PROPOSITIONAL LOGIC

1. Propositional formulas

Propositional logic is a "toy" logic for formalizing statements that are either true or false; the formalized statements are called **formulas**. 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.

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

$$\{P, Q, R\}$$
$$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$
$$\mathbb{N} := \{0, 1, 2, \ldots\} \quad \text{(note: } 0 \in \mathbb{N} \text{ is the convention in logic)}$$
$$\{0, 1, e, \pi, +, -, \sin, \lim, \to, \infty, \textstyle\int, \ldots\}$$

1.2. **Definition.** 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 (these are called **atomic formulas**).
- If ϕ, ψ are $\mathcal{A}$-formulas, then $\phi \wedge \psi, \phi \vee \psi, \neg\phi$ are $\mathcal{A}$-formulas ("and", "or", "not").
- $\top, \bot$ are $\mathcal{A}$-formulas ("true" and "false", thought of as "0-ary" versions of $\wedge, \vee$).

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

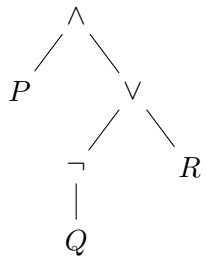
1.A. **Formal representations of formulas.** What exactly do we mean by "syntactic expression"? There are several options for formally representing formulas as familiar mathematical objects:

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

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

(over the extended alphabet $\mathcal{A} \cup \{\wedge, \vee, \neg, \top, \bot, \text{`('}, \text{`)'}\}$). 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.

(b) 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".

(c) 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 (c) 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][3] 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.

1.3. **Remark.** 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 (c), 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.B. **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:

$$\phi \to \psi := \neg\phi \vee \psi,$$
$$\phi \leftrightarrow \psi := (\phi \to \psi) \wedge (\psi \to \phi).$$

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.

[3]For everyone else: what "is" a person? A blob of cells? A blob of atoms? A quantum wavefunction? ...?

1.4. **Remark.** Why do we choose to treat $\to$ as an abbreviation, but not $\phi \vee \psi := \neg(\neg\phi \wedge \neg\psi)$, say? Mostly for pedagogical 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 (per Remark 1.3), but it will turn out to have the same meaning.

1.C. **Inductively defined sets.** Definition 1.2 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 clauses. 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** $n + 1$.[4]

Recall:

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

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

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

Here is an example of a proof by induction you've probably seen:

1.6. **Proposition.** For every $n \in \mathbb{N}$, we have

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}. \quad \Big\}\Gamma(n)$$

Proof. By induction on n.

- When $n = 0$, the LHS is the empty sum, which is 0, which is equal to the RHS.
- Now suppose

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}. \quad \Big\}\text{IH: } \Gamma(n)$$

  Then

$$\left.\begin{aligned} \sum_{k=1}^{n+1} k &= \sum_{k=1}^{n} k + (n+1) \\ &= \frac{n(n+1)}{2} + (n+1) \quad \text{by IH} \\ &= \frac{(n+1)(n+2)}{2}. \end{aligned}\right\}\Gamma(n+1)$$

□

[4] 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 $n \cup \{n\} \in \mathbb{N}$.

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

1.7. **Principle of induction for $\mathcal{L}(\mathcal{A})$.** To prove a statement $\Gamma(\phi)$ for all $\phi \in \mathcal{L}(\mathcal{A})$:

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

Here is a simple example of a proof by induction on formulas. The statement here is rather obvious; the point of the example is to get familiar with the structure of the proof.

1.8. **Proposition.** Each $\mathcal{A}$-formula $\underbrace{\phi \text{ contains only finitely many symbols}}_{\Gamma(\phi)}$.

Proof. By induction on ϕ.

- If $\phi = P \in \mathcal{A}$ is atomic, then (# of symbols in ϕ) $= 1$.
- Suppose that $\overbrace{\#\text{ of symbols in } \phi \text{ and } \#\text{ of symbols in } \psi \text{ are both finite}}^{\text{IHs: } \Gamma(\phi) \text{ and } \Gamma(\psi)}$. Then
$$(\#\text{ of symbols in } \phi \wedge \psi) = (\#\text{ of symbols in } \phi) + (\#\text{ of symbols in } \psi) + 1 < \infty \quad \Big\}\Gamma(\phi\wedge\psi)$$
by the IH.
- The $\vee$ case is similar.
- Similarly, suppose that $\overbrace{\#\text{ of symbols in } \phi < \infty}^{\text{IH: } \Gamma(\phi)}$. Then
$$(\#\text{ of symbols in } \neg\phi) = (\#\text{ of symbols in } \phi) + 1 \overset{\text{IH}}{<} \infty. \quad \Big\}\Gamma(\neg\phi)$$
- Finally, $\top, \bot$ each contain 1 symbol. □

1.9. **Remark.** Unlike induction on $\mathbb{N}$, which involves 1 base case and 1 inductive case, induction on $\mathcal{L}(\mathcal{A})$ involves 6 cases, one for each clause in the definition of $\mathcal{L}(\mathcal{A})$. Later on, we will encounter more complicated types of inductively defined expressions, involving up to 20 cases!

Thus, it is essential to learn when it is appropriate to skip inductive cases which are basically identical to other cases, such as we did for $\vee$ above. As with all proofs in advanced math, this is an art rather than a science: there's no criterion for which cases may be skipped, other than that it should be obvious to any *human* (not computer) reading that they really are similar to what you've done. A good heuristic to keep in mind is that you should always do at least the *hardest* cases; whereas if a case involves no new tricks other than what you've already done, then it may be reasonable to skip it. (So above, we could've probably also skipped the $\neg$ case, as it's simpler than the $\wedge$ case. But this is specific to this proof, not a general rule! Later on we will see proofs by induction where the $\wedge, \vee, \neg$ cases are all different, and so we should do all of them.)

Rather than prove a statement for all n or ϕ by induction, we may also define a mathematical object (number, set, etc.) parametrized by n or ϕ by induction. Recall:

1.10. **Principle of inductive definition for $\mathbb{N}$.** To define an object $g(n)$ depending on $n \in \mathbb{N}$, i.e., a *function* $g : \mathbb{N} \to Y$ with domain $\mathbb{N}$ (and some codomain Y):[5]

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

1.11. **Example.** The factorial function $! : \mathbb{N} \to \mathbb{N}$ is defined inductively as follows:
$$0! := 1,$$
$$(n+1)! := (n+1) \cdot n!.$$

[5]Also known as the **principle of recursion**.

Analogously, we have:

1.12. **Principle of inductive definition for $\mathcal{L}(\mathcal{A})$.** To define an object $g(\phi)$ depending on an $\mathcal{A}$-formula ϕ, i.e., a function g with domain $\mathcal{L}(\mathcal{A})$:

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

1.13. **Example.** For each $\mathcal{A}$-formula ϕ, we define the number $N(\phi) \in \mathbb{N}$ of symbols appearing in ϕ inductively as follows:

$$\begin{aligned} N(P) &:= 1 \quad \text{for } P \in \mathcal{A}, \\ N(\phi \wedge \psi) &:= N(\phi) + N(\psi) + 1, \\ N(\phi \vee \psi) &:= N(\phi) + N(\psi) + 1, \\ N(\neg\phi) &:= N(\phi) + 1, \\ N(\top) &:= N(\bot) := 1. \end{aligned}$$

Note that this makes precise the notion of the "# of symbols appearing in ϕ" used in Proposition 1.8. That is, we should've really defined this N function first, and then referred to it in the statement and proof of Proposition 1.8.

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

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

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

1.15. **Proposition.** For each $\mathcal{A}$-formula ϕ, we have $|\mathrm{AT}(\phi)| \leq N(\phi)$.

This says that the number of atomic formulas appearing in ϕ (not counting repeats) is at most the total number of symbols in ϕ (counting repeats), which is a good "test case" to see that the above inductive definitions actually capture our intuitions.

Proof. By induction on ϕ.

- If $\phi = P \in \mathcal{A}$, then $|\mathrm{AT}(\phi)| = |\{P\}| = 1 = N(\phi)$.
- Suppose $|\mathrm{AT}(\phi)| \leq N(\phi)$ and $|\mathrm{AT}(\psi)| \leq N(\psi)$. Then

$$\begin{aligned} |\mathrm{AT}(\phi \wedge \psi)| &= |\mathrm{AT}(\phi) \cup \mathrm{AT}(\psi)| && \text{by definition of AT} \\ &\leq |\mathrm{AT}(\phi)| + |\mathrm{AT}(\psi)| \\ &\leq N(\phi) + N(\phi) && \text{by IH} \\ &\leq N(\phi) + N(\phi) + 1 \\ &= N(\phi \wedge \psi) && \text{by definition of } N. \end{aligned}$$

  Similarly, $|\mathrm{AT}(\phi \vee \psi)| \leq N(\phi \vee \psi)$.
- Suppose $|\mathrm{AT}(\phi)| \leq N(\phi)$. Then

$$|\mathrm{AT}(\neg\phi)| = |\mathrm{AT}(\phi)| \overset{\text{IH}}{\leq} N(\phi) \leq N(\phi) + 1 = N(\neg\phi).$$

- Finally, $|\mathrm{AT}(\top)| = |\varnothing| = 0 \leq 1 = N(\top)$, and similarly for $\bot$. □

1.16. **Exercise.** What statement $\Gamma(\phi)$ are we proving by induction above, and what are the induction hypotheses in each case?

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

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

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

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

1.18. **Exercise.** What are the subformulas of $P \leftrightarrow \neg Q$, according to this definition?

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

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 $\bot$. □

1.20. **Proposition.** For any $\mathcal{A}$-formulas ϕ, ψ, if $\phi \in \mathrm{SF}(\psi)$, then $\mathrm{SF}(\phi) \subseteq \mathrm{SF}(\psi)$.

Proof. More explicitly, we must prove that every $\theta \in \mathrm{SF}(\phi)$ is in $\mathrm{SF}(\psi)$; in other words, if θ is a subformula of ϕ, and ϕ is a subformula of ψ, then θ is a subformula of ψ. We induct on ψ.

- If $\psi = P \in \mathcal{A}$ is atomic, $\top$, or $\bot$, then $\phi \in \mathrm{SF}(\psi)$ means (by definition of $\mathrm{SF}(\psi)$) that $\phi = \psi$; so $\theta \in \mathrm{SF}(\psi)$ follows from the assumption $\theta \in \mathrm{SF}(\phi)$.
- Suppose $\psi = \psi_1 \wedge \psi_2$, and (IH) every subformula of a subformula of ψ_1 is a subformula of ψ_1, and the same holds for ψ_2. Then for $\phi \in \mathrm{SF}(\psi)$, there are 3 possibilities:
  - $\phi \in \mathrm{SF}(\psi_1)$. Then by the IH, $\theta \in \mathrm{SF}(\phi) \implies \theta \in \mathrm{SF}(\psi_1) \subseteq \mathrm{SF}(\psi)$.
  - $\phi \in \mathrm{SF}(\psi_2)$. This is similar to the previous subcase.
  - $\phi = \psi$. Then as in the atomic case, $\theta \in \mathrm{SF}(\phi) = \mathrm{SF}(\psi)$.
- The $\vee$ case is similar, as is the (slightly simpler) $\neg$ case. □

1.21. **Exercise.** What is the statement Γ that we are proving by induction above?
[Hint: it should have two $\forall$'s and two $\implies$'s.]

1.22. **Exercise.** Are the $\forall$'s and $\implies$'s necessary? [Some are; some aren't.]

1.23. **Exercise.** What happens if we try to do induction on ϕ or θ instead?

2. Propositional semantics

Formulas are a purely syntactical notion, consisting of essentially a string of meaningless symbols. The process of assigning meaning to any kind of syntax yields a function

$$\text{syntax} \longrightarrow \text{semantics},$$

called **interpretation**. Since propositional logic can only express pure statements, the "semantics" in this case would consist only of truth values.

2.1. **Definition.** Let $\mathcal{A}$ be an alphabet. An **$\mathcal{A}$-truth assignment** (also known as **$\mathcal{A}$-valuation**) is an arbitrary 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 $m : \mathcal{L}(\mathcal{A}) \to \{0, 1\}$:

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

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., "$3 < \pi$". It does not make sense to take the min/max of "$m \models \phi$", or to write "$1 - (m \models \phi)$" as above, just as it would not make sense to write "$1 - (3 < \pi)$", say. Instead, in the $\models$ notation, the above inductive definition of m becomes:

$$\begin{aligned} m \models \phi \wedge \psi &\iff m \models \phi \text{ and } m \models \psi, \\ m \models \phi \vee \psi &\iff m \models \phi \text{ or } m \models \psi, \\ m \models \neg\phi &\iff m \not\models \phi, \\ m \models \top &\quad \text{always}, \\ m \models \bot &\quad \text{never}. \end{aligned} \tag{2.2}$$

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.B),

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

Again, here the $\implies$ and $\iff$ on the RHS are "meta" implications.

2.3. **Example.** Even though e.g., $P \wedge Q$ and $Q \wedge P$ are distinct formulas, for any $m : \mathcal{A} \to \{0, 1\}$,

$$m(P \wedge Q) = \min(m(P), m(Q)) = \min(m(Q), m(P)) = m(Q \wedge P).$$

This gives us one way of saying that "$\wedge$ is commutative": namely, swapping the order results in another formula with the same meaning. (We will see another, purely syntactic, way in Example 3.10.) Similarly, we have other familiar "algebraic identities", such as distributivity, de Morgan's laws, etc. We formalize these semantic notions of "truth" as follows:

2.4. **Definition.** We say that a formula $\phi \in \mathcal{L}(\mathcal{A})$ is a **semantic tautology**, written

$$\models \phi,$$

if it is satisfied by all truth assignments $m : \mathcal{A} \to \{0,1\}$.

If $\models \phi \to \psi$, i.e., every m satisfying ϕ also satisfies ψ, then we say ϕ **semantically implies** ψ, or that ψ is a **semantic consequence** of ψ. (People also say "ϕ entails ψ".) We also denote this by

$$\phi \models \psi \ :\Longleftrightarrow \ \models \phi \to \psi.$$

If ϕ, ψ semantically imply each other, then we call them **semantically equivalent**, denoted

$$\phi \models\!\models \psi \ :\Longleftrightarrow \ \phi \models \psi \text{ and } \psi \models \phi \iff \models \phi \leftrightarrow \psi.$$

So the above example shows $P \wedge Q \models\!\models Q \wedge P$ (even though $P \wedge Q \neq Q \wedge P$ as formulas).

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, i.e.,

$$\phi \text{ satisfiable} \iff \not\models \neg\phi.$$

2.5. **Example.** $P \vee \neg P$ is a semantic tautology.

$P \vee \neg Q$ is satisfiable, but not a tautology (it is semantically equivalent to $\neg Q \vee P = Q \to P$).

$P \wedge \neg P$ is not satisfiable.

2.6. **Exercise.** Note that "$\models \neg\phi$" is not equivalent to "$\not\models \phi$" (otherwise "semantic tautology" and "satisfiable" would mean the same thing)! Indeed, whereas by definition (2.2), the relation "$m \models \phi$" for *fixed* m interprets $\wedge$ as "and", $\neg$ as "not", etc., the definition of "$\models \phi$" involves a "$\forall m$".

What relations ($\Longrightarrow$ and/or $\Longleftarrow$), if any, hold between:

(a) "$\models \phi \wedge \psi$" versus "$\models \phi$ and $\models \psi$"?

(b) "$\models \phi \vee \psi$" versus "$\models \phi$ or $\models \psi$"?

(c) "$\models \phi \to \psi$" versus "$\models \phi \implies \models \psi$"?

(d) "$\models \neg\phi$" versus "$\not\models \phi$"?

A good way to conceptualize the preceding notions is to "turn the satisfaction relation $\models$ around". A truth assignment $m : \mathcal{A} \to \{0,1\}$ yields by definition a function $m : \mathcal{L}(\mathcal{A}) \to \{0,1\}$; and we defined $m \models \phi$ to mean $m(\phi) = 1$. Instead, we can fix a ϕ, and ask for which m does $m \models \phi$, yielding a function $m \mapsto m(\phi)$ from the set of all truth assignments, i.e., the set of all functions $\{0,1\}^{\mathcal{A}}$. (See the Appendix for a review of this notation.)

