psi, \theta \in \mathcal{L}_n(\mathcal{A})\} \cup \{\psi \vee \theta \mid \psi, \theta \in \mathcal{L}_n(\mathcal{A})\} \cup \{\neg\psi \mid \psi \in \mathcal{L}_n(\mathcal{A})\} \cup \{\bot, \top\}, \\
\mathcal{L}(\mathcal{A}) &:= \textstyle\bigcup_{n \in \mathbb{N}} \mathcal{L}_n(\mathcal{A});
\end{aligned}$$

intuitively, $\mathcal{L}_n(\mathcal{A})$ consists of the formulas whose syntax tree has height $\le n$. We show by induction that each $\mathcal{L}_n(\mathcal{A})$ is countable. For $n = 0$, $\mathcal{L}_0(\mathcal{A}) = \mathcal{A}$ is countable by assumption. If $\mathcal{L}_n(\mathcal{A})$ is countable, then each of the sets occurring in the definition of $\mathcal{L}_{n+1}(\mathcal{A})$ is countable; e.g., we have a bijection

$$\begin{aligned}
\mathcal{L}_n(\mathcal{A})^2 &\cong \{\psi \wedge \theta \mid \psi, \theta \in \mathcal{L}_n(\mathcal{A})\} \\
(\psi, \theta) &\mapsto \psi \wedge \theta
\end{aligned}$$

and $\mathcal{L}_n(\mathcal{A})^2$ is countable because $\mathcal{L}_n(\mathcal{A})$ is, so the set on the RHS is also countable. So $\mathcal{L}_{n+1}(\mathcal{A})$ is a finite union of countable sets, hence countable. Finally, $\mathcal{L}(\mathcal{A})$ is a countable union of countable sets, hence countable.

So enumerate $\mathcal{L}(\mathcal{A}) = \{\psi_0, \psi_1, \dots\}$. Define the following increasing sequence of theories by induction:

$$\begin{aligned}
\mathcal{T}_0 &:= \mathcal{T}, \\
\mathcal{T}_{n+1} &:= \begin{cases} \mathcal{T}_n \cup \{\psi_n\} & \text{if } \mathcal{T}_n \cup \{\psi_n\} \not\vdash \phi, \\ \mathcal{T}_n \cup \{\neg\psi_n\} & \text{otherwise.} \end{cases}
\end{aligned}$$

By Lemma 4.4, in the second case we have $\mathcal{T}_n \cup \{\neg\psi_n\} \not\vdash \phi$; thus by induction on n, each $\mathcal{T}_n \not\vdash \phi$. Now let

$$\mathcal{T}' := \textstyle\bigcup_n \mathcal{T}_n.$$

Then $\mathcal{T}'$ is complete, since for any $\psi \in \mathcal{L}(\mathcal{A})$, $\psi = \psi_n$ for some n, whence either $\psi = \psi_n$ or $\neg\psi = \neg\psi_n$ is in $\mathcal{T}_{n+1} \subseteq \mathcal{T}'$, whence (by (A)) $\mathcal{T}' \vdash \psi$ or $\mathcal{T}' \vdash \neg\psi$. And $\mathcal{T}' \not\vdash \phi$ by Lemma 4.5.

In the uncountable case, by the well-ordering theorem, we may still transfinitely enumerate $\mathcal{L}(\mathcal{A}) = \{\psi_n \mid n < \alpha\}$ for some fixed ordinal α. We then inductively define a transfinite sequence of theories $\mathcal{T}_n$ for $n \le \alpha$ as above, along with

$$\mathcal{T}_n := \textstyle\bigcup_{m<n} \mathcal{T}_m \quad \text{for limit ordinals } n \le \alpha.$$

We now prove by induction as above that each $\mathcal{T}_n \not\vdash \phi$; for n a limit ordinal, this is by Lemma 4.5 which works just as well for transfinite sequences. Now letting as above

$$\mathcal{T}' := \mathcal{T}_\alpha,$$

we have that $\mathcal{T}'$ is complete and $\mathcal{T}' \not\vdash \phi$.

(Instead of transfinite induction, we can also use Zorn's lemma to find a maximal $\mathcal{T}' \supseteq \mathcal{T}$ such that $\mathcal{T}' \not\vdash \phi$; an analog of Lemma 4.5 is used to verify that the assumptions of Zorn's lemma are satisfied, while Lemma 4.4 is used to verify that such a maximal $\mathcal{T}'$ must be complete.) ∎

This concludes the proof of the completeness theorem for propositional logic.

4.1. **Consequences of completeness.** Soundness and completeness together say

Corollary 4.6. $\mathcal{T} \vdash \phi$ iff $\mathcal{T} \models \phi$.

Special cases of this say that various previously defined syntactic and semantic notions coincide:

Corollary 4.7. A formula ϕ is a provable tautology iff it is a semantic tautology, i.e., $\vdash \phi \iff \models \phi$.

If either (hence both) of these hold, we may call ϕ simply a **tautology**. For example, in Example 2.2 we showed that

$$P \wedge (Q \vee R) \to (P \wedge Q) \vee R$$

is a tautology *semantically*, by drawing a truth table, while in Example 3.6 (after adding an application of ($\to$I) at the end) we showed the same *syntactically*, by giving a formal deduction.

