ELEMENTS OF MATHEMATICAL LOGIC

MICHAEL LIEBERMAN

Contents

1. Semantics of Predicate Logic

Recall that a first-order signature L consists of

- A countably infinite set of variables, $X = \{x_0, x_1, \dots, x_n, \dots\}$.
- A set of constant symbols, $\mathcal{C}$.
- A set of function symbols, $\mathcal{F}$.
- A set of predicate symbols, $\mathcal{R}$.

Example 1.1.
(1) Ordered sets: the natural signature L_{ord} has $\mathcal{R} = \{R\}$, and $\mathcal{C} = \emptyset = \mathcal{F}$.
(2) Groups: the natural signature L_{grp} has $\mathcal{R} = \emptyset$, $\mathcal{C} = \{e\}$, and $\mathcal{F} = \{m\}$.
(3) Natural numbers with successor: the natural signature L_{succ} has $R = \{<\}$, $\mathcal{C} = \{0\}$, $\mathcal{F} = \{S\}$.
(4) Elementary number theory: the natural signature has $\mathcal{R} = \{<\}$, $\mathcal{C} = \{0, 1\}$, and $\mathcal{F} = \{\cdot, +, S\}$.

As mentioned before, we should think of these as formal symbols, and not as the constants, operations, and relations that they typically represent—we will speak about such interpretations in a moment. Often this distinction is enforced through careful notation, but this gets tiresome, and typically leads to more confusion, rather than less.

Date: March 17, 2021.

From the raw ingredients in a signature L, we build:

- Tm_L, the *terms* of L.
- AtFm_L, the *atomic formulas* of L.
- QF_L, the *quantifier-free* (or, more simply, *q.f.*) formulas of L.
- Fm_L, the formulas of L.

1.1. **L-Structures and Interpretation.** Up to this point, we have worked with these symbols in a purely formal way, combining them in specific allowable ways without considering what they actually mean. We now address the latter question, focusing on their *interpretations*: as actual constants, functions, and relations in the everyday world of mathematical objects.

First, we make precise what it means for a mathematical object to interpret the basic symbols in a given signature.

Definition 1.2. Let L be a signature. A *structure for L* (or an L-structure) $\mathcal{M}$ consists of

(1) A set M, known as the *underlying set* or *universe* of $\mathcal{M}$.
(2) For each $c \in \mathcal{C}$, an element $c^{\mathcal{M}} \in M$.
(3) For each $f \in \mathcal{F}$ of arity n, an n-ary function
$$f^{\mathcal{M}} : M \times \cdots \times M \to M,$$
where the domain is an n-fold product of copies of M.
(4) For each $R \in \mathcal{R}$ of arity n, a subset
$$R^{\mathcal{M}} \subseteq M \times \cdots \times M$$
where the set on the right is an n-fold product of copies of M. In general, instead of writing $(m_1, \ldots, m_n) \in R^{\mathcal{M}}$ to signify that the tuple of elements $m_1, \ldots, m_n$ are $R^{\mathcal{M}}$-related, we will simply write $R^{\mathcal{M}} m_1 \ldots m_n$.

Example 1.3. Take $L = L_{succ}$, the signature of the natural numbers with successor: $\mathcal{C} = \{0\}$, $\mathcal{F} = \{S\}$, and $\mathcal{R} = \{<\}$ (more compactly, we often denote this and other signatures simply by the ordered tuple of their symbols; that is, $L_{succ} = \langle 0, S, < \rangle$). We consider several different L_{succ}-structures.

(1) The standard natural numbers are, of course, an L_{succ}-structure, $\mathcal{M}_1$:
   - $M_1 = \mathbb{N}$.
   - $0^{\mathcal{M}_1} = 0$.
   - $S^{\mathcal{M}_1}$ is the usual successor function on the natural numbers; that is, for any $n \in \mathbb{N}$,
   $$S^{\mathcal{M}_1}(n) = n + 1.$$
   - $<^{\mathcal{M}}$ is the standard order relation on the natural numbers; that is,
   $$<^{\mathcal{M}_1} = \{(m, n) \in \mathbb{N}^2 \mid m < n\}.$$
(2) There are all kinds of other variations, though, say $\mathcal{M}_2$ with
   - $M_2 = \mathbb{N}$.
   - $0^{\mathcal{M}_2} = 7$.
   - $S^{\mathcal{M}_2}$ is the function that takes each $n \in \mathbb{N}$ to $2n$.
   - $<^{\mathcal{M}_2} = \{(m, n) \in \mathbb{N}^2 \mid mn > 10\}$.