2.7. **Definition.** For a formula $\phi \in \mathcal{L}(\mathcal{A})$, its **truth table** is the function

$$\begin{aligned} \{0,1\}^{\mathcal{A}} &\longrightarrow \{0,1\} \\ m &\longmapsto m(\phi). \end{aligned}$$

We usually think of the truth table as a table of values (of a function, in the same sense as in calculus), hence the name:

2.8. **Example.** 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

2.9. **Example.** 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

This truth table shows that $P \wedge (Q \vee R) \models (P \wedge Q) \vee R$ (because the fourth column is 1 whenever the fifth column is), i.e., $P \wedge (Q \vee R) \to (P \wedge Q) \vee R$ is a tautology (the last column is all 1); but $(P \wedge Q) \vee R \not\models P \wedge (Q \vee R)$. In general, we have:

$$\begin{aligned}
\phi \text{ tautology} &\iff \models \phi &&\iff \text{truth table of } \phi \text{ is all 1,}\\
\phi \text{ unsatisfiable} &\iff \models \neg\phi &&\iff \text{truth table of } \phi \text{ is all 0,}\\
\phi \text{ satisfiable} &\iff \not\models \neg\phi &&\iff \text{truth table of } \phi \text{ has a 1,}\\
\phi \text{ semantically implies } \psi &\iff \phi \models \psi &&\iff \text{truth table of } \phi \leq \text{truth table of } \psi.
\end{aligned} \tag{2.10}$$

2.11. **Exercise.** Verify, using a truth table, that $(\phi \to \psi) \wedge (\psi \to \phi) \models\!\models (\phi \wedge \psi) \vee (\neg\phi \wedge \neg\psi)$. (Recall that we took the former as the official definition of $\phi \leftrightarrow \psi$.)

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

2.A. **Binary relations and Galois connections.** We have seen two distinct concepts:

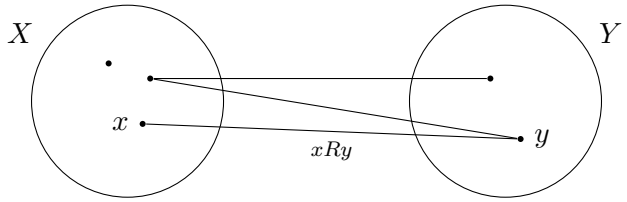
- truth assignments, i.e., functions $m : \mathcal{A} \to \{0, 1\}$, the set of all of which is denoted $\{0, 1\}^{\mathcal{A}}$;
- formulas ϕ, the set of all of which is denoted $\mathcal{L}(\mathcal{A})$.

The satisfaction relation $\models$ is a binary relation between these two sets:

$$\models \subseteq \{0, 1\}^{\mathcal{A}} \times \mathcal{L}(\mathcal{A}).$$

It is helpful at this point to zoom out a bit, and discuss some generalities that apply whenever one has an *arbitrary* binary relation (in any area of math, not just logic).

2.13. **Remark.** One way to visualize a binary relation $R \subseteq X \times Y$ between two sets X, Y is as a "bipartite graph", where we draw X, Y as blobs (as in a Venn diagram), and draw a line or "edge" between $x \in X$ and $y \in Y$ whenever the relation $x\ R\ y$ holds:



2.14. **Definition.** Let $R \subseteq X \times Y$ be a binary relation. For $A \subseteq X$ and $B \subseteq Y$, define

$$A^R := \{y \in Y \mid \forall x \in A\,(x\ R\ y)\}, \qquad {}^RB := \{x \in X \mid \forall y \in B\,(x\ R\ y)\}.$$

We call these the **right R-dual** of A and the **left R-dual** of B, respectively. These two operations

$$\mathcal{P}(X) \underset{{}^RB \hookleftarrow B}{\overset{A \mapsto A^R}{\rightleftarrows}} \mathcal{P}(Y) \tag{2.15}$$

together form the **Galois connection between $\mathcal{P}(X), \mathcal{P}(Y)$ induced by R.**

2.16. **Remark.** When $A = \{x\}$ above is a singleton, we write x^R as an abbreviation for $\{x\}^R$:

$$x^R := \{x\}^R = \{y \in Y \mid x\ R\ y\}.$$

Similarly for $y \in Y$,

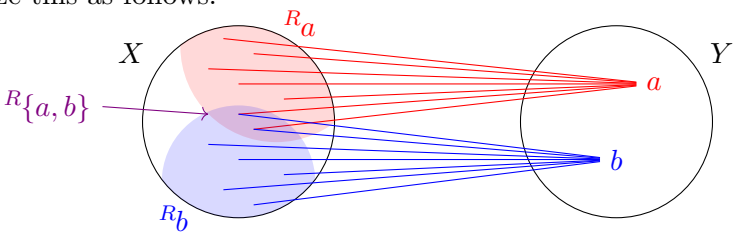
$${}^Ry := {}^R\{y\} = \{x \in X \mid x\ R\ y\}.$$

Note that for general $A \subseteq X$ and $B \subseteq Y$,

$$A^R = \bigcap_{x \in A} x^R, \qquad {}^RB = \bigcap_{y \in B} {}^Ry. \tag{2.17}$$

We may visualize this as follows:

(2.18)

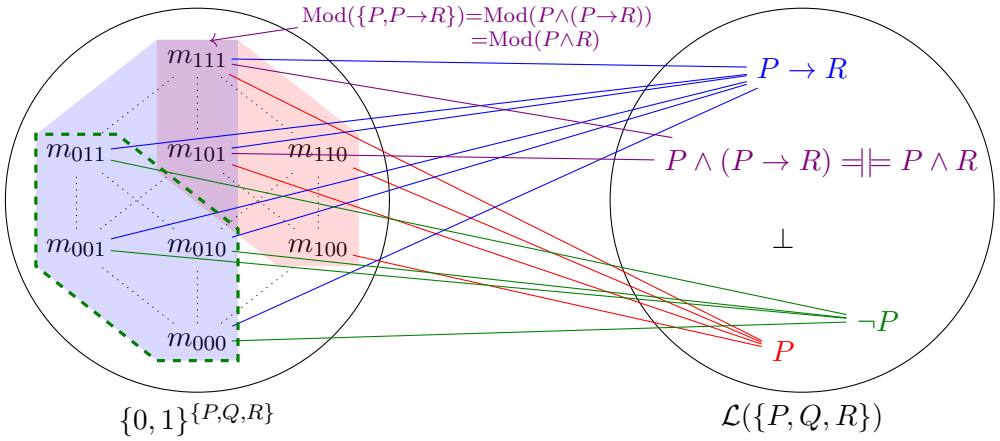


2.19. **Example** (logical satisfaction)**.** When $R = (\models)$ between $X = \{0,1\}^{\mathcal{A}}$ and $Y = \mathcal{L}(\mathcal{A})$:

- For a truth assignment $m \in X$, i.e., $m : \mathcal{A} \to \{0,1\}$,
$$m^{\models} = \{\phi \in \mathcal{L}(\mathcal{A}) \mid m \models \phi\} =: \mathrm{Th}(m)$$
is the set of formulas satisfied by m, called **the theory of** m.
- More generally, if $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is an arbitrary class of truth assignments, we let
$$\mathrm{Th}(\mathcal{K}) := \mathcal{K}^{\models} = \{\phi \in \mathcal{L}(\mathcal{A}) \mid \forall m \in \mathcal{K}\,(m \models \phi)\},$$
and call this set of formulas **the theory of** $\mathcal{K}$.
- On the other side, for a formula $\phi \in \mathcal{L}(\mathcal{A})$,
$${}^{\models}\phi = \{m : \mathcal{A} \to \{0,1\} \mid m \models \phi\} =: \mathrm{Mod}(\phi)$$
is the set of truth assignments satisfying ϕ, also known as **models** of ϕ. Its indicator function $\mathbb{1}_{\mathrm{Mod}(\phi)} : \{0,1\}^{\mathcal{A}} \to \{0,1\}$, taking $m \mapsto 1$ iff $m \models \phi$ (see Appendix), is the truth table of ϕ.
- Thus, recalling (2.10), we have
$$\begin{aligned} \models \phi \quad &\iff \quad \mathrm{Mod}(\phi) = \{0,1\}^{\mathcal{A}}, \\ \phi \models \psi \quad &\iff \quad \mathrm{Mod}(\phi) \subseteq \mathrm{Mod}(\psi), \\ \phi \models\!\!\models \psi \quad &\iff \quad \mathrm{Mod}(\phi) = \mathrm{Mod}(\psi). \end{aligned}$$
- More generally, if $\mathcal{T} \subseteq \mathcal{L}(\mathcal{A})$ is an arbitrary set of formulas, its **models** are
$$\mathrm{Mod}(\mathcal{T}) := {}^{\models}\mathcal{T} = \{m : \mathcal{A} \to \{0,1\} \mid \forall \phi \in \mathcal{T}\,(m \models \phi)\}.$$
An arbitrary set of formulas $\mathcal{T}$ is also known in logic as **a theory** (note: "a", not "the"), especially when treating it as a set of "imposed axioms" such as when talking about models.

The following picture shows this for $\mathcal{A} = \{P, Q, R\}$. Note that there are $2^3 = 8$ truth assignments, mapping each of P, Q, R to 0 or 1; we use m_{ijk} to denote the truth assignment $P \mapsto i, Q \mapsto j, R \mapsto k$.

(2.20)



Notice that, per (2.17), the models of a theory $\mathcal{T}$ are always the intersection of the models of ϕ for each $\phi \in \mathcal{T}$. If $\mathcal{T} = \{\phi_1, \dots, \phi_n\}$, this is in turn the models of the single formula $\phi_1 \wedge \cdots \wedge \phi_n$; however, $\mathcal{T}$ need not be finite in general (see Section 2.B).

2.21. **Exercise.** Verify that for any Galois connection $R \subseteq X \times Y$, generalizing (2.17), we have

$$\Big(\bigcup_{i\in I} A_i\Big)^R = \bigcap_{i\in I} A_i^R, \qquad {}^R\Big(\bigcup_{i\in I} B_i\Big) = \bigcap_{i\in I} {}^R B_i,$$

for any families of subsets $(A_i)_{i\in I}$ of X and $(B_i)_{i\in I}$ of Y. In particular,

$$A \subseteq B \subseteq X \implies A^R \supseteq B^R, \qquad A \subseteq B \subseteq Y \implies {}^R A \supseteq {}^R B. \tag{2.22}$$

There are many examples of Galois connections all over mathematics. Indeed, quite likely you've seen some or all of the following examples before, even if not under the name "Galois connection":

2.23. **Example** (linear equations/orthogonality)**.** Let X be some "space", e.g., 3D space $\mathbb{R}^3$ (= the set of all real triples (x, y, z)), let Y be the set of all equations in the coordinates of X of some form, e.g., all homogeneous linear equations $ax + by + cz = 0$, and let R mean "is a solution of". Then for a system $B \subseteq Y$ of such equations, ${}^R B$ is its set of solutions; while for a subset $A \subseteq \mathbb{R}^3$, A^R is the set of equations that hold for all points in A. For instance,

$${}^R\left\{\begin{aligned} x + y + z &= 0 \\ x + 2y + 3z &= 0 \end{aligned}\right\} = \{(t, -2t, t) \mid t \in \mathbb{R}\}.$$

Another way to view this example: we may identify the equation $ax + by + cz = 0$ with the coefficient vector (a, b, c), so that $X = Y = \mathbb{R}^3$. To say (x, y, z) is a solution of the equation with coefficients (a, b, c) means their dot product is 0, i.e., they are orthogonal vectors, usually denoted $(a, b, c) \perp (x, y, z)$. So the relation R is $\perp$; and $A^\perp = {}^\perp A$ is the *orthogonal complement* of A.

Recall from linear algebra that an orthogonal complement $A^\perp$ (in any inner product space) is always a linear subspace; this is a trivial observation (e.g., it is the kernel of a linear transformation). Less trivial is the fact that the *double* orthogonal complement $A^{\perp\perp}$ of a subset $A \subseteq \mathbb{R}^n$ is precisely the linear span of A; this depends on a Gaussian elimination argument, hence is highly specific to finite-dimensional linear algebra.[6] It follows that

$$A = A^{\perp\perp} \iff A = B^\perp \text{ for some } B \subseteq \mathbb{R}^n \iff A \text{ is a linear subspace.}$$

[6]For instance, it fails in infinite-dimensional Hilbert spaces: we must instead take the *closure* of the span.

2.24. **Example** (polynomial equations)**.** An analogous example to the above is given by again taking $X = \mathbb{R}^3$, but taking Y to be a set of more complicated equations than homogeneous linear equations, e.g., polynomial equations $f(x, y, z) = 0$ where $f \in \mathbb{R}[x, y, z] =: Y$ is a polynomial, and R means "is a root of". Of course, we can no longer think of A^R and ${}^R B$ as "orthogonal complement". But there is a more complicated characterization of ${}^R(A^R)$ for a set of points $A \subseteq \mathbb{R}^3$, called the *Zariski closure* of A, as well as of $({}^R B)^R$ for a set of polynomials $B \subseteq \mathbb{R}[x, y, z]$, called the *real radical ideal* generated by B; these are studied in the area of math known as *real algebraic geometry*. (Replacing $\mathbb{R}$ with an algebraically closed field, such as $\mathbb{C}$, yields classical algebraic geometry.)

2.25. **Remark.** Note the conceptual similarity between the preceding two examples and our main Example 2.19 of $\models$: instead of vectors of real numbers in $\mathbb{R}^3$, we have vectors of truth values in $2^{\mathcal{A}}$; and instead of linear/polynomial equations in 3 variables, we have Boolean equations in $\mathcal{A}$ variables.

2.26. **Example** (Galois theory)**.** Let X be a field, let $Y = \mathrm{Aut}(X)$ be the set of field automorphisms $f : X \to X$ (i.e., maps preserving $+, -, \cdot, /$), and let $R \subseteq X \times Y$ mean "is a fixed point of":

$$x \; R \; f \;\; :\Longleftrightarrow \;\; f(x) = x.$$

Then for $A \subseteq X$, $A^R \subseteq \mathrm{Aut}(X)$ is the set of field automorphisms fixing A; while for $B \subseteq Y$, ${}^R B \subseteq X$ is the set of fixed points of B.

It is trivially checked that for any subset $A \subseteq X$, $A^R \subseteq \mathrm{Aut}(X)$ is a subgroup of automorphisms (e.g., the composition $g \circ f$ of two automorphisms fixing each $x \in A$ clearly also fixes each $x \in A$), and that for any subset $B \subseteq Y = \mathrm{Aut}(X)$, ${}^R B \subseteq X$ is a subfield (e.g., if a field automorphism $f \in B$ fixes x, y, then it also fixes $x + y$). The *fundamental theorem of Galois theory* says that

$$A = {}^R(A^R), \qquad\qquad B = ({}^R B)^R$$

for subfields $A \subseteq X$ and subgroups $B \subseteq \mathrm{Aut}(X)$ obeying certain conditions.[7] This was the original example of a Galois connection, named after Galois who famously used it to determine when a polynomial may be solved using repeated radicals (and invented group theory in the process).

As the above examples indicate, even though the two maps in a Galois connection (2.15) need not be inverses, it is often of interest to consider the subsets on which they *do* form inverses.

2.27. **Exercise.** Let $R \subseteq X \times Y$ be a binary relation. Verify that:

(a) For $A \subseteq X$ and $B \subseteq Y$, we have $B \subseteq A^R \iff \forall x \in A, y \in B \, (x \; R \; y) \iff A \subseteq {}^R B$.
(b) For $A \subseteq X$, we have $A \subseteq {}^R(A^R)$; similarly, for $B \subseteq Y$, we have $B \subseteq ({}^R B)^R$.
(c) For $A \subseteq B \subseteq X$, we have ${}^R(A^R) \subseteq {}^R(B^R)$; similarly for subsets of Y. [Recall (2.22).]
(d) For $A \subseteq X$, we have $({}^R(A^R))^R = A^R$; similarly, for $B \subseteq Y$, we have ${}^R(({}^R B)^R) = {}^R B$.
(e) Thus for $A \subseteq X$, we have ${}^R(({}^R(A^R))^R) = {}^R(A^R)$; similarly for $B \subseteq Y$.
(f) Thus, the two maps (2.15) restrict to inverse (order-reversing) bijections

$$\{A \subseteq X \mid A = {}^R(A^R)\} \underset{{}^R B \hookleftarrow B}{\overset{A \mapsto A^R}{\underset{\longleftarrow}{\overset{\longrightarrow}{\cong}}}} \{B \subseteq Y \mid B = ({}^R B)^R\}. \tag{2.28}$$

An operation $A \mapsto {}^R(A^R)$ on subsets of a set X obeying (b), (c) and (e) is called a **closure operator** on X, and can be thought of as a very general notion of "closing under certain operations" (e.g., under linear combinations in Example 2.23). Thus, the above shows that a Galois connection between $\mathcal{P}(X), \mathcal{P}(Y)$ induces a closure operator on either side; and the Galois connection restricts to a bijection (2.28) between the sets of "closed" sets on either side. What exactly this closure operator "closes under" is highly context-dependent; see Definition 2.57 and Corollary 3.36.