Corollary 4.8. A theory $\mathcal{T}$ is consistent iff it is satisfiable, i.e., $\mathcal{T} \nvdash \bot \iff \mathcal{T} \not\models \bot$.

For a theory $\mathcal{T}$, we let

$$\mathrm{Mod}(\mathcal{T}) := \{m : \mathcal{A} \to \{0,1\} \mid m \models \mathcal{T}\} \subseteq \{0,1\}^{\mathcal{A}} = \{\text{all } m : \mathcal{A} \to \{0,1\}\}.$$

The above corollary says $\mathcal{T}$ is consistent iff $\mathrm{Mod}(\mathcal{T}) \neq \varnothing$, i.e., iff $|\mathrm{Mod}(\mathcal{T})| \geq 1$. On the other hand,

Exercise 4.9 (HW4)**.** A theory $\mathcal{T}$ is complete iff it has at most one model, i.e., iff $|\mathrm{Mod}(\mathcal{T})| \leq 1$.

Corollary 4.10. A theory $\mathcal{T}$ is consistent and complete iff it has a unique model, i.e., iff $\mathrm{Mod}(\mathcal{T}) = \{m\}$ for some $m : \mathcal{A} \to \{0,1\}$.

Exercise 4.11. Verify that in this case, the unique model m is that defined in Lemma 4.2.

The following yields a kind of converse to Corollary 4.10:

Proposition 4.12. For any truth assignment $m : \mathcal{A} \to \{0,1\}$, the singleton $\{m\}$ is $\mathrm{Mod}(\mathcal{T})$ for some theory $\mathcal{T}$ (which is then necessarily complete and consistent, by Corollary 4.10).

Proof. There are several possibilities: one is to take

$$\mathcal{T} = \{P \in \mathcal{A} \mid m \models P\} \cup \{\neg P \mid m \not\models P\},$$

which exactly forces each of its models $n \models \mathcal{T}$ to agree with m on all atomic formulas, hence be equal to m.

Another possibility is

$$\mathrm{Th}(m) := \{\phi \in \mathcal{L}(\mathcal{A}) \mid m \models \phi\};$$

clearly this latter theory is also satisfied by m, and it contains the above $\mathcal{T}$, hence any $n \models \mathrm{Th}(m)$ must also be a model of $\mathcal{T}$, hence equal m.

(If $\mathcal{A}$ is finite, we could also take the single formula which is the conjunction of all the formulas in $\mathcal{T}$. However, if $\mathcal{A}$ is infinite, then no single formula can work; see Example 4.19 below.) □

The theory $\mathrm{Th}(m)$ defined in the preceding proof is called **the complete theory of** m. Note that this is yet *another* use of the term "complete"; but at least $\mathrm{Th}(m)$ is *a complete* (as well as consistent) theory in the previously-defined sense.

In general, we say that an arbitrary set of truth assignments $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is **axiomatizable** if

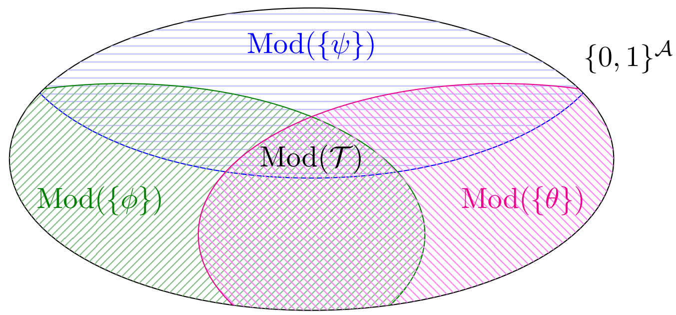
$$\mathcal{K} = \mathrm{Mod}(\mathcal{T}) \quad \text{for some } \mathcal{T},$$

in which case we say $\mathcal{T}$ **axiomatizes** $\mathcal{K}$. Thus, Proposition 4.12 says that any singleton $\mathcal{K} = \{m\}$ is axiomatizable. In Examples 2.6 to 2.8, and on HW2, we saw several other examples of axiomatizable sets of truth assignments. Informally, these are sets of "possible worlds" which can be "described" in propositional logic.

Remark 4.13. It is important to understand the distinction between the theory $\mathcal{T}$ and the set $\mathcal{K} = \mathrm{Mod}(\mathcal{T})$ it axiomatizes. These are two entirely different kinds of mathematical objects: $\mathcal{T}$ is a set of *formulas*, while $\mathcal{K}$ is a set of *truth assignments*. Each $\phi \in \mathcal{T}$ imposes a *constraint* on the truth assignments which are allowed. For example, say $\mathcal{T}$ consists of the three formulas

$$\mathcal{T} = \{\phi, \psi, \theta\}.$$

We can visualize the sets of truth assignments satisfying each of these formulas as follows:



Note that the *more* axioms are in $\mathcal{T}$, the *fewer* models there will be in $\mathcal{K} = \mathrm{Mod}(\mathcal{T})$. This "duality" between $\mathcal{T}$ and $\mathcal{K}$ is well-illustrated by the following list of basic properties of axiomatizable sets:

Proposition 4.14.