Clearly one of these structures inteprets the symbols in a reasonable, mathematically useful way, while the other does not. One of our major goals moving forward will be to use formulas to force proper behavior in the structures we consider.

First, we must extend our interpretation of the basic symbols to arbitrary terms.

Definition 1.4. A valuation of the variables of a signature L in an L-structure $\mathcal{M}$ is a map

$$\begin{aligned} v : X &\to M \\ x &\mapsto a \in M. \end{aligned}$$

Definition 1.5. Given an L-structure $\mathcal{M}$ and valuation v, we denote by $\tau[v]$ the value of the term τ in $\mathcal{M}$ under assignment v. This value is determined recursively, according to the following rules:

(1) If τ is a variable x, $\tau[v] = v(x)$.
(2) If $c \in \mathcal{C}$, $c[v] = c^{\mathcal{M}}$.
(3) If $f \in \mathcal{F}$ is n-ary, and $\tau_1, \ldots, \tau_n$ are terms,

$$(f\tau_1 \ldots \tau_n)[v] = f^{\mathcal{M}}(\tau_1[v], \ldots, \tau_n[v]).$$

This, of course, is the inductive step.

Example 1.6. Take $L = L_{grp}$. Consider the L-structure $\mathcal{M}$ given by $M = \mathbb{Z}$, $e^{\mathcal{M}} = 0$, and $*^{\mathcal{M}} = +$, with assignment $v : X \to M$ given by $v(x_n) = n$. Let $\tau = m(x_1, m(x_2, e))$. Then

$$\begin{aligned} \tau[v] &= m(x_1, m(x_2, e))[v] \\ &= m^{\mathcal{M}}(x_1[v], m(x_2, e)[v]) \\ &= m^{\mathcal{M}}(1, m^{\mathcal{M}}(x_2[v], e[v])) \\ &= m^{\mathcal{M}}(1, m^{\mathcal{M}}(2, 0)) \\ &= (1 + (2 + 0)) \\ &= 3. \end{aligned}$$

Note that the value of a term is highly sensitive to the structure and the valuation, in general:

Example 1.7. Take $L = L_{grp}$, as before, but consider the L-structure $\mathcal{M}_1$ with $M = \mathbb{Z}$, $m^{\mathcal{M}_1}$ the standard multiplication on $\mathbb{Z}$, and $e^{\mathcal{M}_1} = 7$. Let v be the constant valuation with $v(x_i) = 1$ for every $x_i \in X$. Then

$$\begin{aligned} \tau[v] &= m(x_1, m(x_2, e))[v] \\ &= m^{\mathcal{M}_1}(x_1[v], m(x_2, e)[v]) \\ &= m^{\mathcal{M}_1}(1, m^{\mathcal{M}_1}(x_2[v], e[v])) \\ &= m^{\mathcal{M}_1}(1, m^{\mathcal{M}_1}(1, 7)) \\ &= (1 \cdot (1 \cdot 7)) \\ &= 7. \end{aligned}$$

Remark 1.8. If we fix $\mathcal{M}$ and v, the value of τ in $\mathcal{M}$ under the valuation v, $\tau[v]$, is determined entirely by the value of v on variables that occur in τ. That is, if $\tau = \tau(x_1, \ldots, x_n)$ contains only the variables $x_1, \ldots, x_n$, then if valuations v and v' satisfy $v(x_i) = v'(x_i)$ for $i = 1, \ldots, n$, then $\tau[v] = \tau[v']$.

Now that we can interpret terms, we can start to make sense of the truth of formulas, which consist, roughly speaking, of equations or relations between terms, as we know.