[7] Namely, it suffices for X to have characteristic 0 (e.g., $X \supseteq \mathbb{Q}$), to be a finite-dimensional vector space over A (e.g., $A = \mathbb{Q}$, $X = \mathbb{Q}[i]$), and to contain all conjugates of every root of a polynomial with coefficients in A (e.g., if $\sqrt[3]{2} \in X$, then its conjugate $e^{\frac{2\pi i}{3}}$ is also in X); and for B to be closed in the pointwise convergence topology on $\mathrm{Aut}(X)$.

2.B. **Examples of theories.** We now return to the context of propositional logic, and specialize the above machinery of Galois connections to this case.

2.29. **Definition.** Let $\mathcal{A}$ be an alphabet. Recall from Example 2.19 that an **$\mathcal{A}$-theory**[8] $\mathcal{T}$ is any set of $\mathcal{A}$-formulas $\mathcal{T} \subseteq \mathcal{L}(\mathcal{A})$; a formula $\phi \in \mathcal{T}$ is called an **axiom** of $\mathcal{T}$. A truth assignment $m : \mathcal{A} \to \{0,1\}$ is a **model** of $\mathcal{T}$ if it satisfies every axiom in $\mathcal{T}$, denoted

$$m \models \mathcal{T} \iff \forall \phi \in \mathcal{T}\, (m \models \phi).$$

We denote the set of all models of $\mathcal{T}$ by

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

When $\mathcal{T} = \{\phi\}$ is a singleton, we write $\mathrm{Mod}(\phi)$ instead of $\mathrm{Mod}(\{\phi\})$.

Recall also that for any set of truth assignments $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$, the **theory of** $\mathcal{K}$ is

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

Again when $\mathcal{K} = \{m\}$ is a single truth assignment, we write simply $\mathrm{Th}(m)$.

Extending Definition 2.4, we say that a theory $\mathcal{T}$ **semantically implies** a formula ϕ, or that ϕ is a **semantic consequence** of $\mathcal{T}$, if ϕ is true in all models of $\mathcal{T}$, denoted

$$\begin{aligned} \mathcal{T} \models \phi &\iff \forall m \in \mathrm{Mod}(\mathcal{T})\, (m \models \phi) \\ &\iff \phi \in \mathrm{Th}(\mathrm{Mod}(\mathcal{T})). \end{aligned}$$

Thus, the closure operator on theories $\mathcal{T} \mapsto \mathrm{Th}(\mathrm{Mod}(\mathcal{T}))$ closes $\mathcal{T}$ under semantic consequences, by definition. (We will see a more "constructive" characterization of semantic consequence in Corollary 3.36.) We also say that $\mathcal{T}$ is **satisfiable** if it has at least one model. Note that

$$\mathcal{T} \text{ satisfiable} \iff \mathcal{T} \not\models \bot.$$

Finally, we say that a set of truth assignments $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is **axiomatizable**[9] if it is equal to $\mathrm{Mod}(\mathcal{T})$ for *some* $\mathcal{T}$. If $\mathcal{K} = \mathrm{Mod}(\mathcal{T})$, then $\mathcal{K}$ is **axiomatized by** $\mathcal{T}$. By Exercise 2.27(d), if $\mathcal{K}$ is axiomatizable, then there is a canonical $\mathcal{T}$ axiomatizing it, namely $\mathcal{T} = \mathrm{Th}(\mathcal{K})$.

When we speak of a "theory $\mathcal{T} \subseteq \mathcal{L}(\mathcal{A})$", and the elements $\phi \in \mathcal{T}$ as "axioms", the connotation is that we are interested in working under the assumption that each $\phi \in \mathcal{T}$ is "true". Semantically, this means we restrict from looking at all truth assignments $m : \mathcal{A} \to \{0,1\}$, to just the models of $\mathcal{T}$.

This is essential in semantic applications of mathematical logic to other areas. In a typical such application, we start with some *a priori* given set of objects of interest (e.g., numbers, graphs), call it S. To study S using the tools of propositional logic, we can try to come up with an alphabet $\mathcal{A}$ of propositions which we think of as representing certain important features of elements $x \in S$, so that each x is represented by a truth assignment $m : \mathcal{A} \to \{0,1\}$ giving all such properties of x. But since $\mathcal{A}$ consists of meaningless symbols, there will usually be many other "junk" truth assignments $m : \mathcal{A} \to \{0,1\}$ which do not describe any $x \in S$. The goal then is to find an $\mathcal{A}$-theory $\mathcal{T}$ carving out those "meaningful" truth assignments m which do correspond to an $x \in S$; in other words, so that we have a bijection

$$\mathrm{Mod}(\mathcal{T}) \cong S.$$

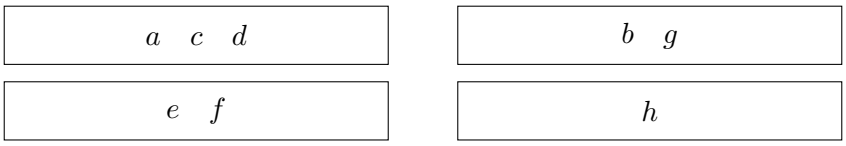
Intuitively, this means we've represented our original set of interest S as a subset of the left half of the picture (2.20), which we can then study using all the tools of propositional logic, i.e., the rest of the picture.

[8] Unfortunately, the terms *theory* and *model* have well-established meanings in logic that seem to contradict their informal meanings in science, where a "model" usually refers to a linguistic description of an underlying reality (e.g., a system of differential equations describing the motion of an atom), which logicians call a "theory".

[9] This is also sometimes known as an **elementary** set of truth assignments.

We will illustrate this idea via several examples. First, a toy example that at least pretends to be "real-world" (not just in the mathematical sense):

2.30. **Example.** Let C be a class of students, seated in a classroom with multiple horizontal tables:



We would like to describe all possible seating arrangements as models of a theory.

To do so, first we have to specify an alphabet $\mathcal{A}$, where each $P \in \mathcal{A}$ represents a single binary feature (true or false) of a seating arrangement. An essential such feature is whether two students $x, y \in C$ are seated next to each other. Thus, we can take

$$\mathcal{A} := \{P_{x,y} \mid x, y \in C\}$$

where each $P_{x,y}$ is a different symbol, thought of as

$$P_{x,y} = \text{"}x, y \text{ are seated beside each other"}.$$

Note that the above equation has absolutely no mathematical content at this point! Even though we were motivated to define the alphabet $\mathcal{A}$ in this way, the symbols $P_{x,y} \in \mathcal{A}$ are meaningless and don't know anything about x, y, or what it means for them to be "seated beside each other". In particular, there are many "junk" truth assignments $m : \mathcal{A} \to \{0, 1\}$ that say e.g., that all students are seated next to each other ($m(P) = 1\ \forall P$); that a student is seated "beside" themself; etc.

Note also that the above choice of $\mathcal{A}$ captures only certain features of the seating arrangements we are trying to describe. For example, we have chosen not to distinguish "x is seated on the left of y" from "seated on the right" (whatever that might mean: whose left/right?). We have also chosen not to include any *explicit* information about the tables themselves; the $P_{x,y}$'s allow us to determine if two students are seated at the same table (if there is a chain of students $x = x_0, x_1, \ldots, x_n = y$ between them where each $P_{x_i,x_{i+1}}$ is true), but do not distinguish "x, y are seated at table A and z is seated at table B" from "x, y are seated at table B and z is seated at table A", and also do not include any information about empty tables. *We must decide which information to encode in our alphabet $\mathcal{A}$; there is no such thing as a "complete" formalization of a real-world situation.* Even if we choose to encode tables and left/right, we still haven't encoded the GPS coordinates of each table, or how much space there is between adjacent students, or the height and posture of each student determining whether they're blocking the view of others behind them, etc.

With this in mind, let us try to axiomatize the $\mathcal{A}$-truth assignments $m : \mathcal{A} \to \{0, 1\}$ corresponding to seating arrangements, only insofar as the relevant information is encoded in m. Clearly, no student can be seated beside themself, i.e., $P_{x,x}$ should be false for all x; so we impose the axioms

$$\{\neg P_{x,x} \mid x \in C\}.$$

Since we can't distinguish left from right, we should also impose

$$\{P_{x,y} \to P_{y,x} \mid x, y \in C\}.$$

Since the rows are horizontal, no student can be seated beside three others:

$$\{\neg(P_{x,y_1} \wedge P_{x,y_2} \wedge P_{x,y_3}) \mid x, y_1, y_2, y_3 \in C \text{ and } y_1, y_2, y_3 \text{ are distinct}\}.$$

Finally, since the rows are horizontal rather than circular, we cannot have x_1 beside x_2 beside x_3 beside $\ldots$ x_n beside x_1, *provided that* (i) $n \geq 3$ (since two students can be beside each other), and (ii) the x_i are distinct (since x can be beside y who is beside x who is beside y who is beside x):

$$\{\neg(P_{x_1,x_2} \wedge P_{x_2,x_3} \wedge \cdots \wedge P_{x_{n-1},x_n} \wedge P_{x_n,x_1}) \mid n \geq 3 \text{ and } x_1, \ldots, x_n \in C \text{ are distinct}\}.$$

The union of these four sets is a theory $\mathcal{T}$ axiomatizing the desired set of truth assignments.

2.31. **Remark.** Nothing above requires the class C to be finite, or the tables to be finite. For example, we could have x_0 seated beside x_1 who is seated beside x_2 who is seated beside ....

Can we impose additional axioms to require the tables to be finite? For example, the axioms

$$\{\neg(P_{x_1,x_2} \wedge P_{x_2,x_3} \wedge P_{x_3,x_4}) \mid x_1, x_2, x_3, x_4 \in C \text{ distinct}\}$$

(in addition to the above axioms) will say that no 4 students can be seated in a row, i.e., each table can seat at most 3 students. Similarly, for any finite $n \in \mathbb{N}$, we can write down (infinitely many, if C is infinite) axioms that say there are $\leq n$ students per table. However, if C is infinite, it turns out that there is no way to say that every table is finite! Intuitively, we would need to say each table has $\leq n$ students for *some* n; but there is no way to express this infinite disjunction via a theory (unlike an infinite conjunction, where we can just take a set of axioms). See Exercise 2.65.

Similarly, we might want to say that there are only finitely many tables. It turns out that here, we cannot even say that there is ≤ 1 table, i.e., any pair of students x, y is connected by a finite chain $x = x_0, x_1, \ldots, x_n = y$ where each $P_{x_i,x_{i+1}}$ is satisfied; intuitively, we would again need an infinite $\bigvee_{n\in\mathbb{N}}$. (See Example 2.64.) But if C is finite, we can easily say this in a brute-force manner:

$$\bigvee_{\substack{\text{bijective enumerations} \\ C=\{x_1,\ldots,x_n\}}} (P_{x_1,x_2} \wedge P_{x_2,x_3} \wedge \cdots \wedge P_{x_{n-1},x_n}).$$

This is a finite formula, since there are only $n!$ ways of listing the elements of C, where $n = |C|$ is the cardinality. And if the n students are seated in a single row, then by taking $(x_1, \ldots, x_n)$ to be the listing in order along this row, the above conjunction will be satisfied.

2.32. **Exercise.** Similarly, write down a formula which axiomatizes "no student is seated by themself at a table". Does your formula depend on C being finite?

This is a general phenomenon: axiomatizability is trivial when the alphabet $\mathcal{A}$ is finite, as a consequence of the following two facts.

2.33. **Proposition.** For any $\mathcal{A}$ and $m : \mathcal{A} \to \{0, 1\}$, the singleton $\{m\} \subseteq \{0, 1\}^{\mathcal{A}}$ is axiomatizable.

Proof. The theory $\{P \in \mathcal{A} \mid m \models P\} \cup \{\neg P \mid m \not\models P\}$ works. □

Proof 2, if $\mathcal{A}$ is finite. Take the conjunction of the formulas above:

$$\bigwedge_{m\models P\in\mathcal{A}} P \wedge \bigwedge_{m\not\models P\in\mathcal{A}} \neg P. \tag{2.34}$$

□

2.35. **Proposition.** A finite union of axiomatizable $\mathcal{K}_1, \ldots, \mathcal{K}_n \subseteq \{0, 1\}^{\mathcal{A}}$ is axiomatizable.[10]

Proof. It suffices (by induction on n) to consider $n = 0, 2$.

For $n = 0$, the nullary union is $\varnothing$, which is axiomatized by $\mathcal{T} = \{\bot\}$.

For $n = 2$, suppose $\mathcal{K}_1, \mathcal{K}_2$ are axiomatized by $\mathcal{T}_1, \mathcal{T}_2$ respectively. We claim that

$$\mathcal{T}_1 \vee \mathcal{T}_2 := \{\phi_1 \vee \phi_2 \mid \phi_1 \in \mathcal{T}_1 \text{ and } \phi_2 \in \mathcal{T}_2\}$$

axiomatizes $\mathcal{K}_1 \cup \mathcal{K}_2$. We have $\mathcal{K}_1 \subseteq \operatorname{Mod}(\mathcal{T}_1 \vee \mathcal{T}_2)$, since for every $m \in \mathcal{K}_1$ and $\phi_1 \vee \phi_2 \in \mathcal{T}_1 \vee \mathcal{T}_2$ where $\phi_1 \in \mathcal{T}_1$ and $\phi_2 \in \mathcal{T}_2$, we have $m \models \phi_1$, hence $m \models \phi_1 \vee \phi_2$. Similarly, $\mathcal{K}_2 \subseteq \operatorname{Mod}(\mathcal{T}_1 \vee \mathcal{T}_2)$; thus $\mathcal{K}_1 \cup \mathcal{K}_2 \subseteq \operatorname{Mod}(\mathcal{T}_1 \vee \mathcal{T}_2)$. Conversely, if $m \notin \mathcal{K}_1 \cup \mathcal{K}_2$, then there are $\phi_1 \in \mathcal{T}_1$ and $\phi_2 \in \mathcal{T}_2$ such that $m \not\models \phi_1$ and $m \not\models \phi_2$, whence $m \not\models \phi_1 \vee \phi_2 \in \mathcal{T}_1 \vee \mathcal{T}_2$, whence $m \notin \operatorname{Mod}(\mathcal{T}_1 \vee \mathcal{T}_2)$. □

2.36. **Corollary** (of Propositions 2.33 and 2.35)**.** Any finite $\mathcal{K} \subseteq \{0, 1\}^{\mathcal{A}}$ is axiomatizable. In particular, if $\mathcal{A}$ is finite, then every $\mathcal{K} \subseteq \{0, 1\}^{\mathcal{A}}$ is axiomatizable. □

[10] Exercise 2.21 and Proposition 2.35 say that the axiomatizable $\mathcal{K} \subseteq \{0, 1\}^{\mathcal{A}}$ form the closed sets of a topology on $\{0, 1\}^{\mathcal{A}}$, namely the product topology; see Exercise 2.66.