(a) If $\mathcal{T} \subseteq \mathcal{T}'$ are two theories, then $\mathrm{Mod}(\mathcal{T}) \supseteq \mathrm{Mod}(\mathcal{T}')$.
(b) $\varnothing$ and $\{0,1\}^{\mathcal{A}}$ are axiomatizable, e.g., by $\{\bot\}$ and $\varnothing$ respectively.
(c) Let $\mathcal{K}_i \subseteq \{0,1\}^{\mathcal{A}}$ be an arbitrary indexed family of axiomatizable sets (for all i in some index set I, possibly infinite). Then $\bigcap_{i\in I} \mathcal{K}_i$ is axiomatizable. Namely, if $\mathcal{K}_i = \mathrm{Mod}(\mathcal{T}_i)$, then

$$\bigcap_{i\in I} \mathcal{K}_i = \mathrm{Mod}(\bigcup_{i\in I} \mathcal{T}_i).$$

(d) If $\mathcal{K}_1, \mathcal{K}_2 \subseteq \{0,1\}^{\mathcal{A}}$ are axiomatizable, then so is $\mathcal{K}_1 \cup \mathcal{K}_2$.

Note that (d), together with the first part of (b) (which is the case of "0-ary union"), implies that a finite union of axiomatizable sets is axiomatizable, by induction.[8]

Proof. (a) This is rather obvious (and was already noted above): if $\mathcal{T} \subseteq \mathcal{T}'$, then every model of $\mathcal{T}'$ will in particular satisfy all axioms in $\mathcal{T}$.

(b) Clearly $\mathrm{Mod}(\{\bot\}) = \varnothing$ (no m satisfies $\bot$) and $\mathrm{Mod}(\varnothing) = \{0,1\}^{\mathcal{A}}$ (vacuously). We note also that there are several other natural theories that work here, e.g., $\mathrm{Mod}(\mathcal{L}(\mathcal{A})) = \varnothing$ (by (a), since $\{\bot\} \subseteq \mathcal{L}(\mathcal{A})$) and $\mathrm{Mod}(\{\top\}) = \{0,1\}^{\mathcal{A}}$ (every m satisfies $\top$).

(c) $m \in \bigcap_i \mathcal{K}_i$ iff $\forall i\,(m \models \mathcal{T}_i)$ iff $\forall i\,\forall \phi \in \mathcal{T}_i\,(m \models \phi)$ iff $\forall \phi \in \bigcup_i \mathcal{T}_i\,(m \models \phi)$ iff $m \in \mathrm{Mod}(\bigcup_{i\in I} \mathcal{T}_i)$.

(d) Let $\mathcal{K}_1 = \mathrm{Mod}(\mathcal{T}_1)$ and $\mathcal{K}_2 = \mathrm{Mod}(\mathcal{T}_2)$. We claim that

$$\mathcal{T} := \{\phi \vee \psi \mid \phi \in \mathcal{T}_1,\ \psi \in \mathcal{T}_2\}$$

axiomatizes $\mathcal{K}_1 \cup \mathcal{K}_2$. Indeed, if $m \models \mathcal{T}_1$, then for every $\phi \vee \psi \in \mathcal{T}$, we have $m \models \phi \vee \psi$ because $m \models \phi$, which shows $\mathcal{K}_1 = \mathrm{Mod}(\mathcal{T}_1) \subseteq \mathrm{Mod}(\mathcal{T})$; similarly $\mathcal{K}_2 \subseteq \mathrm{Mod}(\mathcal{T})$, so $\mathcal{K}_1 \cup \mathcal{K}_2 \subseteq \mathrm{Mod}(\mathcal{T})$. Conversely, if $m \notin \mathcal{K}_1 \cup \mathcal{K}_2 = \mathrm{Mod}(\mathcal{T}_1) \cup \mathrm{Mod}(\mathcal{T}_2)$, i.e., $m \not\models \mathcal{T}_1$ and $m \not\models \mathcal{T}_2$, then there is $\phi \in \mathcal{T}_1$ such that $m \not\models \phi$ and $\psi \in \mathcal{T}_2$ such that $m \not\models \psi$, whence $m \not\models \phi \vee \psi \in \mathcal{T}$, and so $m \notin \mathrm{Mod}(\mathcal{T})$. $\square$

Exercise 4.15. Show that in general, only one inclusion between the following holds:

$$\mathrm{Mod}(\mathcal{T}_1) \cup \mathrm{Mod}(\mathcal{T}_2) \text{ vs. } \mathrm{Mod}(\mathcal{T}_1 \cap \mathcal{T}_2).$$

[8] For topologists: this Proposition says that the axiomatizable sets, being closed under arbitrary intersections and finite unions, form the closed sets for a topology on $\{0,1\}^{\mathcal{A}}$. This topology turns out to be the same as the product topology; see Theorem 4.22 and Proposition 4.21 below.

Corollary 4.16 (of Proposition 4.14 and Proposition 4.12)**.** Any finite $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is axiomatizable.

Thus, if $\mathcal{A}$ is finite, then every $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is axiomatizable.