Definition 1.9. Let $\mathcal{M}$ be an L-structure, v a valuation, and $\phi \in \mathrm{Fm}_L$. We define what it means for ϕ to be satisfied in $\mathcal{M}$ under the valuation v[1], denoted

$$\mathcal{M}, v \models \phi$$

by induction on the complexity of ϕ.

(1) If $\phi \in \mathrm{AtFm}_L$, there are two possibilities:
   (a) ϕ is $\tau_1 = \tau_2$ for some $\tau_1, \tau_2 \in \mathrm{Tm}_L$. Then
   $$\mathcal{M}, v \models \phi \text{ if and only if } \tau_1[v] = \tau_2[v].$$
   (b) ϕ is $R\tau_1, \dots \tau_n$, where $R \in \mathcal{R}$ and $\tau_1, \dots \tau_n \in \mathrm{Tm}_L$. Then
   $$\mathcal{M}, v \models \phi \text{ if and only if } (\tau_1[v], \dots, \tau_n[v]) \in R^{\mathcal{M}}.$$

(2) Propositional connectives:
   (a) Say $\phi = \neg\psi$. Then
   $$\mathcal{M}, v \models \phi \text{ if and only if } \mathcal{M}, v \not\models \psi.$$
   (b) Say $\phi = \psi \vee \chi$. Then
   $$\mathcal{M}, v \models \phi \text{ if and only if either } \mathcal{M}, v \models \psi \textbf{ or } \mathcal{M}, v \models \chi.$$
   (c) Say $\phi = \psi \wedge \chi$. Then
   $$\mathcal{M}, v \models \phi \text{ if and only if } \mathcal{M}, v \models \psi \textbf{ and } \mathcal{M}, v \models \chi.$$
   (d) Say $\phi = \psi \to \chi$. Then
   $$\mathcal{M}, v \models \phi \text{ if and only if } \mathcal{M}, v \not\models \psi \text{ or } \mathcal{M}, v \models \chi,$$
   where we have used the fact that $\psi \to \chi$ is tautologically equivalent to $\neg\psi \vee \chi$, together with earlier clauses in this definition.
   (e) If $\phi = \psi \leftrightarrow \chi$, we use the fact that $\psi \leftrightarrow \chi$ is tautologically equivalent to $\psi \to \chi \wedge \psi \leftarrow \chi$, and earlier clauses in this definition.

   Note: since $\{\vee, \neg\}$ is a base of connectives, we could have made do with just the clauses (a) and (b).

(3) Quantifiers: here we require a bit of extra notation. Given a valuation $v : X \to M$, variable x and $a \in M$, we denote by v(x/a) the following valuation:
$$v(x/a)(y) = \begin{cases} v(y) & \text{if } x \neq y \\ a & \text{if } x = y \end{cases}$$
That is, $v(x/a)$ is the same as v, except that we adjust it so that the variable x is assigned value a. We now turn to our cases:
   (a) Say $\phi = \forall x\psi$ for some ψ. Then $\mathcal{M}, v \models \phi$ if and only if **for all** $a \in M$,
   $$\mathcal{M}, v(x/a) \models \psi.$$

[1]We use certain less formal expressions for this relation as well: "$\mathcal{M}$ believes ϕ under v," "ϕ is true in $\mathcal{M}$ under v," "$\mathcal{M}, v$ satisfies ϕ," and so on.

(b) Say $\phi = \exists x\psi$ for some ψ. Then $\mathcal{M}, v \models \phi$ if and only if **there is some** $a \in M$ such that

$$\mathcal{M}, v(x, a) \models \phi.$$

Notice that this definition tells us precisely how to convert syntax to semantics; from formal logical and mathematical symbols to their concrete meanings.

Remark 1.10. In place of $\mathcal{M}, v \models \phi$, I (and many others) prefer to write

$$\mathcal{M} \models \phi[v].$$

Example 1.11. Consider $L = L_{ord}$, the signature of ordered sets, which consists of a single binary relation symbol R. Let $\mathcal{M}$ be the structure with $M = \mathbb{N}$ and $R^{\mathcal{M}}$ the usual strict ordering on the natural numbers; that is,

$$R^{\mathcal{M}}mn \text{ if and only if } m < n.$$

Let v be the constant valuation with $v(x) = 0$ for all $x \in X$.