2.37. **Remark.** When $\mathcal{A}$ is finite, and we take each singleton $\{m\} \subseteq \mathcal{K}$ to be axiomatized by a single formula as in (2.34), then the above proof shows that $\mathcal{K}$ is axiomatized by a disjunction of such formulas, one for each $m \in \mathcal{K}$. Such a disjunction of conjunctions is called the **disjunctive normal form (DNF)** of the truth table $\mathbb{1}_{\mathcal{K}} : \{0,1\}^{\mathcal{A}} \to \{0,1\}$. For example, the truth table

P	Q	R	ϕ
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	0

corresponding to $\mathcal{K} = \{(P \mapsto 0, Q \mapsto 0, R \mapsto 0), (P \mapsto 0, Q \mapsto 1, R \mapsto 1), (P \mapsto 1, Q \mapsto 0, R \mapsto 1)\}$ has DNF given by a disjunction over the 3 rows with a 1 in the truth table:

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

There are other formulas with the same truth table, e.g., here $\phi = (\neg P \wedge (Q \leftrightarrow R)) \vee (P \wedge \neg Q \wedge R)$ also works; but the DNF is a brute-force axiomatization that's guaranteed to work.

2.38. **Exercise.** In the above situation, show that if we take each singleton $\{m\} \subseteq \mathcal{K}$ to be axiomatized by a set of atomic and negated atomic formulas as in the first proof of Proposition 2.33, then the proof of Proposition 2.35 produces an axiomatization of $\mathcal{K}$ by what is essentially the **conjunctive normal form (CNF)** of the truth table $\mathbb{1}_{\mathcal{K}}$, i.e., the dual of the DNF, with the roles of $\wedge, \vee$ and $0, 1$ swapped.

2.39. **Exercise.**

(a) Give an alternative proof of Proposition 2.35 by showing that

$$\mathrm{Th}(\mathcal{K}_1) \cap \cdots \cap \mathrm{Th}(\mathcal{K}_n)$$

axiomatizes $\mathcal{K}_1 \cup \cdots \cup \mathcal{K}_n$.

(b) If each $\mathcal{K}_i$ is axiomatized by $\mathcal{T}_i$, what is the relationship between

$$\mathrm{Mod}(\mathcal{T}_1 \cap \cdots \cap \mathcal{T}_n) \qquad \text{vs.} \qquad \mathcal{K}_1 \cup \cdots \cup \mathcal{K}_n = \mathrm{Mod}(\mathcal{T}_1) \cup \cdots \cup \mathrm{Mod}(\mathcal{T}_n)?$$

Due to Corollary 2.36, the question of axiomatizability is only interesting when the set of objects S we are trying to axiomatize is (at least potentially) infinite. Thus, we will switch to describing such sets mathematically, rather than trying to make them sound "real-world"; that is, we will assume we have already formalized any "real-world" situation into a mathematically defined set S, and the remaining task is to identify S with the models of some theory $\mathcal{T}$.

A common type of such S is, as in Example 2.30, the set of all relations R (in this case binary) on some set X (in this case the class C) obeying certain axioms, in this case

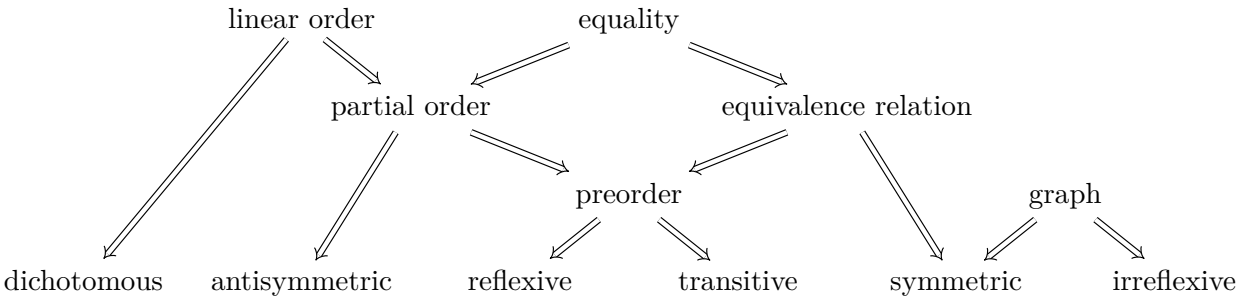
- $\forall x \in X\, (x \not R\, x)$ (no student is beside themself);
- $\forall x, y \in X\, (xRy \implies yRx)$ (we don't remember left and right);

etc. These kinds of axioms, about *relations* between *elements of sets* and involving *quantifiers* like $\forall, \exists$ as well as connectives, belong to first-order logic, not propositional logic. However, for particularly simple axioms such as those above, involving a single $\forall$ quantifier on the outside, and for a *fixed* underlying set X, we can translate those axioms into a propositional theory whose models describe such relations *on the set* X, using the technique of Example 2.30. Here are some common examples of axioms on binary relations to which this technique applies:

2.40. **Definition.** A binary relation R on a set X is:

- **reflexive** if $\forall x \in X\,(x\ R\ x)$;
- **irreflexive** if $\forall x \in X\,(x \not R\ x)$ (note: *not* the same as "not reflexive");
- **dichotomous** if $\forall x, y \in X\,(x\ R\ y \text{ or } y\ R\ x)$;
- **symmetric** if $\forall x, y \in X\,(x\ R\ y \implies y\ R\ x)$;
- **antisymmetric** if $\forall x, y \in X\,(x\ R\ y\ R\ x \implies x = y)$ (*not* the same as "not symmetric");
- **transitive** if $\forall x, y, z \in X\,(x\ R\ y\ R\ z \implies x\ R\ z)$;
- a **(simple undirected) graph** if it is irreflexive and symmetric;
- a **preorder** if it is reflexive and transitive;
- a **partial order** if it is a preorder and antisymmetric;
- a **linear order** (or **total order**) if it is a dichotomous partial order;
- an **equivalence relation** if it is a preorder and symmetric.

The following diagram depicts the relationships between these notions:



2.41. **Example.** $\leq$ is a linear order on $\mathbb{R}$ (or $\mathbb{Q}, \mathbb{Z}, \ldots$).

2.42. **Example.** For any set X, $\subseteq$ is a partial order on $\mathcal{P}(X)$. It is linear iff $|X| \leq 1$.

2.43. **Example.** For any alphabet $\mathcal{A}$, the relation $\models$ (semantic implication) is a preorder on $\mathcal{L}(\mathcal{A})$. It is never antisymmetric, since $\top \models\!\!\!\models \neg\neg\top$ but $\top \neq \neg\neg\top$.

2.44. **Example.** For any set X, the relation $\hookrightarrow$, i.e., "having cardinality $\leq$", is a preorder on $\mathcal{P}(X)$.

2.45. **Example.** For any $1 \leq n \in \mathbb{N}$, $\equiv \pmod n$ is an equivalence relation on $\mathbb{Z}$.

For any of the above axioms, and a fixed set X, it is possible to formulate a theory whose models correspond to relations on X obeying said axioms, in a manner similar to Example 2.30:

2.46. **Example.** Let X be a set. We want to describe equivalence relations $\sim$ on X via a propositional theory. Since $\sim$ is a binary relation, as was "seated beside" in Example 2.30, we take the alphabet

$$\mathcal{A} := \{P_{x,y} \mid x, y \in X\}$$

where $P_{x,y}$ is thought of as "$x \sim y$". So far, we have a bijection

$$\{0,1\}^{\mathcal{A}} \cong \mathcal{P}(X^2) = \{\text{all binary relations on } X\}$$
$$m \mapsto \{(x,y) \mid m(P_{x,y}) = 1\}$$

(m is essentially the indicator function of the relation, up to replacing $P_{x,y}$ with (x,y)). To capture just the equivalence relations, take the $\mathcal{A}$-theory

$$\begin{aligned}\mathcal{T} := &\{P_{x,x} \mid x \in X\}\\ &\cup \{P_{x,y} \to P_{y,x} \mid x, y \in X\}\\ &\cup \{P_{x,y} \wedge P_{y,z} \to P_{x,z} \mid x, y, z \in X\}.\end{aligned}$$

Then the above bijection restricts to $\mathrm{Mod}(\mathcal{T}) \cong \{\text{equivalence relations on } X\}$.

2.47. **Example.** We may similarly handle partial orders $\leq$ on a set X: with the $\mathcal{A}$ from above, let

$$\begin{aligned}\mathcal{T} := &\{P_{x,x} \mid x \in X\} \\ &\cup \{P_{x,y} \wedge P_{y,z} \to P_{x,z} \mid x, y, z \in X\};\end{aligned}$$

then models of $\mathcal{T}$ correspond to preorders (i.e., reflexive and transitive) $\leq$ on X. The antisymmetry axiom "$x \leq y \leq x \implies x = y$" is a bit tricky though; we cannot write

$$\cup \{P_{x,y} \wedge P_{y,x} \to (x = y) \mid x, y \in X\},$$

since $x = y$ is not a propositional formula.

One way around this is to take the contrapositive: instead of "$x \leq y \leq x \implies x = y$", we can say "$x \neq y \implies x \not\leq y$ or $y \not\leq x$". Of course, $x \neq y$ isn't a propositional formula either; but we can simply omit the axiom when $x = y$, yielding

$$\cup \{\neg P_{x,y} \vee \neg P_{y,x} \mid x, y \in X \text{ and } x \neq y\}.$$

A niftier way is to note that "$x = y$" is effectively a constant truth value for fixed x, y. Thus we can simply say

$$\cup \{P_{x,y} \wedge P_{y,x} \to \phi_{x,y} \mid x, y \in X\} \quad \text{where } \phi_{x,y} := \begin{cases} \top & \text{if } x = y, \\ \bot & \text{else.} \end{cases}$$

When $x \neq y$, we get the axiom $P_{x,y} \wedge P_{y,x} \to \bot$ which is semantically equivalent to $\neg(P_{x,y} \wedge P_{y,x})$. When $x = y$, we instead get a tautology $P_{x,y} \wedge P_{y,x} \to \top$, so there is no harm including it. (It would be "inefficient" if we were somehow executing these axioms on a computer; but we aren't.)

Finally, if we wanted linear orders, then we may simply add the axioms

$$\cup \{P_{x,y} \vee P_{y,x} \mid x, y \in X\}.$$

2.48. **Example.** Now consider the following situation: we already have a partial order $\leq$ on X; and we want to extend this to a linear order $\leq'$ containing $\leq$. (For example: X is a set of candidates in an election, $x \leq y$ means "every voter prefers y to x", and we want to find a linear ranking of the candidates consistent with these preferences.) Let $\mathcal{T}$ be as above, and let

$$\mathcal{T}' := \mathcal{T} \cup \{P_{x,y} \mid x \leq y\}.$$

Then the models of $\mathcal{T}'$ correspond to linear orders $\leq'$ such that $x \leq y \implies x \leq' y$, as desired.

2.49. **Exercise.** For each of the following conditions, determine if the set of partial orders $\leq$ obeying this additional condition is axiomatizable. (For the axiomatizable ones, you should be able to give the axioms; for the non-axiomatizable ones, you'd need the methods of the following subsection to prove rigorously that it's impossible to axiomatize.)

(a) Every element has at most 2 elements strictly greater than it.
(b) Every element has at least 2 elements strictly greater than it.
(c) $\leq$ is the same as $=$.
(d) $\leq$ is not a linear order.
(e) $\leq$ is irreflexive.

2.50. **Exercise.** Let X be a set.

(a) Give a theory (in some alphabet $\mathcal{A}$) whose models correspond to graphs on X.
(b) Describe the additional graph axioms (informally, not in propositional logic), such that graphs satisfying them correspond to models of the theory given in Example 2.30.
(c) A **bipartite graph** (on vertex set X) is a graph together with a partition $X = A \sqcup B$ such that there are no edges between two vertices in A, or two vertices in B. Give a theory (in some alphabet $\mathcal{B}$) whose models correspond to bipartite graphs on X.

2.C. **Axiomatizability and topology.** We now give an alternate, "geometric", characterization of which sets of truth assignments $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ are axiomatizable, that will in particular yield a practical method of proving that a given $\mathcal{K}$ is *not* axiomatizable.

Recall (Definition 2.29) that $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is **axiomatizable** iff

$$\begin{aligned} \mathcal{K} &= \mathrm{Mod}(\mathcal{T}) && \text{for some } \mathcal{T} \subseteq \mathcal{L}(\mathcal{A}) \\ &= \bigcap_{\phi \in \mathcal{T}} \mathrm{Mod}(\phi) && \text{by (2.17).} \end{aligned}$$

In other words, sets of the form $\mathrm{Mod}(\phi) \subseteq \{0,1\}^{\mathcal{A}}$ are axiomatizable, although they are not *all* of the axiomatizable sets; those are arbitrary intersections of $\mathrm{Mod}(\phi)$'s. Recall also that this is

$$\begin{aligned} &\iff \mathcal{K} = \mathrm{Mod}(\mathrm{Th}(\mathcal{K})) && \text{by Exercise 2.27(d)} \\ &\iff \mathcal{K} \supseteq \mathrm{Mod}(\mathrm{Th}(\mathcal{K})) && \text{by Exercise 2.27(b)} \\ &= \big\{m : \mathcal{A} \to \{0,1\} \;\big|\; \forall \phi \in \mathrm{Th}(\mathcal{K})\,(m \models \phi)\big\} \\ &= \big\{m : \mathcal{A} \to \{0,1\} \;\big|\; \forall \phi \in \mathcal{L}(\mathcal{A})\,(\forall n \in \mathcal{K}\,(n \models \phi) \implies m \models \phi)\big\} && (2.51) \\ &= \big\{m : \mathcal{A} \to \{0,1\} \;\big|\; \forall \phi \in \mathcal{L}(\mathcal{A})\,(m \not\models \phi \implies \exists n \in \mathcal{K}\,(n \not\models \phi))\big\} \\ &= \big\{m : \mathcal{A} \to \{0,1\} \;\big|\; \forall \psi \in \mathcal{L}(\mathcal{A})\,(m \models \psi \implies \exists n \in \mathcal{K}\,(n \models \psi))\big\} && (2.52) \end{aligned}$$

where the last equivalence is by taking $\psi = \neg\phi$.

2.53. **Proposition.** This is also

$$= \big\{m : \mathcal{A} \to \{0,1\} \;\big|\; \forall \text{ finite } \mathcal{F} \subseteq \mathcal{L}(\mathcal{A}),\, \exists n \in \mathcal{K} \text{ s.t. } \forall \phi \in \mathcal{F}\,(m \models \phi \iff n \models \phi)\big\} \tag{2.54}$$

$$= \big\{m : \mathcal{A} \to \{0,1\} \;\big|\; \forall \text{ finite } \mathcal{F} \subseteq \mathcal{A},\, \exists n \in \mathcal{K} \text{ s.t. } (m|_{\mathcal{F}} = n|_{\mathcal{F}} : \mathcal{F} \to \{0,1\})\big\}. \tag{2.55}$$

(Here $m|_{\mathcal{F}}$ denotes the restriction of $m : \mathcal{A} \to \{0,1\}$ to the subset $\mathcal{F} \subseteq \mathcal{A}$ of its domain.)

Proof. (2.52) $\subseteq$ (2.54): Given $m \in$ (2.52) and finite $\mathcal{F} \subseteq \mathcal{L}(\mathcal{A})$, take ψ in (2.52) to be

$$\psi := \bigwedge_{m \models \phi \in \mathcal{F}} \phi \wedge \bigwedge_{m \not\models \phi \in \mathcal{F}} \neg\phi.$$

Then clearly $m \models \psi$, so since $m \in$ (2.52), there is some $n \in \mathcal{K}$ such that $n \models \psi$, which means by definition of ψ that $m \models \phi \iff n \models \phi$ for all $\phi \in \mathcal{F}$. Thus $m \in$ (2.54).

(2.54) $\subseteq$ (2.55) is obvious.

(2.55) $\subseteq$ (2.52): Given $m \in$ (2.55) and $\psi \in \mathcal{L}(\mathcal{A})$ such that $m \models \psi$, take $\mathcal{F} \subseteq \mathcal{A}$ in (2.55) to be $\mathrm{AT}(\psi)$, the set of atomic formulas appearing in ψ, which is finite by Propositions 1.8 and 1.15. Since $m \in$ (2.55), there is an $n \in \mathcal{K}$ such that $m|_{\mathcal{F}} = n|_{\mathcal{F}}$. This clearly implies $(*)$ that $m(\psi) = n(\psi)$, whence $n \models \psi$ since $m \models \psi$. Thus $m \in$ (2.52). □

2.56. **Exercise.** Prove $(*)$ above.