Exercise 4.17. If $\mathcal{A}$ is finite, then a subset $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ corresponds to selecting a subset of the rows of the truth table over the atomic formulas $\mathcal{A}$.

Show that in this case, if we unravel the parts of the proofs of Proposition 4.14 and Proposition 4.12 used in the preceding Corollary, then depending on which theory we use in Proposition 4.12, we get either the disjunctive normal form (DNF) formula with 1's in the rows of the truth table selected by $\mathcal{K}$, or the "conjunctive normal form (CNF)" but with the conjuncts written as separate formulas in the theory.

4.2. **Compactness and topology.** In order to show that a set of truth assignments is *not* axiomatizable, we can use the semantic counterpart of syntactic compactness (Proposition 3.21), which follows from it via soundness and completeness:

Corollary 4.18 ((semantic) compactness)**.** If $\mathcal{T} \models \phi$, then there is a finite $\mathcal{T}' \subseteq \mathcal{T}$ such that $\mathcal{T}' \models \phi$.

In particular, taking $\phi = \bot$, we get: if every finite $\mathcal{T}' \subseteq \mathcal{T}$ is satisfiable, then so is $\mathcal{T}$.

This direct semantic translation of syntactic completeness is a much deeper and more useful result, hence usually simply called "the **compactness theorem**". Intuitively, it says that in order to construct some mathematical object satisfying some conditions expressible in propositional logic (namely, the axioms in $\mathcal{T}$), it suffices to construct "approximations" satisfying only finitely many of those conditions at once; we can then take a "limit" of these approximations.

Example 4.19. Let $\mathcal{A} = \{P_0, P_1, \dots\}$. Then for any single truth assignment $m : \mathcal{A} \to \{0,1\}$, we claim that the complement

$$\mathcal{K} := \{0,1\}^{\mathcal{A}} \setminus \{m\} \subseteq \{0,1\}^{\mathcal{A}}$$

is *not* axiomatizable. For example, by taking $m(P_i) := 0$ for all i, we get that the set of all $n : \mathcal{A} \to \{0,1\}$ which satisfy at least one P_i is not axiomatizable.

To prove this claim, suppose for contradiction that $\mathcal{K}$ were axiomatizable, say $\mathcal{K} = \mathrm{Mod}(\mathcal{T})$ for some theory $\mathcal{T}$. So by definition of $\mathcal{K}$, every $n \in \{0,1\}^{\mathcal{A}} \setminus \{m\}$ must be a model of $\mathcal{T}$. We claim that m must then also be a model of $\mathcal{T}$, contradicting the definition of $\mathcal{K}$. In other words, we claim that there is a model of $\mathcal{T}$ which is equal to m; this latter condition can be enforced via any of the theories axiomatizing $\{m\}$ from the proof of Proposition 4.12, say the first one. So we are claiming

$$\mathcal{T}' := \mathcal{T} \cup \{P_i \mid m \models P_i\} \cup \{\neg P_i \mid m \not\models P_i\}$$

has a model. By the compactness theorem, it suffices to check that every finite $\mathcal{T}'' \subseteq \mathcal{T}'$ has a model. But since $\mathcal{T}''$ only contains finitely many of the axioms $P_i, \neg P_i \in \mathcal{T}'$ which impose "equals to m", a model n of $\mathcal{T}''$ only has to be *approximately* equal to m, on those finitely many P_i. We can thus find such a model n which is also in $\mathrm{Mod}(\mathcal{T}) = \{0,1\}^{\mathcal{A}} \setminus \{m\}$, say

$$n(P_i) := \begin{cases} 1 - m(P_i) & \text{if } i \text{ is least s.t. } P_i, \neg P_i \notin \mathcal{T}'', \\ m(P_i) & \text{otherwise,} \end{cases}$$

which will satisfy all the added axioms $P_i, \neg P_i \in \mathcal{T}'' \setminus \mathcal{T}$ since n agrees with m on all of those P_i, but will also satisfy $\mathcal{T}$ since n differs from m on the first P_i which $\mathcal{T}''$ doesn't mention. Thus, we've shown that every finite $\mathcal{T}'' \subseteq \mathcal{T}'$ is satisfiable, and hence $\mathcal{T}'$ is, by compactness; but by definition of $\mathcal{T}'$, a model $n \models \mathcal{T}'$ must be equal to m and yet also be in $\mathrm{Mod}(\mathcal{T}) = \{0,1\}^{\mathcal{A}} \setminus \{m\}$, a contradiction.

Exercise 4.20. Verify that the same argument shows that the set (from Example 2.8) of all $n : \mathcal{A} \to \{0,1\}$ which satisfy *exactly one* P_i is not axiomatizable.

The general outline of the proof of non-axiomatizability of $\mathcal{K}$ used above is: there is a truth assignment (namely m) which is not in $\mathcal{K}$, but which can be approximated by $n \in \mathcal{K}$; thus if $\mathcal{K}$ were axiomatizable, then using compactness we could show that m must actually be in $\mathcal{K}$, a contradiction. We now make this general idea precise.

For a set $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ of truth assignments, we say that a truth assignment $m : \mathcal{A} \to \{0,1\}$ is a **limit of truth assignments in** $\mathcal{K}$ (or more simply, a **limit point of** $\mathcal{K}$[9]) if for any finite set $\mathcal{F} \subseteq \mathcal{A}$ of atomic formulas, there is an $n \in \mathcal{K}$ which agrees with m on $\mathcal{F}$, i.e., $n(P) = m(P)$ for all $P \in \mathcal{F}$. Even though this definition only mentions atomic formulas,

Proposition 4.21. This definition is equivalent to either of:[10]

(i) for any finite set $\mathcal{F} \subseteq \mathcal{L}(\mathcal{A})$ of formulas, there is an $n \in \mathcal{K}$ which agrees with m on $\mathcal{F}$;
(ii) for any formula $\phi \in \mathcal{L}(\mathcal{A})$ such that $m \models \phi$, there is an $n \in \mathcal{K}$ such that $n \models \phi$.

Proof. Clearly (i) implies the original definition above.

Suppose the original definition holds, i.e., for any finite $\mathcal{F} \subseteq \mathcal{A}$, there is $n \in \mathcal{K}$ agreeing with m on $\mathcal{F}$; we prove (ii). Given $\phi \in \mathcal{L}(\mathcal{A})$, let $\mathcal{F}$ be the set of all atomic formulas appearing in ϕ (in the notation of Example 1.6, $\mathcal{F} := \mathrm{AT}(\phi)$). Then we get $n \in \mathcal{K}$ agreeing with m on $\mathcal{F}$, which clearly implies that n agrees with m on ϕ, so in particular $m \models \phi \implies n \models \phi$.

Finally, suppose (ii) holds; we prove (i). Let $\mathcal{F} \subseteq \mathcal{L}(\mathcal{A})$ be finite, let $\phi_1, \dots, \phi_k \in \mathcal{F}$ be those formulas which m satisfies, and let $\psi_1, \dots, \psi_l \in \mathcal{F}$ be those formulas which m does not satisfy. Then

$$m \models \phi_1 \wedge \cdots \wedge \phi_k \wedge \neg\psi_1 \wedge \cdots \wedge \neg\psi_l,$$

so by (ii), there is $n \in \mathcal{K}$ which satisfies the same conjunction on the RHS, which is to say that n agrees with m on every formula in $\mathcal{F}$. $\square$

We say that $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is **closed under limits**[11] if every limit point of $\mathcal{K}$ is already in $\mathcal{K}$. The following is now the general version of the argument given in Example 4.19:

Theorem 4.22. A set $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ of truth assignments is axiomatizable iff it is closed under limits.

Proof. ($\Longrightarrow$) Suppose for contradiction that $\mathcal{K}$ is axiomatized by a theory $\mathcal{T}$, yet is not closed under limits. Let m be a limit point of $\mathcal{K}$ which is not in $\mathcal{K}$. Let

$$\mathcal{T}' := \mathcal{T} \cup \{P \in \mathcal{A} \mid m \models P\} \cup \{\neg P \mid m \not\models P\}.$$

Then as in Example 4.19, any finite $\mathcal{T}'' \subseteq \mathcal{T}'$ is satisfiable, since we may take $n \in \mathcal{K} = \mathrm{Mod}(\mathcal{T})$ which agrees with m on all of the finitely many $P, \neg P \in \mathcal{T}'' \setminus \mathcal{T}$ (by the assumption that m is a limit point of $\mathcal{K}$), whence $n \models \mathcal{T}$ and also $n \models \mathcal{T}'' \setminus \mathcal{T}$, whence $n \models \mathcal{T}''$. Thus by compactness, $\mathcal{T}'$ is satisfiable; but a model of it must be in $\mathrm{Mod}(\mathcal{T}) = \mathcal{K}$ and yet be equal to m, which means $m \in \mathcal{K}$, a contradiction.

($\Longleftarrow$) Suppose $\mathcal{K}$ is closed under limits. We claim that $\mathcal{K}$ is axiomatized by

$$\mathrm{Th}(\mathcal{K}) := \{\phi \in \mathcal{L}(\mathcal{A}) \mid \forall n \in \mathcal{K}\, (n \models \phi)\}$$

(note that $\mathrm{Th}(m)$ from Proposition 4.12 is the special case $\mathrm{Th}(\{m\})$). Clearly $\mathcal{K} \subseteq \mathrm{Mod}(\mathrm{Th}(\mathcal{K}))$: every $n \in \mathcal{K}$ satisfies every $\phi \in \mathrm{Th}(\mathcal{K})$, by definition of $\mathrm{Th}(\mathcal{K})$. Conversely, for $m \in \mathrm{Mod}(\mathrm{Th}(\mathcal{K}))$, m satisfies every formula satisfied by every $n \in \mathcal{K}$, i.e., if m does *not* satisfy some formula, then some $n \in \mathcal{K}$ also does not satisfy it; now given ϕ such that $m \models \phi$, we have $m \not\models \neg\phi$, whence there is $n \in \mathcal{K}$ such that $n \not\models \neg\phi$, i.e., $n \models \phi$, which by Proposition 4.21(ii) means m is a limit point of $\mathcal{K}$, hence in $\mathcal{K}$ since $\mathcal{K}$ is closed under limits. Thus $\mathcal{K} = \mathrm{Mod}(\mathrm{Th}(\mathcal{K}))$, as desired. $\square$

[9] This is the usual term used in topology.