The main point of Proposition 2.53 is the characterization (2.55) of $\mathrm{Mod}(\mathrm{Th}(\mathcal{K}))$, which is much easier to check as it involves only atomic formulas $\in \mathcal{A}$, not arbitrary formulas $\in \mathcal{L}(\mathcal{A})$.

2.57. **Definition.** For a set of truth assignments $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$, we say $m \in \{0,1\}^{\mathcal{A}}$ is **a limit of** $\mathcal{K}$ if $m \in$ (2.55). (See Remark 2.59 for an explanation of this term.) Thus Proposition 2.53 says

$$\mathrm{Mod}(\mathrm{Th}(\mathcal{K})) = \{\text{limits of } \mathcal{K}\}.$$

2.58. **Corollary.** A set of truth assignments $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is axiomatizable iff it is **closed under limits**,[11] i.e., every limit of $\mathcal{K}$ is in $\mathcal{K}$. □

[11] A topologist would simply say **closed**; but there are too many distinct notions of "closure" in math!

2.59. **Remark.** Recall from calculus/real analysis that for a set $K \subseteq \mathbb{R}$ of real numbers, $x \in \mathbb{R}$ is a **limit** of points in K if x can be approximated up to any positive error by a point in K:

$$\forall \varepsilon > 0,\ \exists y \in K \text{ s.t. } (|x - y| < \varepsilon). \tag{2.60}$$

Comparing this with (2.55), we can interpret a limit m of a set $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ of truth assignments as meaning that m can be "approximated up to any error" by some $n \in \mathcal{K}$, where we consider $m, n : \mathcal{A} \to \{0,1\}$ "approximately equal" if they agree on a large finite subset $\mathcal{F} \subseteq \mathcal{A}$.

Recall also that in $\mathbb{R}$, an equivalent definition of x being a limit of points in $K \subseteq \mathbb{R}$ is

$$\forall \text{ open intervals } (a,b),\ (a < x < b \implies \exists y \in K\,(a < y < b)).$$

This is analogous to (2.52), where the open interval neighborhood $(a,b) \ni x$ becomes the set of models "neighborhood" $\mathrm{Mod}(\psi) \ni m$.

We will say more about the connection between limits of truth assignments, and other kinds of "limits" in mathematics, in Exercises 2.63 and 2.66.

2.61. **Example.** Let $\mathcal{A} := \{P_0, P_1, P_2, \dots\}$, and consider the truth assignments m_n for $n \in \mathbb{N}$ where

$$m_n \models P_i \ :\iff\ i < n.$$

Here is a truth table:

	P_0	P_1	P_2	P_3	$\cdots$
m_0	0	0	0	0	$\cdots$
m_1	1	0	0	0	$\cdots$
m_2	1	1	0	0	$\cdots$
m_3	1	1	1	0	$\cdots$
m_4	1	1	1	1	$\cdots$

Intuitively, these m_n seem to be "converging to" $m(P_i) := 1$ for all i. Formally, using Proposition 2.53, for any finite $\mathcal{F} \subseteq \mathcal{A}$, there is some m_n agreeing with m on each $P_i \in \mathcal{F}$, namely by taking $n > i$ for each i with $P_i \in \mathcal{F}$. Thus m is a limit point of $\{m_0, m_1, \dots\}$, but is not in that set, hence that set is not axiomatizable by Corollary 2.58. Moreover, any set containing each $m_0, m_1, \dots$ but not m is not axiomatizable either, e.g., $\{0,1\}^{\mathcal{A}} \setminus \{m\}$.

Generalizing the last sentence, we have

2.62. **Exercise.** If $\mathcal{A}$ is infinite, then for any $m \in \{0,1\}^{\mathcal{A}}$, $\{0,1\}^{\mathcal{A}} \setminus \{m\}$ is not axiomatizable.

2.63. **Exercise.** In $\mathbb{R}$, limits are more commonly discussed in terms of *sequences*, rather than sets. Recall that for a sequence $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, we say that

$$\lim_{n\to\infty} x_n = x \ :\iff\ \forall \varepsilon > 0,\ \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N,\ (|x_n - x| < \varepsilon).$$

This notion is interchangeable with the limit of a set as defined in 2.60: x is a limit of a set K iff it is a limit of a sequence contained in K; and conversely, x is a limit of $(x_n)_n$ iff it is the unique point which is a limit of $\{x_N, X_{N+1}, x_{N+2}, \dots\}$ for all $N \in \mathbb{N}$.

(a) For a sequence of truth assignments $(m_n)_{n\in\mathbb{N}} \in (\{0,1\}^{\mathcal{A}})^{\mathbb{N}}$ and $m \in \{0,1\}^{\mathcal{A}}$, we say that

$$\lim_{n\to\infty} m_n = m \ :\iff\ \forall \text{ finite } \mathcal{F} \subseteq \mathcal{A},\ \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N,\ (m_n|_{\mathcal{F}} = m|_{\mathcal{F}}).$$

Show that if $\mathcal{A}$ is *countable*, then m is a limit of $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ iff it is a limit of a sequence in $\mathcal{K}$; and conversely, m is a limit of a sequence $(m_n)_{n\in\mathbb{N}} \in (\{0,1\}^{\mathcal{A}})^{\mathbb{N}}$ iff _____.

(b) Conclude that for countable $\mathcal{A}$, a set of truth assignments $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is axiomatizable iff it is closed under limits of sequences.

(c) Verify that in the preceding example, the sequence m_n indeed converges to m.

(d) (requires some set theory) Give counterexamples to (a) and (b) if $\mathcal{A}$ is uncountable.

It is often easier to think about limits of sequences of truth assignments in examples:

2.64. **Example.** The set of seating arrangements with a single table for an infinite class $C = \mathbb{N}$ from Remark 2.31 is not axiomatizable. Consider the following single-table seating arrangements:

$$\begin{array}{ll} m_0: & \boxed{0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5\ \ \cdots} \\ m_1: & \boxed{1\ \ 0\ \ 2\ \ 3\ \ 4\ \ 5\ \ \cdots} \\ m_2: & \boxed{1\ \ 2\ \ 0\ \ 3\ \ 4\ \ 5\ \ \cdots} \\ m_3: & \boxed{1\ \ 2\ \ 3\ \ 0\ \ 4\ \ 5\ \ \cdots} \\ m_4: & \boxed{1\ \ 2\ \ 3\ \ 4\ \ 0\ \ 5\ \ \cdots} \end{array}$$

The truth assignments $m_n : \mathcal{A} = \{P_{x,y} \mid x, y \in \mathbb{N}\} \to \{0, 1\}$ representing these are as described in Example 2.30. For example,

$$m_0 \models P_{0,1}, P_{1,2}, P_{2,3}, \ldots, \qquad m_0 \not\models P_{0,0}, P_{0,2}, P_{0,3}, P_{1,3}, \ldots .$$

We claim that the m_n are converging to the truth assignment m representing the seating arrangement

$$m: \quad \boxed{1\ \ 2\ \ 3\ \ 4\ \ 5\ \ \cdots} \qquad \boxed{0}$$

where 0 is seated at a table by itself. Indeed, for any finite set $\mathcal{F} \subseteq \mathcal{A}$ of atomic formulas, let $n \in \mathbb{N}$ be sufficiently large so that $n > x, y$ for all $P_{x,y} \in \mathcal{F}$; then in the seating arrangement m_n, we have x beside y iff they are beside each other in m (namely $x, y > 0$ and $|x - y| = 1$) for all $x, y < n$, so m and m_n agree on $\mathcal{F}$. Thus m is a limit of the set of seating arrangements with a single table, but is not itself in that set, and so that set is not axiomatizable by Corollary 2.58.

(Note that student 0 is *not* "converging to infinity"! The only information that the truth assignment remembers is who's beside who, and this information is what's converging.)

2.65. **Exercise.** Show that the seating arrangements for a countably infinite class with finite tables, or finitely many tables, are not axiomatizable either.

We close this section by explaining how the notion of "limit of truth assignments" is a special case of the general "limit" in topology, which the following exercise assumes you are familiar with:

2.66. **Exercise.** Recall that a topology on a set X is specified by a collection of "open sets", which is an arbitrary collection of sets closed under finite intersections and arbitrary unions, or equivalently by their complements which are called "closed sets", which are closed under finite unions and arbitrary intersections. By Exercise 2.21 and Proposition 2.35, the axiomatizable sets form the closed sets of a topology on $\{0, 1\}^{\mathcal{A}}$, such that $\mathrm{Mod}(\mathcal{T}(\mathcal{K}))$ is the topological closure of $\mathcal{K}$.

(a) Verify that this topology on $\{0, 1\}^{\mathcal{A}}$ is none other than the **product topology** of $\mathcal{A}$ many copies of the discrete topology on $\{0, 1\}$, i.e., the coarsest topology such that $m \mapsto m(P) : \{0, 1\}^{\mathcal{A}} \to \{0, 1\}$ is continuous for each $P \in \mathcal{A}$.

(b) Verify that for $\mathcal{A} = \{P_1, P_2, P_3, \ldots\}$, we have a continuous embedding

$$\begin{aligned} \{0, 1\}^{\mathcal{A}} &\hookrightarrow \mathbb{R} \\ m &\longmapsto \sum_{n=1}^{\infty} \frac{2m(P_n)}{3^n}. \end{aligned}$$

Its image is known as the **Cantor set**.

3. Deductive systems and proofs

We now return to the syntactic side of the picture of propositional logic (2.20). We have discussed *formulas*, which are the syntactic formalization of "statements" in mathematics; and we have discussed what it means for a formula to be "true" semantically. But another important part of mathematical language is *syntactic reasoning* about truth, i.e., *proofs*. We now formalize these.

As with formulas, we need a different word for the formalized syntactic expressions representing proofs, versus the proofs themselves (which live in the "real world"); we call the former **deductions**. To give an idea, is an informal "real-world" proof; try to imagine how it might be formalized.

3.1. **Example.** Assume P, and also Q or R. Then either P and Q, or R.

Proof. We know Q or R, so there are two cases:

Case 1: Q holds.

Then both P and Q hold, so we have the first alternative of the conclusion.

Case 2: R holds.

Then we have the second alternative. □

3.2. **Definition.** Let $\mathcal{S}$ be a set (whose elements will usually be syntactic expressions of some form). A **deductive system over** $\mathcal{S}$ is a set of expressions of the form

$$(L)\ \frac{S_1 \quad S_2 \quad \cdots \quad S_n}{T}$$

where $S_1, \ldots, S_n, T \in \mathcal{S}$ (possibly with $n = 0$). Such an expression is called an **inference rule**, and is informally thought of as meaning

"from the **hypotheses** $S_1, \ldots, S_n$, we may **conclude** T".

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

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

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

  $$(L)\ \frac{S_1 \quad S_2 \quad \cdots \quad S_n}{T}$$

  is an inference rule, and

  $$\mathcal{D}_1 = \frac{\vdots}{S_1}, \qquad \mathcal{D}_2 = \frac{\vdots}{S_2}, \qquad \cdots, \qquad \mathcal{D}_n = \frac{\vdots}{S_n}$$

  are deductions ending in $S_1, \ldots, S_n$ respectively, then we have a deduction

  $$(L)\ \frac{\mathcal{D}_1 \quad \mathcal{D}_2 \quad \cdots \quad \mathcal{D}_n}{T}.$$

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

  $$(L)\ \frac{\mathcal{D}_1 \quad \mathcal{D}_2 \quad \cdots \quad \mathcal{D}_n}{T}$$

  where $\mathcal{D}_1, \ldots, \mathcal{D}_n$ are sub-deductions, the conclusion is T, 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. When we just say a **deduction of** T, we mean from *no* hypotheses.

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

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

Here is a deduction of R from P:

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

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

3.4. **Remark.** As with any kind of inductively defined expression, we can formally represent deductions as nested tuples, as in Section 1.A, although we rarely bother to do so. For example, the above deduction may be represented as

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

while the one after adding rule (J) is

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

3.5. **Example.** Here is a rather meaningless example of a deductive system. The set $\mathcal{S}$ is $\mathbb{Z}$ (the integers), and the inference rules are

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

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

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

When it is clear what the parameters are, we may omit them from the label.

Since deductions are inductively defined, we may prove things about them by induction:

3.6. **Proposition.** In the above deductive system, if there is a deduction $\mathcal{D}$ of $n \in \mathbb{Z}$ from hypotheses which are all positive numbers, then n is positive.

Proof. We prove by induction on $\mathcal{D}$ that if all hypotheses of $\mathcal{D}$ are positive, then so is its conclusion. If $\mathcal{D}$ is a single hypothesis n, then n is positive by assumption. If $\mathcal{D}$ ends in either of

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

then by the IH applied to the sub-deductions of m, n, both are > 0, hence $m + n, mn > 0$. □

3.7. **Example.** Let $\mathcal{A}$ be an alphabet. Define a deductive system over $\mathcal{L}(\mathcal{A})$ with the rule schemas:

$$\frac{\phi \quad \psi}{\phi \wedge \psi} \qquad \frac{\phi \wedge \psi}{\phi} \qquad \frac{\phi \wedge \psi}{\psi} \qquad \frac{\phi}{\phi \vee \psi} \qquad \frac{\psi}{\phi \vee \psi} \qquad \frac{\phi \quad \neg\phi}{\bot}$$

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

$$\frac{\dfrac{P \quad Q}{P \wedge Q}}{(P \wedge Q) \vee R}$$

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 formulas 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 3.7, since in the subproofs for each case, we have to temporarily introduce a new assumption (e.g., Q in Case 1 of 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.A. **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**.[12] 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)**

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

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"):

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

- For $\vee$, there are two introduction rules and one elimination rule ("case analysis"):

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

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

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

  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"):

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

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"):

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

3.8. **Definition.** 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*.[13] We also say in this case that ϕ is a **provable consequence** of $\mathcal{T}$. As with $\models$ (see Definition 2.4), when $\mathcal{T} = \{\psi\}$, we drop the braces. We say ϕ, ψ are **provably equivalent** if

$$\phi \dashv\vdash \psi \;:\Longleftrightarrow\; \phi \vdash \psi \text{ and } \psi \vdash \phi.$$

And we say ϕ is a **provable tautology** if

$$\vdash \phi \;:\Longleftrightarrow\; \varnothing \vdash \phi.$$

3.9. **Example.** In Example 2.9, we showed that $P \wedge (Q \vee R) \models (P \wedge Q) \vee R$. In Example 3.1, we gave an informal proof showing the *provable* implication $P \wedge (Q \vee R) \vdash (P \wedge Q) \vee R$. However, that proof skipped a lot of trivial steps, like going from "P and ..." to just P, that we normally wouldn't bother to write out, but are needed in the fully formalized 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 from $P \wedge (Q \vee R)$, we get P. Hence, $P \wedge Q$, (∧I) and so $(P \wedge Q) \vee R$. (∨I1)

Case 2: R holds.

(∨I2) Then we have the second alternative in $(P \wedge Q) \vee R$.

Here is the formal deduction corresponding to the above informal proof:

$$\dfrac{\dfrac{\dfrac{(\mathrm{A})}{P \wedge (Q \vee R) \vdash P \wedge (Q \vee R)}}{(\wedge\mathrm{E2})\ P \wedge (Q \vee R) \vdash Q \vee R} \qquad \dfrac{\dfrac{\dfrac{(\mathrm{A})}{\{P \wedge (Q \vee R), Q\} \vdash P \wedge (Q \vee R)}}{(\wedge\mathrm{E1})\ \{P \wedge (Q \vee R), Q\} \vdash P} \qquad \dfrac{(\mathrm{A})}{\{P \wedge (Q \vee R), Q\} \vdash Q}}{(\wedge\mathrm{I})\ \{P \wedge (Q \vee R), Q\} \vdash P \wedge Q}}{(\vee\mathrm{I1})\ \{P \wedge (Q \vee R), Q\} \vdash (P \wedge Q) \vee R} \qquad \dfrac{(\mathrm{A})}{\{P \wedge (Q \vee R), R\} \vdash R}}{(\vee\mathrm{I2})\ \{P \wedge (Q \vee R), R\} \vdash (P \wedge Q) \vee R}$$

$$(\vee\mathrm{E})\ P \wedge (Q \vee R) \vdash (P \wedge Q) \vee R$$

Note the excessive width; this tends to happen in all but the simplest deductions, due to the need to repeatedly write the assumptions. We can mitigate this somewhat by naming the assumptions:

$$\dfrac{\dfrac{\dfrac{(\mathrm{A})}{\phi \vdash P \wedge (Q \vee R)}}{(\wedge\mathrm{E2})\ \phi \vdash Q \vee R} \qquad \dfrac{\dfrac{\dfrac{(\mathrm{A})}{\mathcal{T} \vdash P \wedge (Q \vee R)}}{(\wedge\mathrm{E1})\ \mathcal{T} \vdash P} \qquad \dfrac{(\mathrm{A})}{\mathcal{T} \vdash Q}}{(\wedge\mathrm{I})\ \mathcal{T} \vdash P \wedge Q}}{(\vee\mathrm{I1})\ \mathcal{T} := \{\phi, Q\} \vdash (P \wedge Q) \vee R} \qquad \dfrac{(\mathrm{A})}{\{\phi, R\} \vdash R}}{(\vee\mathrm{I2})\ \{\phi, R\} \vdash (P \wedge Q) \vee R}$$

$$(\vee\mathrm{E})\ \phi := P \wedge (Q \vee R) \vdash (P \wedge Q) \vee R$$

[12] 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.