[10] For topologists: the original definition above yields the product topology on $\{0,1\}^{\mathcal{A}}$ induced by the discrete topology on $\{0,1\}$, while (ii) below yields the topology generated by the basic open sets $\mathrm{Mod}(\{\phi\})$.

[11] The usual term in topology is simply "closed".

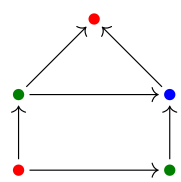
Exercise 4.23. Show that the use of compactness in the above proof is actually overkill:

(a) For any formula ϕ, $\mathrm{Mod}(\{\phi\}) \subseteq \{0,1\}^{\mathcal{A}}$ is closed under limits.
[Hint: this is trivial, using Proposition 4.21(ii) and contrapositives.]

(b) Thus, for any theory $\mathcal{T}$, $\mathrm{Mod}(\mathcal{T}) = \bigcap_{\phi \in \mathcal{T}} \mathrm{Mod}(\{\phi\})$ is closed under limits.
[In general, an intersection of sets closed under $\langle$blah$\rangle$ is always still closed under $\langle$blah$\rangle$.]

A more essential use of compactness happens when the object represented by a truth assignment we're trying to construct is not uniquely determined (unlike the m in Example 4.19); rather, we only have a theory specifying *some* conditions that have to be fulfilled. In that case, compactness can be used to give a rather mysterious proof of existence, that does not actually tell us how to get our hands on any particular instance of the object we're after. Here is a nice example of this.

A **graph** $G = (V, E)$ is (for present purposes) a **vertex** set V together with an arbitrary binary **edge** relation $E \subseteq V^2$. For $k \in \mathbb{N}$, a **k-coloring** of G is a function $c : V \to \{1, \dots, k\}$ such that whenever $(x, y) \in E$, then $c(x) \neq c(y)$. Here is a 3-coloring of a 5-vertex graph:



$({\color{red}\bullet} = 1,\ {\color{green}\bullet} = 2,\ {\color{blue}\bullet} = 3)$

For a subset of vertices $W \subseteq V$, the **induced subgraph** $G|W$ is $(W, E \cap W^2)$. Clearly, a k-coloring G restricts to a k-coloring of any induced subgraph.

Theorem 4.24 (de Bruijn–Erdős)**.** A graph $G = (V, E)$ has a k-coloring iff every finite induced subgraph of it has a k-coloring.

Proof. The idea is to represent a k-coloring of G as a model of a theory, analogous to how we represented an equivalence relation in Example 2.7. Let

$$\mathcal{A} := \{P_{x,i} \mid x \in V \text{ and } i \in \{1, \dots, k\}\},$$

where we think of $P_{x,i}$ as "$c(x) = i$". Let

$$\begin{aligned}
\mathcal{T} := \{&P_{x,1} \vee \cdots \vee P_{x,k} \mid x \in V\} && (\text{"}c(x) \text{ is defined"}) \\
\cup \{&\neg(P_{x,i} \wedge P_{x,j}) \mid x \in V \text{ and } i \neq j \in \{1, \dots, k\}\} && (\text{"}c(x) \text{ is unique"}) \\
\cup \{&\neg(P_{x,i} \wedge P_{y,i}) \mid (x, y) \in E \text{ and } i \in \{1, \dots, k\}\} && (\text{"}c \text{ is a coloring"}).
\end{aligned}$$

Given a k-coloring $c : V \to \{1, \dots, k\}$ of G, we get a model $m \models \mathcal{T}$ given by

$$(*) \qquad m \models P_{x,i} \;:\Longleftrightarrow\; c(x) = i.$$

Conversely, given $m \models \mathcal{T}$, we may define a coloring

$$\begin{aligned}
c : V &\to \{1, \dots, k\} \\
x &\mapsto \text{unique } i \text{ s.t. } m \models P_{x,i}.
\end{aligned}$$

These two constructions yield a bijection

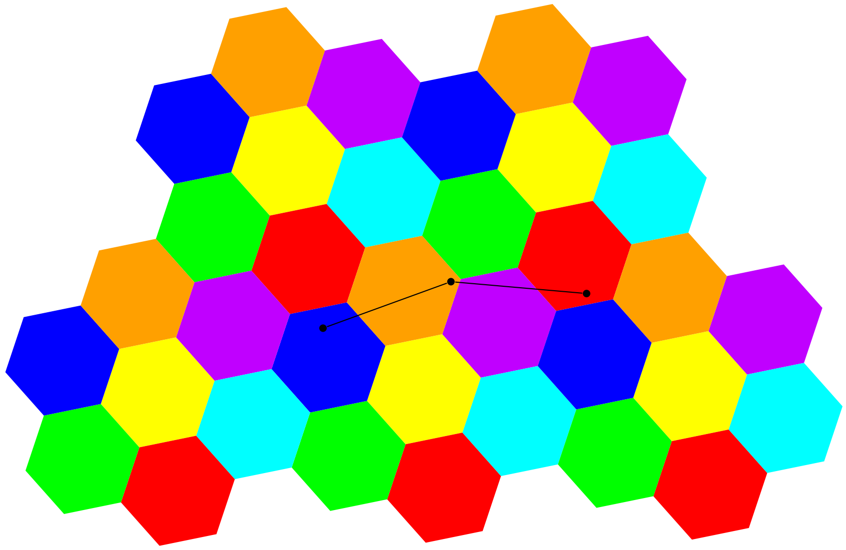
$$\mathrm{Mod}(\mathcal{T}) \longleftrightarrow \{k\text{-colorings of } G\}.$$

Thus, G has a k-coloring iff $\mathcal{T}$ is satisfiable. By compactness, $\mathcal{T}$ is satisfiable iff every finite $\mathcal{T}' \subseteq \mathcal{T}$ is satisfiable. Suppose every finite induced subgraph of G has a k-coloring. Then given finite $\mathcal{T}' \subseteq \mathcal{T}$, let $W \subseteq V$ be all finitely many vertices mentioned in $\mathcal{T}'$, let $c : W \to \{1, \dots, k\}$ be a coloring of $G|W$, and extend c arbitrarily to $V \to \{1, \dots, k\}$ (e.g., by mapping every vertex in $V \setminus W$ to 1); then the truth assignment m defined by $(*)$ satisfies $\mathcal{T}'$, since all the axioms in $\mathcal{T}'$ correspond to a constraint on the finite subgraph $G|W$ on which c is indeed a coloring. Thus, every finite $\mathcal{T}' \subseteq \mathcal{T}$ is satisfiable, and so $\mathcal{T}$ is satisfiable, i.e., G has a k-coloring. $\square$

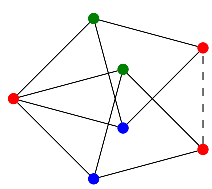
A famous example of an infinite graph is the **unit distance graph** $G = (\mathbb{R}^2, E)$ where

$$E := \{(x, y) \in \mathbb{R}^2 \times \mathbb{R}^2 \mid \mathrm{dist}(x, y) = 1\}.$$

Via a hexagonal tiling, it is easy to 7-color this graph:



It is currently an open problem what is the minimum number of colors needed. By the de Bruijn–Erdős theorem, if k colors are not enough, then there must be a *finite* configuration of points which shows that k colors are not enough. The following configuration (known as the **Moser spindle**) shows that 3 colors are not enough:



This was the best known lower bound until very recently, when de Grey (2018) found (with the help of a computer program) a 1581-point configuration which cannot be 4-colored. So it is now known that the minimum number of colors needed is one of $5, 6, 7$, but nobody knows which!

4.3. **Definability.** We close with another important theoretical application of compactness:

Theorem 4.25. Let $\mathcal{T}_1, \mathcal{T}_2$ be theories such that $\mathrm{Mod}(\mathcal{T}_1 \cup \mathcal{T}_2) = \mathrm{Mod}(\mathcal{T}_1) \cap \mathrm{Mod}(\mathcal{T}_2) = \varnothing$. Then there is a formula ϕ such that $\mathcal{T}_1 \models \phi$ and $\mathcal{T}_2 \models \neg\phi$.

Proof. First, consider the case where $\mathcal{T}_1$ is finite, say $\mathcal{T}_1 = \{\phi_1, \dots, \phi_n\}$. Then $\phi := \phi_1 \wedge \cdots \wedge \phi_n$ works: clearly $\mathrm{Mod}(\{\phi\}) = \mathrm{Mod}(\mathcal{T}_1)$, thus $\mathcal{T}_1 \models \phi$ (every model of $\mathcal{T}_1$ satisfies ϕ) and $\mathcal{T}_2 \models \neg\phi$ since $\mathrm{Mod}(\{\phi\}) \cap \mathrm{Mod}(\mathcal{T}_2) = \mathrm{Mod}(\mathcal{T}_1) \cap \mathrm{Mod}(\mathcal{T}_2) = \varnothing$.

In the general case, if $\mathrm{Mod}(\mathcal{T}_1 \cup \mathcal{T}_2) = \varnothing$, then by compactness there is a finite $\mathcal{T}' \subseteq \mathcal{T}_1 \cup \mathcal{T}_2$ which is unsatisfiable; now apply the above to $\mathcal{T}_1 \cap \mathcal{T}', \mathcal{T}_2 \cap \mathcal{T}'$. ☐

Corollary 4.26. If $\mathcal{K} \subseteq \{0, 1\}^{\mathcal{A}}$ and its complement are both axiomatizable, then it is axiomatized by a single formula.