[13] This is analogous to how $A \cong B$ can denote a bijection between two sets, or the claim that such a bijection exists.

3.10. **Example.** For any formulas ϕ, ψ, we have $\phi \wedge \psi \vdash \psi \wedge \phi$:

$$\dfrac{\dfrac{\dfrac{}{\phi \wedge \psi \vdash \phi \wedge \psi}(\mathrm{A})}{\phi \wedge \psi \vdash \psi}(\wedge\mathrm{E2}) \qquad \dfrac{\dfrac{}{\phi \wedge \psi \vdash \phi \wedge \psi}(\mathrm{A})}{\phi \wedge \psi \vdash \phi}(\wedge\mathrm{E1})}{\phi \wedge \psi \vdash \psi \wedge \phi}(\wedge\mathrm{I})$$

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).

3.11. **Exercise.** Give deductions for the following laws:

(commutativity) $\quad (\phi \wedge \psi \dashv\vdash \psi \wedge \phi,) \qquad \phi \vee \psi \dashv\vdash \psi \vee \phi,$

(associativity) $\quad (\phi \wedge \psi) \wedge \theta \dashv\vdash \phi \wedge (\psi \wedge \theta), \qquad (\phi \vee \psi) \vee \theta \dashv\vdash \phi \vee (\psi \vee \theta),$

(idempotence) $\quad \phi \wedge \phi \dashv\vdash \phi, \qquad \phi \vee \psi \dashv\vdash \phi,$

(distributivity) $\quad \phi \wedge (\psi \vee \theta) \dashv\vdash (\phi \wedge \psi) \vee (\phi \wedge \theta), \qquad \phi \vee (\psi \wedge \theta) \dashv\vdash (\phi \vee \psi) \wedge (\phi \vee \theta).$

3.12. **Example.** For any formula ϕ, we have $\phi \vdash \neg\neg\phi$:

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

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

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

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

3.13. **Exercise.** Give deductions for

(de Morgan's laws) $\quad \neg(\phi \vee \psi) \dashv\vdash \neg\phi \wedge \neg\psi, \qquad \neg(\phi \wedge \psi) \dashv\vdash \neg\phi \vee \neg\psi.$

[Hint: exactly one of the four $\vdash$'s requires (C).]

3.14. **Remark.** 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 3.23). 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.B. **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 $S_1, \dots, S_n$.

3.15. **Example.** To say that $\mathcal{T} \vdash \phi$ is the same as saying that $\dfrac{}{\mathcal{T} \vdash \phi}$ is derivable (or admissible).

3.16. **Example.** Recalling our convention that $\phi \wedge \psi \wedge \theta := (\phi \wedge \psi) \wedge \theta$ (Section 1.B), the following ternary version of $(\wedge\mathrm{I})$

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

is derivable:

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

3.17. **Proposition** (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 $\mathcal{D} = (\mathrm{A})\,\dfrac{}{\mathcal{T} \vdash \phi}$, then $\phi \in \mathcal{T} \subseteq \mathcal{T}'$, so $\mathcal{D}' := (\mathrm{A})\,\dfrac{}{\mathcal{T}' \vdash \phi}$ 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 $(\wedge\mathrm{I})$, 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 $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$ (which say respectively that any $\mathcal{T}' \supseteq \mathcal{T}$ proves $\phi \vee \psi$, any $\mathcal{T}' \supseteq \mathcal{T} \cup \{\phi\}$ proves θ, and any $\mathcal{T}' \supseteq \mathcal{T} \cup \{\psi\}$ proves θ), we have

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

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

The rest of the cases are similar. □

3.18. **Remark.** (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 $\mathcal{T}' \nvdash \phi$, e.g., $\mathcal{T} := \varnothing$, $\mathcal{T}' := \{Q\}$, and $\phi := P$ (by soundness, Proposition 3.27).

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.

3.19. **Example.** 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:[14]

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

3.20. **Example.** Analogously to Example 3.16, the following ternary version of ($\vee$E)

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

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.

The following concrete applications of this rule, versus that of Example 3.16, illustrate the distinction between derivable versus admissible. Using the above ternary ($\vee$E), we have

$$\dfrac{\dfrac{}{\phi \vdash \phi}\,(\mathrm{A}) \qquad \dfrac{\dfrac{\dfrac{}{\{\phi, P \wedge Q\} \vdash P \wedge Q}\,(\mathrm{A})}{\{\phi, P \wedge Q\} \vdash P}\,(\wedge\mathrm{E1})}{\{\phi, P \wedge Q\} \vdash P \vee Q}\,(\vee\mathrm{I1}) \qquad \dfrac{\dfrac{\dfrac{}{\{\phi, P \wedge \neg Q\} \vdash P \wedge \neg Q}\,(\mathrm{A})}{\{\phi, P \wedge \neg Q\} \vdash P}\,(\wedge\mathrm{E1})}{\{\phi, P \wedge \neg Q\} \vdash P \vee Q}\,(\vee\mathrm{I1}) \qquad \dfrac{\dfrac{\dfrac{}{\{\phi, \neg P \wedge Q\} \vdash \neg P \wedge Q}\,(\mathrm{A})}{\{\phi, \neg P \wedge Q\} \vdash Q}\,(\wedge\mathrm{E2})}{\{\phi, \neg P \wedge Q\} \vdash P \vee Q}\,(\vee\mathrm{I2})}{\phi := (P \wedge Q) \vee (P \wedge \neg Q) \vee (\neg P \wedge Q) \vdash P \vee Q}\,(\vee\mathrm{E})$$

If we "plug in" the above derivation of ternary ($\vee$E) using (W), we get

$$\dfrac{\dfrac{}{\phi \vdash \phi}\,(\mathrm{A}) \qquad \dfrac{}{\{\phi, \psi\} \vdash \psi}\,(\mathrm{A}) \qquad \dfrac{\dfrac{\dfrac{}{\{\phi, P \wedge Q\} \vdash P \wedge Q}\,(\mathrm{A})}{\{\phi, P \wedge Q\} \vdash P}\,(\wedge\mathrm{E1})}{\{\phi, P \wedge Q\} \vdash P \vee Q}\,(\vee\mathrm{I1})\,(\mathrm{W}) \qquad \dfrac{\dfrac{\dfrac{}{\{\phi, P \wedge \neg Q\} \vdash P \wedge \neg Q}\,(\mathrm{A})}{\{\phi, P \wedge \neg Q\} \vdash P}\,(\wedge\mathrm{E1})}{\{\phi, P \wedge \neg Q\} \vdash P \vee Q}\,(\vee\mathrm{I1})\,(\mathrm{W}) \qquad \dfrac{\dfrac{\dfrac{}{\{\phi, \neg P \wedge Q\} \vdash \neg P \wedge Q}\,(\mathrm{A})}{\{\phi, \neg P \wedge Q\} \vdash Q}\,(\wedge\mathrm{E2})}{\{\phi, \neg P \wedge Q\} \vdash P \vee Q}\,(\vee\mathrm{I2})}{\{\phi, \psi := (P \wedge Q) \vee (P \wedge \neg Q)\} \vdash P \vee Q}\,(\vee\mathrm{E})}{\phi := ((P \wedge Q) \vee (P \wedge \neg Q)) \vee (\neg P \wedge Q) \vdash P \vee Q}\,(\vee\mathrm{E})$$

where the blue bits are from the deduction of ternary ($\vee$E); now eliminating the uses of (W) following the proof of Proposition 3.17, we get

$$\dfrac{\dfrac{}{\phi \vdash \phi}\,(\mathrm{A}) \qquad \dfrac{}{\{\phi, \psi\} \vdash \psi}\,(\mathrm{A}) \qquad \dfrac{\dfrac{\dfrac{}{\{\phi, \psi, P \wedge Q\} \vdash P \wedge Q}\,(\mathrm{A})}{\{\phi, \psi, P \wedge Q\} \vdash P}\,(\wedge\mathrm{E1})}{\{\phi, \psi, P \wedge Q\} \vdash P \vee Q}\,(\vee\mathrm{I1}) \qquad \dfrac{\dfrac{\dfrac{}{\{\phi, \psi, P \wedge \neg Q\} \vdash P \wedge \neg Q}\,(\mathrm{A})}{\{\phi, \psi, P \wedge \neg Q\} \vdash P}\,(\wedge\mathrm{E1})}{\{\phi, \psi, P \wedge \neg Q\} \vdash P \vee Q}\,(\vee\mathrm{I1}) \qquad \dfrac{\dfrac{\dfrac{}{\{\phi, \neg P \wedge Q\} \vdash \neg P \wedge Q}\,(\mathrm{A})}{\{\phi, \neg P \wedge Q\} \vdash Q}\,(\wedge\mathrm{E2})}{\{\phi, \neg P \wedge Q\} \vdash P \vee Q}\,(\vee\mathrm{I2})}{\{\phi, \psi := (P \wedge Q) \vee (P \wedge \neg Q)\} \vdash P \vee Q}\,(\vee\mathrm{E})}{\phi := ((P \wedge Q) \vee (P \wedge \neg Q)) \vee (\neg P \wedge Q) \vdash P \vee Q}\,(\vee\mathrm{E})$$

where the red bits are from (W).

[14] We are using here that the above proof of Proposition 3.17 continues to work when we remove the ($\bot$E) rule, since the inductive case for each rule only appeals to that same rule.

On the other hand, the following similar deduction uses the ternary $(\wedge\mathrm{I})$ from Example 3.16:

$$\dfrac{\dfrac{\dfrac{\dfrac{}{P\wedge Q\vdash P\wedge Q}(\mathrm{A})}{P\wedge Q\vdash P}(\wedge\mathrm{E1})}{P\wedge Q\vdash P\vee Q}(\vee\mathrm{I1}) \quad \dfrac{\dfrac{\dfrac{}{P\wedge Q\vdash P\wedge Q}(\mathrm{A})}{P\wedge Q\vdash P}(\wedge\mathrm{E1})}{P\wedge Q\vdash P\vee\neg Q}(\vee\mathrm{I1}) \quad \dfrac{\dfrac{\dfrac{}{P\wedge Q\vdash P\wedge Q}(\mathrm{A})}{P\wedge Q\vdash Q}(\wedge\mathrm{E2})}{P\wedge Q\vdash \neg P\vee Q}(\vee\mathrm{I2})}{P\wedge Q\vdash (P\vee Q)\wedge(P\vee\neg Q)\wedge(\neg P\vee Q)}$$

"Plugging in" the derivation from Example 3.16 yields

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{P\wedge Q\vdash P\wedge Q}(\mathrm{A})}{P\wedge Q\vdash P}(\wedge\mathrm{E1})}{P\wedge Q\vdash P\vee Q}(\vee\mathrm{I1}) \quad \dfrac{\dfrac{\dfrac{}{P\wedge Q\vdash P\wedge Q}(\mathrm{A})}{P\wedge Q\vdash P}(\wedge\mathrm{E1})}{P\wedge Q\vdash P\vee\neg Q}(\vee\mathrm{I1})}{P\wedge Q\vdash (P\vee Q)\wedge(P\vee\neg Q)}(\wedge\mathrm{I}) \quad \dfrac{\dfrac{\dfrac{}{P\wedge Q\vdash P\wedge Q}(\mathrm{A})}{P\wedge Q\vdash Q}(\wedge\mathrm{E2})}{P\wedge Q\vdash \neg P\vee Q}(\vee\mathrm{I2})}{P\wedge Q\vdash ((P\vee Q)\wedge(P\vee\neg Q))\wedge(\neg P\vee Q)}(\wedge\mathrm{I})$$

3.21. **Exercise.** 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 $(\vee\mathrm{E})$, which allows us to do casework depending on whether an arbitrary formula is true or false.

3.22. **Example.** 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":

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

The derivation of $(\to\mathrm{I})$ uses (LEM) from Exercise 3.21:

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

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

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

3.23. **Example.** Recalling Examples 3.10 and 3.12, we have

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

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

3.24. **Exercise.** The following **cut rule** is admissible:

$$(\textsc{Cut})\ \frac{\mathcal{T} \vdash \phi \quad \mathcal{T} \cup \{\phi\} \vdash \psi}{\mathcal{T} \vdash \psi}$$

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 $\mathcal{T}' = \{\phi\}$). [For the last part, use Proposition 3.26 below.]

3.25. **Exercise.** A rule

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

is called **invertible** if the **inverse rules**

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

are all admissible. For example, ($\wedge$I) is invertible, with the inverse rules being ($\wedge$E1) and ($\wedge$E2).

Intuitively, this means that if there is a deduction of the conclusion T, then there must be such a deduction which ends with the rule in question. Thus, when trying to prove T, it is always safe to take the rule as the last step. For example, when proving a $\wedge$, we may always apply ($\wedge$I) first.

Show that (A), ($\wedge$I), ($\top$I), ($\neg$I), ($\neg$E) and (C) are invertible, while the other basic inference rules are not invertible. [Use soundness, Proposition 3.27.] What about ($\to$I), ($\to$E), (Cut)?

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 reduce to finitely many assumptions.

3.26. **Proposition** (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

  $$(\wedge\text{I})\ \frac{\mathcal{T} \vdash \phi \quad \mathcal{T} \vdash \psi}{\mathcal{T} \vdash \phi \wedge \psi},$$

  then by the induction hypotheses (for the sub-deductions of $\mathcal{T} \vdash \phi$ and $\mathcal{T} \vdash \psi$), there are finite $\mathcal{T}_1 \subseteq \mathcal{T}$ and $\mathcal{T}_2 \subseteq \mathcal{T}$ such that

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

  By (W),

  $$\mathcal{T}_1 \cup \mathcal{T}_2 \vdash \phi, \qquad \mathcal{T}_1 \cup \mathcal{T}_2 \vdash \psi,$$

  whence by ($\wedge$I), $\mathcal{T}_1 \cup \mathcal{T}_2 \vdash \phi \wedge \psi$.
- If the deduction ends with

  $$(\vee\text{E})\ \frac{\mathcal{T} \vdash \phi \vee \psi \quad \mathcal{T} \cup \{\phi\} \vdash \theta \quad \mathcal{T} \cup \{\psi\} \vdash \theta}{\mathcal{T} \vdash \theta},$$

  then by the induction hypotheses (for 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 $\mathcal{T}_1 \subseteq \mathcal{T}'$, $\mathcal{T}_2 = \mathcal{T}_2 \cap (\mathcal{T} \cup \{\phi\}) \subseteq (\mathcal{T}_2 \cap \mathcal{T}) \cup \{\phi\} \subseteq \mathcal{T}' \cup \{\phi\}$, and similarly $\mathcal{T}_3 \subseteq \mathcal{T}' \cup \{\psi\}$, 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 ($\vee$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.C. **Soundness and completeness.** We have discussed both semantic and syntactic truth:

$$\mathcal{T} \models \phi \quad \text{vs} \quad \mathcal{T} \vdash \phi.$$

We now show that these are equivalent notions, i.e., "true" iff "provable".

3.27. **Proposition** (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 induct on $\mathcal{D}$. We need to write out most of the cases, because the argument is different for each rule: namely, in order to prove a rule is sound, we need to apply the corresponding *informal* principle of reasoning (in the metatheory) that the rule formalizes.

- If $\mathcal{D} = {}^{(\mathrm{A})}\dfrac{}{\mathcal{T} \vdash \phi}$, then $\phi \in \mathcal{T}$, so since $m \models \mathcal{T}$, we have $m \models \phi$.
- If $\mathcal{D}$ ends with
$$(\wedge\mathrm{I})\,\frac{\mathcal{T} \vdash \phi \quad \mathcal{T} \vdash \psi}{\mathcal{T} \vdash \phi \wedge \psi},$$
then by the induction hypotheses for 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 $(\wedge\mathrm{E2})$.
- 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 $(\vee\mathrm{I2})$.
- If $\mathcal{D}$ ends with
$$(\vee\mathrm{E})\,\frac{\mathcal{T} \vdash \phi \vee \psi \quad \mathcal{T} \cup \{\phi\} \vdash \theta \quad \mathcal{T} \cup \{\psi\} \vdash \theta}{\mathcal{T} \vdash \theta},$$
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 $(\top\mathrm{I})$, 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
$$(\neg\mathrm{E})\,\frac{\mathcal{T} \vdash \phi \quad \mathcal{T} \vdash \neg\phi,}{\mathcal{T} \vdash \bot}$$
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 $(\neg\mathrm{I})$ case. □

The converse direction is significantly harder:

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

Our proof strategy will be as follows. We prove the contrapositive:

$$\mathcal{T} \nvdash \phi \implies \mathcal{T} \not\models \phi \iff \exists m \models \mathcal{T} \text{ s.t. } m \not\models \phi.$$

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

$$m \models P \ :\iff \mathcal{T} \vdash P. \tag{3.29}$$

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 $\mathcal{T} \nvdash P$. Thus, we would like to know also

$$m \models \phi \iff \mathcal{T} \vdash \phi \tag{3.30}$$

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 $(\top\text{I})$.
- We have $m \not\models \bot$ always, and also $\mathcal{T} \nvdash \bot$, or else by $(\bot\text{E})$ we would have $\mathcal{T} \vdash \phi$ for all ϕ; but we are assuming $\mathcal{T} \nvdash \phi$ 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
$$\begin{aligned} m \models \neg\phi &\iff m \not\models \phi \\ &\iff \mathcal{T} \nvdash \phi && \text{by IH} \\ &\impliedby \mathcal{T} \vdash \neg\phi \end{aligned}$$
since if $\mathcal{T} \vdash \neg\phi$ and also $\mathcal{T} \vdash \phi$, then by $(\neg\text{E})$, $\mathcal{T} \vdash \bot$, which as we noted above is impossible if $\mathcal{T} \nvdash \phi$. 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 $\mathcal{T} \nvdash \bot$, or equivalently (by the $(\bot\text{E})$ rule), $\mathcal{T} \nvdash \phi$ for at least one formula ϕ;
- **complete**[15] if for all formulas ϕ, either $\mathcal{T} \vdash \phi$ or $\mathcal{T} \vdash \neg\phi$.

[15]This unfortunate terminology is distinct from the usage in the "completeness theorem". There is only a high-level conceptual similarity between the two notions: both say that a "syntax fully captures its intended semantics". The completeness theorem says that the deductive system for *propositional logic as a whole* captures its semantics *under all possible* 2*-valued semantics*; while completeness of a theory means *that particular set of axioms* captures the semantics *of a single model*.

3.31. **Lemma.** $\mathcal{T}$ is complete iff whenever $\mathcal{T} \vdash \phi \vee \psi$, then $\mathcal{T} \vdash \phi$ or $\mathcal{T} \vdash \psi$.

Proof. $\Longleftarrow$ is because $\mathcal{T} \vdash \phi \vee \neg\phi$ by (LEM). Conversely, if $\mathcal{T}$ is complete and $\mathcal{T} \nvdash \phi$, then $\mathcal{T} \vdash \neg\phi$ by completeness, whence $\mathcal{T} \vdash \psi$ by the deduction

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

□

3.32. **Lemma.** Let $\mathcal{T}$ be a theory, and let $m : \mathcal{A} \to \{0, 1\}$ be defined as in (3.29)

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

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

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

Proof. By the induction argument sketched above. □

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

3.33. **Lemma.** 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 3.33, there is a complete theory $\mathcal{T}' \supseteq \mathcal{T}$ such that $\mathcal{T}' \nvdash \phi$; thus $\mathcal{T}'$ is also consistent. By Lemma 3.32, 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 3.33 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.

3.34. **Lemma.** 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

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

□

3.35. **Lemma.** Let $\mathcal{T}_0 \subseteq \mathcal{T}_1 \subseteq \cdots$ be 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 3.26), there are finitely many formulas $\psi_1, \ldots, \psi_k \in \bigcup_n \mathcal{T}_n$ such that $\{\psi_1, \ldots, \psi_k\} \vdash \phi$. Let n be large enough so that $\phi_1, \ldots, \phi_k \in \mathcal{T}_n$. Then by weakening, $\mathcal{T}_n \vdash \phi$, a contradiction. □

Proof of Lemma 3.33. First, assume that $\mathcal{A}$ is countable, and so we can enumerate all $\mathcal{A}$-formulas

$$\mathcal{L}(\mathcal{A}) = \{\psi_n\}_{n\in\mathbb{N}} = \{\psi_0, \psi_1, \psi_2, \dots\}.$$

Define an increasing sequence of theories $\mathcal{T}_0 \subseteq \mathcal{T}_1 \subseteq \mathcal{T}_2 \subseteq \cdots$ inductively as follows:

$$\mathcal{T}_0 := \mathcal{T},$$

$$(*) \qquad \mathcal{T}_{n+1} := \begin{cases} \mathcal{T}_n \cup \{\psi_n\} & \text{if } \mathcal{T}_n \cup \{\psi_n\} \nvdash \phi, \\ \mathcal{T}_n \cup \{\neg\psi_n\} & \text{otherwise.} \end{cases}$$

By Lemma 3.34, in the second case we have $\mathcal{T}_n \cup \{\neg\psi_n\} \nvdash \phi$; thus by induction, each $\mathcal{T}_n \nvdash \phi$. Now

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

is complete, since any $\psi \in \mathcal{L}(\mathcal{A})$ is $\psi = \psi_n$ 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 $\mathcal{T}' \nvdash \phi$ by Lemma 3.35, so we're done.

In the uncountable case, by the well-ordering theorem, we may still transfinitely enumerate

$$\mathcal{L}(\mathcal{A}) = \{\psi_0, \psi_1, \dots, \psi_\omega, \psi_{\omega+1}, \dots\} = \{\psi_n \mid n < \alpha\}$$

for an ordinal α. We inductively define a transfinite sequence of theories $(\mathcal{T}_n)_{n\leq\alpha}$ by $(*)$ above and

$$\mathcal{T}_n := \bigcup_{k<n} \mathcal{T}_k \quad \text{for limit ordinals } n \leq \alpha.$$

We now prove by induction as above that each $\mathcal{T}_n \nvdash \phi$; for n a limit ordinal, this is by Lemma 3.35 which works just as well for transfinite sequences. Now letting

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

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

(Instead of transfinite induction, we can also use Zorn's lemma to find a maximal $\mathcal{T}' \supseteq \mathcal{T}$ such that $\mathcal{T}' \nvdash \phi$. An analogue of Lemma 3.35 is used to verify that the assumptions of Zorn's lemma are satisfied, while Lemma 3.34 is used to verify that such a maximal $\mathcal{T}'$ must be complete.)[16] □

This concludes the proof of the Completeness Theorem 3.28.

3.36. **Corollary** (of soundness and completeness)**.**

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

In other words, the closure $\mathrm{Th}(\mathrm{Mod}(\mathcal{T}))$ of $\mathcal{T}$ under semantic consequences from Definition 2.29 consists of precisely the *provable* consequences of $\mathcal{T}$, i.e., we close $\mathcal{T}$ under consequences by adding to it all formulas that can be deduced from it in finitely many steps via the inference rules.

3.37. **Example.** The law $P \wedge (Q \vee R) \vdash (P \wedge Q) \vee R$ that we gave a syntactic proof of in Example 3.9 follows from the corresponding semantic implication we verified via truth table in Example 2.9.

Special cases of Corollary 3.36 say that various other syntactic and semantic notions coincide:

3.38. **Corollary.** A formula ϕ is a provable tautology iff it is a semantic tautology:

$$\vdash \phi \iff \models \phi.$$

3.39. **Corollary.** A theory $\mathcal{T}$ is consistent iff it is satisfiable:

$$\mathcal{T} \nvdash \bot \iff \mathcal{T} \not\models \bot \iff \mathrm{Mod}(\mathcal{T}) \neq \varnothing.$$

3.40. **Exercise.** A theory $\mathcal{T}$ is complete iff it has at most one model.

3.41. **Corollary.** A theory $\mathcal{T}$ is consistent and complete iff it has a unique model.

3.42. **Exercise.** Verify that in this case, the unique model m is that defined in Lemma 3.32.

[16] It is impossible to prove the completeness theorem without the Axiom of Choice in some form; see Theorem 4.7.

4. Compactness

By syntactic compactness (Proposition 3.26) and Corollary 3.36, we have

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

4.2. **Corollary.** If every finite $\mathcal{T}' \subseteq \mathcal{T}$ is satisfiable, then so is $\mathcal{T}$.

Proof. Take $\phi := \bot$. □

4.3. **Corollary.** Let $(\mathcal{T}_i)_{i\in I}$ be a family of theories. Suppose that the union of any finitely many $\mathcal{T}_{i_1} \cup \cdots \cup \mathcal{T}_{i_n}$ is satisfiable, for any $i_1, \ldots, i_n \in I$. Then $\bigcup_{i\in I} \mathcal{T}_i$ is satisfiable.

Proof. Each finite $\mathcal{T}' \subseteq \bigcup_i \mathcal{T}_i$ is contained in a finite union. □

4.4. **Corollary.** Let $(\mathcal{K}_i)_{i\in I}$ be a family of axiomatizable sets of truth assignments $\mathcal{K}_i \subseteq \{0,1\}^{\mathcal{A}}$. Suppose that the intersection of any finitely many $\mathcal{K}_{i_1} \cap \cdots \cap \mathcal{K}_{i_n}$ is nonempty. Then $\bigcap_{i\in I} \mathcal{K}_i \neq \varnothing$.

Proof. Let $\mathcal{K}_i = \mathrm{Mod}(\mathcal{T}_i)$, and apply the preceding corollary, using Exercise 2.21. □

These results are usually simply called the **Compactness Theorem** (for propositional logic), and are much deeper than the corresponding syntactic Proposition 3.26. They say that in order to show that a certain type of mathematical object exists, *if it can be described in propositional logic*, then it suffices to check only finitely many conditions at once. Here is a typical application:

4.5. **Example.** For any set X, recall from Example 2.47 that we can make a theory $\mathcal{T}$ (over the alphabet $\mathcal{A} := \{P_{x,y} \mid x, y \in X\}$) whose models correspond to linear orders on X. Now every finite $\mathcal{T}' \subseteq \mathcal{T}$ is easily seen to be satisfiable: if $X' = \{x_1, \ldots, x_n\} \in X$ are all finitely many $x, y \in X$ for $P_{x,y}$ appearing in a formula in $\mathcal{T}'$, then we may define a linear order $\leq$ on X', namely $x_1 < x_2 < \cdots < x_n$, and then regard this as a binary relation on $X \supseteq X'$, which is no longer a linear order, but $\mathcal{T}'$ can't tell since it only mentions the elements of X'. Thus, *every set has a linear order.*

4.6. **Exercise.** Show that more generally, every partial order $\leq$ extends to a linear order $\leq'$.

But even if something exists, that doesn't mean you can actually get your hands on one:

4.7. **Theorem.** It is impossible to write down an example of a linear order on $\mathcal{P}(\mathbb{R})$.[17]

Very rough proof sketch. Let $\sim$ be the equivalence relation on $\mathbb{R}$ defined by

$$x \sim y \;:\Longleftrightarrow\; x - y \in \mathbb{Q}.$$

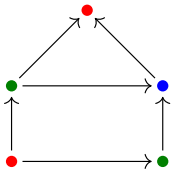
Note that $x \sim -x \iff 2x \in \mathbb{Q} \iff x \in \mathbb{Q}$; thus for every equivalence class $[x] \in \mathbb{R}/\!\sim$ other than $[0] = \mathbb{Q}$, we have $[x] \cap [-x] = \varnothing$. Now if $\leq$ were an example of a linear order on $\mathcal{P}(\mathbb{R})$, then

$$A := \{x \in \mathbb{R} \mid [x] < [-x]\}$$

together with $-A$ would form a partition of the irrationals $\mathbb{R} \setminus \mathbb{Q}$. Also, the negation map $- : \mathbb{R} \to \mathbb{R}$ interchanges A with $-A$, and clearly also maps each interval to another interval of the same length. But, it turns out that for every set $A \subseteq \mathbb{R}$ that you can actually write down, and which is also invariant under the equivalence relation $\sim$, either A or $\mathbb{R} \setminus A$ must have "length 0" in the sense that it is contained in a countable union of intervals with total length $< \varepsilon$ for any $\varepsilon > 0$. (For example, $\mathbb{Q}$ has length 0, since we can cover the first rational with an interval of length $\varepsilon/2$, then the second with an interval of length $\varepsilon/4$, etc.) So one of A or $-A = \mathbb{R} \setminus \mathbb{Q} \setminus A \subseteq \mathbb{R} \setminus A$ has length 0, hence both do since $-$ preserves lengths, hence $\mathbb{R} = A \cup -A \cup \mathbb{Q}$ has length 0, which is absurd. □

[17] The reality is even more confusing: (a) for any partial order $\leq$ on $\mathcal{P}(\mathbb{R})$ that we can write down, it is impossible to prove that it is a linear order; and (b) we can write down a specific example of a partial order on $\mathcal{P}(\mathbb{R})$, for which it is impossible to *disprove* that it is a linear order. So it is impossible to write down an example of a linear order on $\mathcal{P}(\mathbb{R})$, but you cannot prove that this is impossible! To find out what all of this means, take a course in set theory.

Here is another application. Recall from Definition 2.40 that a **graph** G on the set of vertices X is an irreflexive, symmetric relation $G \subseteq X^2$. For $k \in \mathbb{N}$, a **k-coloring** of G is a function $c : X \to \{1, \dots, k\}$ such that whenever $x\ G\ y$, then $c(x) \neq c(y)$. Here is a 3-coloring:



($\bullet = 1$, $\bullet = 2$, $\bullet = 3$)

4.8. **Theorem** (de Bruijn–Erdős)**.** A graph (X, G) has a k-coloring iff every finite subgraph does.