Proof. Let $\mathcal{K} = \mathrm{Mod}(\mathcal{T}_1)$ and $\{0, 1\}^{\mathcal{A}} \setminus \mathcal{K} = \mathrm{Mod}(\mathcal{T}_2)$. Then clearly $\mathrm{Mod}(\mathcal{T}_1) \cap \mathrm{Mod}(\mathcal{T}_2) = \varnothing$, whence there is a ϕ such that $\mathcal{T}_1 \models \phi$, i.e., $\mathcal{K} = \mathrm{Mod}(\mathcal{T}_1) \subseteq \mathrm{Mod}(\{\phi\})$, and $\mathcal{T}_2 \models \neg\phi$, i.e., $\{0, 1\}^{\mathcal{A}} \setminus \mathcal{K} = \mathrm{Mod}(\mathcal{T}_2) \subseteq \mathrm{Mod}(\{\neg\phi\}) = \{0, 1\}^{\mathcal{A}} \setminus \mathrm{Mod}(\{\phi\})$, i.e., $\mathrm{Mod}(\{\phi\}) \subseteq \mathcal{K}$. ☐

Now by Theorem 4.22, $\mathcal{K}$ and its complement are both axiomatizable iff both are closed under limits; we call such $\mathcal{K}$ **clopen**.[12] Intuitively, this means that $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is a property of truth assignments which is "boundaryless": if two truth assignments m, n are really close together, then they're either both in $\mathcal{K}$ or both out; no truth assignment can be "just barely" in or out of $\mathcal{K}$.

Exercise 4.27 (for those who've seen some real analysis)**.** Show that the only clopen sets in $\mathbb{R}$ (with respect to the usual notion of limit from calculus) are $\varnothing, \mathbb{R}$.

Note that sets of the form $\mathrm{Mod}(\{\phi\})$ are indeed clopen, since (as noted above) its complement is $\mathrm{Mod}(\{\neg\phi\})$. We thus get a function

$$\begin{aligned} \mathcal{L}(\mathcal{A}) &\longrightarrow \mathcal{CO}(\{0,1\}^{\mathcal{A}}) := \{\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}} \mid \mathcal{K} \text{ is clopen}\} \subseteq \mathcal{P}(\{0,1\}^{\mathcal{A}}) \\ \phi &\longmapsto \mathrm{Mod}(\{\phi\}). \end{aligned}$$

The above Corollary says that this function is surjective. It is not injective, however, since there will be many formulas which are satisfied by exactly the same truth assignments, e.g., P and $\neg\neg P$. In fact, two formulas will be mapped to the same clopen set of truth assignments satisfying them iff they are semantically equivalent (by definition), which by soundness and completeness is the same as saying they're provably equivalent. Thus if we define

$$\phi \equiv \psi \ :\Longleftrightarrow\ \phi, \psi \text{ are (provably) equivalent},$$

then $\equiv$ will be an equivalence relation on $\mathcal{L}(\mathcal{A})$, such that the above map $\phi \mapsto \mathrm{Mod}(\{\phi\})$ descends to a bijection

$$\begin{aligned} \mathcal{L}(\mathcal{A})/{\equiv} &\cong \mathcal{CO}(\{0,1\}^{\mathcal{A}}) \\ [\phi] &\mapsto \mathrm{Mod}(\{\phi\}). \end{aligned}$$

This bijection can be seen as an ultimate expression of the fact that

$$\text{syntax} \longleftrightarrow \text{semantics}$$

for propositional logic: it says that syntactic expressions (formulas) denoting statements, modulo the syntactic notion of provable equivalence between them, exactly correspond to semantically "nice" (clopen) properties of possible semantic universes (truth assignments).

Exercise 4.28. Deduce the completeness theorem, in the special case of an empty theory, from the assertion that the above map $[\phi] \mapsto \mathrm{Mod}(\{\phi\})$ is injective.

Exercise 4.29. For a theory $\mathcal{T}$, we say that a set $\mathcal{K} \subseteq \mathrm{Mod}(\mathcal{T})$ of models of $\mathcal{T}$ is **clopen in** $\mathrm{Mod}(\mathcal{T})$ if both $\mathcal{K}$ and $\mathrm{Mod}(\mathcal{T}) \setminus \mathcal{K}$ are closed under limits. Show that

$$\begin{aligned} \mathcal{L}(\mathcal{A}) &\longrightarrow \mathcal{CO}(\mathrm{Mod}(\mathcal{T})) := \{\mathcal{K} \subseteq \mathrm{Mod}(\mathcal{T}) \mid \mathcal{K} \text{ is clopen in } \mathrm{Mod}(\mathcal{T})\} \\ \phi &\longmapsto \mathrm{Mod}(\mathcal{T}) \cap \mathrm{Mod}(\{\phi\}) = \mathrm{Mod}(\mathcal{T} \cup \{\phi\}) \end{aligned}$$

is surjective, with two formulas mapped to the same set iff they are **provably equivalent modulo** $\mathcal{T}$, i.e., iff the following relation holds:

$$\phi \equiv_{\mathcal{T}} \psi \ :\Longleftrightarrow\ \mathcal{T} \vdash \phi \leftrightarrow \psi.$$

Thus, we get a bijection

$$\mathcal{L}(\mathcal{A})/{\equiv_{\mathcal{T}}} \cong \mathcal{CO}(\mathrm{Mod}(\mathcal{T})).$$

Show that the assertion that this map is injective is equivalent to the (full) completeness theorem.

[12] "Clopen" is an abbreviation for "closed and open"; "open" means the complement is closed, *not* that the set itself is not closed.