Proof. We can represent a k-coloring $c : X \to \{1, \dots, k\}$ as k many unary relations $C_i \subseteq X$, where $C_i = c^{-1}(i)$ is the set of vertices with color i, obeying the axioms

$$\forall x \in X\, (x \in C_i \text{ and } x \in C_j \implies i = j),$$
$$\forall x \in X\, (x \in C_1 \text{ or } \cdots \text{ or } x \in C_k),$$
$$\forall (x, y) \in G\ \neg(x \in C_i \text{ and } y \in C_i).$$

To describe this in propositional logic as in Section 2.B, let $\mathcal{A} := \{P_{x,i} \mid x \in X \text{ and } i \in \{1, \dots, k\}\}$, where $P_{x,i}$ denotes "$x \in C_i$" (i.e., "x has color i"), and let

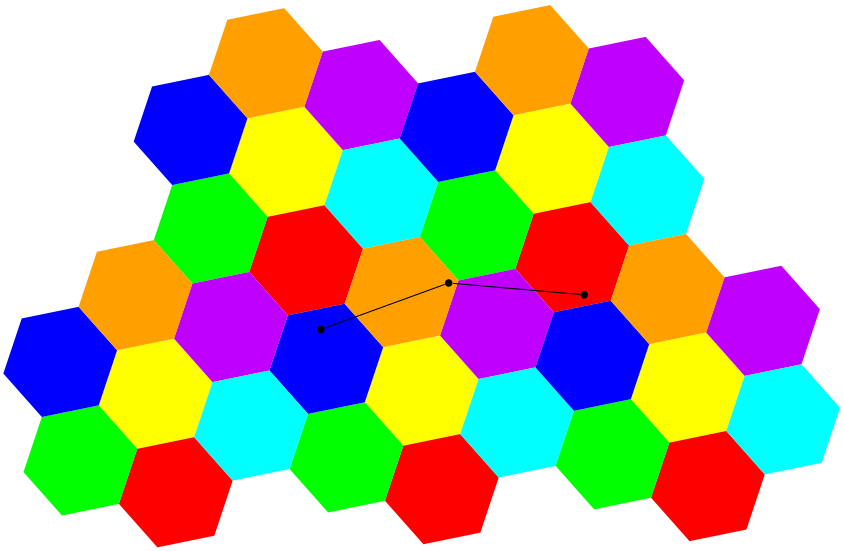
$$\begin{aligned}\mathcal{T} := &\{\neg(P_{x,i} \wedge P_{x,j}) \mid x \in X \text{ and } i \neq j \in \{1, \dots, k\}\} \\ &\cup \{P_{x,1} \vee \cdots \vee P_{x,k} \mid x \in X\} \\ &\cup \{\neg(P_{x,i} \wedge P_{y,i})\}(x, y) \in G \text{ and } i \in \{1, \dots, k\}.\end{aligned}$$

Then each model $m \models \mathcal{T}$ corresponds to the coloring $c : X \to \{1, \dots, k\}$ where $c(x)$ is the unique i (by the first two sets of axioms) such that $m \models P_{x,i}$. Now every finite $\mathcal{T}' \subseteq \mathcal{T}$ has a model, since we may take a coloring $c : X' \to \{1, \dots, k\}$ on the finite subgraph consisting of all finitely many vertices $X' \subseteq X$ mentioned in $\mathcal{T}'$, and extend it arbitrarily to $c : X \to \{1, \dots, k\}$, which won't be a coloring anymore, but $\mathcal{T}'$ can't tell. Thus by compactness, $\mathcal{T}$ has a model. □

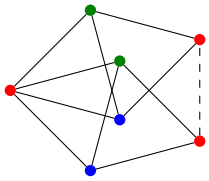
4.9. **Example.** The **unit distance graph** G on $\mathbb{R}^2$ has edges

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

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



The 70-year-old open **Hadwiger–Nelson problem**[18] asks 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:

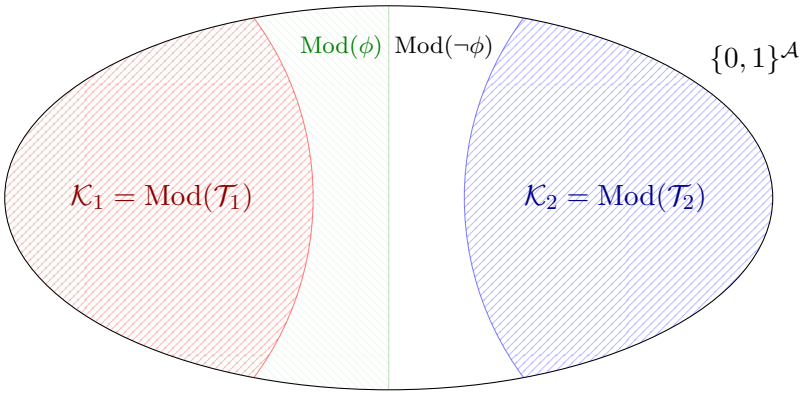


This was the best known lower bound until recently, when de Grey (2018) found a 1581-point configuration which cannot be 4-colored. So it is now known that 5, 6, or 7 colors are needed.

In fact, before de Grey's result, it was already known (Falconer 1981) that it is impossible to *write down* a 4-coloring of the plane. The reason is somewhat analogous to that in Theorem 4.7: in any 4-coloring you can write down, you can reason about the "areas" of the different-colored pieces in order to reach a contradiction.

4.A. **Finite axiomatizability and Stone duality.** Here is an important but more theoretical application of the compactness theorem:

4.10. **Theorem** (separation)**.** Let $\mathcal{K}_1 = \mathrm{Mod}(\mathcal{T}_1)$ and $\mathcal{K}_2 = \mathrm{Mod}(\mathcal{T}_2)$ be disjoint axiomatizable sets. Then there is a formula ϕ such that $\mathcal{T}_1 \models \phi$ and $\mathcal{T}_2 \models \neg\phi$, i.e., $\mathcal{K}_1 \subseteq \mathrm{Mod}(\phi)$ and $\mathcal{K}_2 \cap \mathrm{Mod}(\phi) = \varnothing$.



Proof. First, consider the special case where $\mathcal{T}_1 = \{\phi_1, \dots, \phi_n\}$ is finite. Then $\phi := \bigwedge \mathcal{T}_1 = \phi_1 \wedge \cdots \wedge \phi_n$ works: we have $\mathrm{Mod}(\phi) = \mathrm{Mod}(\mathcal{T}_1) = \mathcal{K}_1$, which is disjoint from $\mathcal{K}_2$ by assumption.

In the general case, since $\varnothing = \mathcal{K}_1 \cap \mathcal{K}_2 = \mathrm{Mod}(\mathcal{T}_1 \cup \mathcal{T}_2)$, by compactness, some finite $\mathcal{T}' \subseteq \mathcal{T}_1 \cup \mathcal{T}_2$ is already unsatisfiable; replace $\mathcal{T}_1$ with $\mathcal{T}_1 \cap \mathcal{T}'$ in the previous case. □

4.11. **Corollary.** $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is axiomatizable by a single formula iff both $\mathcal{K}$ and its complement are axiomatizable.

Proof. $\Longrightarrow$: If $\mathcal{K} = \mathrm{Mod}(\phi)$, then $\{0,1\}^{\mathcal{A}} \setminus \mathcal{K} = \mathrm{Mod}(\neg\phi)$.
$\Longleftarrow$: Apply the preceding theorem to $\mathcal{K}, \{0,1\}^{\mathcal{A}} \setminus \mathcal{K}$. □

Recall from Corollary 2.58 that $\mathcal{K}$ is axiomatizable iff it is closed under limits. We call $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ **clopen** if both it and its complement are closed under limits.

4.12. **Corollary.** $\mathcal{K} \subseteq \{0,1\}^{\mathcal{A}}$ is axiomatizable by a single formula iff it is clopen. □

4.13. **Exercise.** What are the clopen sets in $\mathbb{R}$ (with respect to the usual notion of limit)?

[18] For detailed history of this problem, see A. Soifer, *The Mathematical Coloring Book*, Springer, 2008.

In other words, Corollary 4.12 says that the image of the function

$$\mathrm{Mod} : \mathcal{L}(\mathcal{A}) \longrightarrow \mathcal{P}(\{0,1\}^{\mathcal{A}})$$

consists of precisely the clopen sets $\mathcal{K}$. We may strengthen this statement by combining it with the completeness theorem, as follows.

4.14. **Lemma.** $\vdash$ is a preorder on $\mathcal{L}(\mathcal{A})$.

Proof. We already know from Example 2.43 that $\models$ is trivially a preorder, and from soundness and completeness that $\vdash$ is the same as $\models$. However, since we wish to reformulate the completeness theorem, we give a purely syntactic proof. Reflexivity follows immediately from the (A) rule. For transitivity, we use the admissible (in fact derivable) (Cut) rule (Exercise 3.24):

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

Take $\mathcal{T} := \varnothing$. (The version for general $\mathcal{T}$ will be used later; see Exercise 4.19.) □

4.15. **Corollary.** $\dashv\vdash$ is an equivalence relation on $\mathcal{L}(\mathcal{A})$. □

4.16. **Definition.** The quotient set $\mathcal{L}(\mathcal{A})/\dashv\vdash$ of provable equivalence classes of formulas is called the **Lindenbaum–Tarski algebra** of the alphabet $\mathcal{A}$. The preorder $\vdash$ on $\mathcal{L}(\mathcal{A})$ descends to a partial order on $\mathcal{L}(\mathcal{A})/\dashv\vdash$:

$$[\phi] \leq [\psi] \ :\Longleftrightarrow\ \phi \vdash \psi.$$

(Transitivity ensures that this is well-defined, i.e., does not depend on the representatives ϕ, ψ.)

The term "algebra" is explained by

4.17. **Exercise.** Show that the following rules are admissible:

$$\frac{\mathcal{T} \cup \{\phi_1\} \vdash \psi_1 \quad \mathcal{T} \cup \{\phi_2\} \vdash \psi_2}{\mathcal{T} \cup \{\phi_1 \wedge \phi_2\} \vdash \psi_1 \wedge \psi_2} \qquad \frac{\mathcal{T} \cup \{\phi_1\} \vdash \psi_1 \quad \mathcal{T} \cup \{\phi_2\} \vdash \psi_2}{\mathcal{T} \cup \{\phi_1 \vee \phi_2\} \vdash \psi_1 \vee \psi_2} \qquad \frac{\mathcal{T} \cup \{\phi\} \vdash \psi}{\mathcal{T} \cup \{\neg\psi\} \vdash \neg\phi}$$

Thus (for $\mathcal{T} := \varnothing$)

$$\begin{aligned} \phi_1 \dashv\vdash \psi_1 \text{ and } \phi_2 \dashv\vdash \psi_2 &\implies \phi_1 \wedge \phi_2 \dashv\vdash \psi_1 \wedge \psi_2, \\ \phi_1 \dashv\vdash \psi_1 \text{ and } \phi_2 \dashv\vdash \psi_2 &\implies \phi_1 \vee \phi_2 \dashv\vdash \psi_1 \vee \psi_2, \\ \phi \dashv\vdash \psi &\implies \neg\phi \dashv\vdash \neg\psi, \end{aligned}$$

and so the operations $\wedge, \vee, \neg$ on $\mathcal{L}(\mathcal{A})$ descend to the quotient set $\mathcal{L}(\mathcal{A})/\dashv\vdash$, i.e.,

$$[\phi] \wedge [\psi] := [\phi \wedge \psi], \qquad [\phi] \vee [\psi] := [\phi \vee \psi], \qquad \neg[\phi] := [\neg\phi]$$

are well-defined operations on provable equivalence classes of formulas.

Note that by Exercise 3.11, (¬E), and (LEM), these operations obey the laws

$$\begin{array}{lll} \text{(commutativity)} & [\phi] \wedge [\psi] = [\psi] \wedge [\phi], & [\phi] \vee [\psi] = [\psi] \vee [\phi], \\ \text{(associativity)} & ([\phi] \wedge [\psi]) \wedge [\theta] = [\phi] \wedge ([\psi] \wedge [\theta]), & ([\phi] \vee [\psi]) \vee [\theta] = [\phi] \vee ([\psi] \vee [\theta]), \\ \text{(idempotence)} & [\phi] \wedge [\phi] = [\phi], & [\phi] \vee [\psi] = [\phi], \\ \text{(distributivity)} & [\phi] \wedge ([\psi] \vee [\theta]) = ([\phi] \wedge [\psi]) \vee ([\phi] \wedge [\theta]), & [\phi] \vee ([\psi] \wedge [\theta]) = ([\phi] \vee [\psi]) \wedge ([\phi] \vee [\theta]), \\ \text{(complements)} & [\phi] \wedge \neg[\phi] = [\bot], & [\phi] \vee \neg[\phi] = [\top]. \end{array}$$

So the "familiar identities" like commutativity you've probably wanted to write all this time actually become true in $\mathcal{L}(\mathcal{A})/\dashv\vdash$! We express these identities by saying that $\mathcal{L}(\mathcal{A})/\dashv\vdash$ is a **Boolean algebra**; this is a type of abstract algebraic structure, analogous to a group, ring, vector space, etc.

4.18. **Theorem** (soundness + completeness + axiomatizability)**.** The function

$$\mathrm{Mod} : \mathcal{L}(\mathcal{A}) \to \mathcal{P}(\{0,1\}^{\mathcal{A}})$$

descends to an isomorphism of Boolean algebras

$$\mathrm{Mod} : \mathcal{L}(\mathcal{A})/\dashv\vdash \;\cong \mathrm{Clopen}(\{0,1\}^{\mathcal{A}}) \subseteq \mathcal{P}(\{0,1\}^{\mathcal{A}})$$

between provable equivalence classes of formulas and clopen sets of truth assignments, that preserves order both ways and takes the operations $\wedge, \vee, \neg, \top, \bot$ to set operations $\cap, \cup, \neg, \{0,1\}^{\mathcal{A}}, \varnothing$.

Proof. By soundness, $\phi \vdash \psi \implies \mathrm{Mod}(\phi) \subseteq \mathrm{Mod}(\psi)$, which implies in particular that $\phi \dashv\vdash \psi \implies \mathrm{Mod}(\phi) = \mathrm{Mod}(\psi)$, whence Mod is order-preserving and descends to the quotient. By completeness, $\mathrm{Mod}(\phi) \subseteq \mathrm{Mod}(\psi) \implies \phi \vdash \psi \iff [\phi] \leq [\psi]$, whence the descended Mod is injective and its inverse is also order-preserving. Its image consists precisely of the clopen sets by Corollary 4.12. Finally, the Boolean operations correspond to each other by Definition 2.1 of $\models$. □

This further clarifies the picture (2.20) of the relationship between syntax and semantics, by showing that the right, syntactic, half of the picture is "equivalent" modulo the equivalence relation $\dashv\vdash$ (which is also defined syntactically) to the clopen sets in the left, semantic, space of models.

4.19. **Exercise.** There is also a "relative" form of the above correspondence, where on the left semantic side we restrict to models of a fixed theory $\mathcal{T}$, and on the right syntactic side we do everything modulo the theory.

(a) Verify that

$$\phi \vdash^{\mathcal{T}} \psi \;:\iff\; \mathcal{T} \cup \{\phi\} \vdash \psi,$$

pronounced "ϕ proves ψ **modulo** $\mathcal{T}$" (people also say "**under** $\mathcal{T}$"), likewise defines a preorder on $\mathcal{L}(\mathcal{A})$. [See Lemma 4.14.] Thus, its symmetric part $\dashv\vdash^{\mathcal{T}}$ defines an equivalence relation; the quotient $\mathcal{L}(\mathcal{A})/\dashv\vdash^{\mathcal{T}}$ is called the **Lindenbaum–Tarski algebra of** $\mathcal{T}$.

(b) Verify that the logical connectives on $\mathcal{L}(\mathcal{A})$ likewise respect this equivalence relation, hence descend to operations on the quotient. [See Exercise 4.17.] Moreover, they again make $\mathcal{L}(\mathcal{A})/\dashv\vdash^{\mathcal{T}}$ into a Boolean algebra.

(c) Show that we have an isomorphism

$$\mathcal{L}(\mathcal{A})/\dashv\vdash^{\mathcal{T}} \;\cong \mathrm{Clopen}(\mathrm{Mod}(\mathcal{T})) \subseteq \mathcal{P}(\mathrm{Mod}(\mathcal{T}))$$
$$[\phi] \mapsto \mathrm{Mod}(\mathcal{T} \cup \{\phi\}) = \mathrm{Mod}(\mathcal{T}) \cap \mathrm{Mod}(\phi).$$

(A **clopen set in** $\mathrm{Mod}(\mathcal{T})$ means a set such that it and its relative complement inside $\mathrm{Mod}(\mathcal{T})$ are closed under limits.)

(d) Show that the existence of this isomorphism is equivalent to the soundness (Proposition 3.27), completeness (Theorem 3.28), and axiomatizability (Corollary 4.12) theorems. That is, show that these three theorems follow easily from the above isomorphism.

The above results are (one formulation of what's) known as the **Stone representation theorem**. In general, a *representation theorem* in math shows that an abstract structure of some kind can be "represented" or "understood" in terms of more familiar structures. In this case, we get that the syntactic structure of propositional logic ($\vdash, \wedge, \vee, \ldots$) can be understood in terms of the familiar structure of set operations ($\subseteq, \cap, \cup, \ldots$). As part (d) above shows, in some sense the Stone representation theorem "fully captures" the main mathematical content of propositional logic, by encapsulating the three major theorems (soundness, completeness, axiomatizability) into one single structural isomorphism between the syntax and semantics of the logic.