We consider the formula $\phi(y) = \forall x(x \neq y \to Ryx)$. Is it the case that $\mathcal{M} \models \phi[v]$? From the definition of satisfaction,

$$\begin{array}{lll} \mathcal{M} \models \phi[v] & \text{if and only if} & \text{for all } a \in M, \mathcal{M} \models (x \neq y \to Ryz)[v(x/a)] \\ & \text{if and only if} & \text{for all } a \in \mathbb{N}, \text{ if } \mathcal{M} \models (x \neq y)[v(x/a)], \text{ then } \mathcal{M} \models Ryz[v(x/a)] \\ & \text{if and only if} & \text{for all } a \in \mathbb{N}, \text{ if } a \neq 0, \text{ then } \mathcal{M} \models R^{\mathcal{M}}0a \\ & \text{if and only if} & \text{for all } a \in \mathbb{N}, \text{ if } a \neq 0, \text{ then } a > 0. \end{array}$$

This final statement is certainly true in this structure, the natural numbers, so yes: $\mathcal{M} \models \phi[v]$.

Again, satisfaction is sensitive to both $\mathcal{M}$ and v.

Example 1.12. (1) Take everything precisely as in Example 1.11, but now with $M = \mathbb{Z}$. Then $\mathcal{M} \not\models \phi[v]$, since, for example,

$$\mathcal{M} \not\models (x \neq y \to Rxy)[v(x/-1)].$$

Check this!

(2) Similarly, if we take the situation of Example 1.11, but now with constant valuation $v(x) = 1$ for all $x \in X$, we find that $\mathcal{M} \not\models \phi[v]$ since, for example,

$$\mathcal{M} \not\models (x \neq y \to Rxy)[v(x/0)].$$

At the start, it is best to work directly from the definition of satisfaction, but before long we will be able to relax a bit and work in a more informal—and more straightforwardly mathematical—way. This is the great benefit of semantics, after all.

Notes 1.13. (1) If a formula ϕ is satisfied in $\mathcal{M}$ for any valuation v, we simply write $\mathcal{M} \models \phi$ and say that ϕ is true in $\mathcal{M}$, or that $\mathcal{M}$ believes ϕ.

(2) If we fix $\mathcal{M}$, the truth value of a formula ϕ under valuation v will only depend on the values of v on variables that occur free in ϕ. That is, if ϕ has free variables in the list $x_1, \ldots, x_n$ and valuations v and v' satisfy $v(x_i) = v'(x_i)$ for $i = 1, \ldots, n$, then

$$\mathcal{M} \models \phi[v] \text{ if and only if } \mathcal{M} \models \phi[v'].$$

(3) As a very simple corollary of (2), if ϕ is a sentence—it contains no free variables—the choice of valuation has no effect whatsoever: either $\mathcal{M} \models \phi$ or $\mathcal{M} \models \neg\phi$.

1.2. **Theories and Models.** In the remarks following Example 1.3, we noted that we often wish to restrict to structures that interpret the symbols in some useful, mathematically-sensible way: we may wish, for example, to consider only those structures for L_{grp} where the binary function symbol is interpreted as an associative (or commutative) operation. We do this by restricting our attention to those structures who satisfy appropriate sentences—axioms, essentially—in the relevant signature.

Definition 1.14. Fix a signature L.

(1) An *L-theory* (or *theory in L*) is a set of sentences in L.
(2) Given an L-theory T, we say that an L-structure $\mathcal{M}$ is a *model of T* if for any $\phi \in T$, $\mathcal{M} \models \phi$. We denote the class of all models of T by $\mathbf{Mod}(T)$

Many important classes of mathematical structures are models of first-order theories in the above sense. For example:

Example 1.15. Let $L = L_{grp}$.

(1) Recall that semigroups are sets with an associative binary operation. Hence the class of semigroups can be realized as $\mathbf{Mod}(T_{sgp})$, where
$$T_{sgp} = \{\forall x \forall y \forall z [m(x, m(y, z)) = m(m(x, y), z)]\},$$
where this single sentence encodes associativity of the binary operation.
(2) A monoid is a semigroup with a well-behaved identity element. That is, the class of monoids is $\mathbf{Mod}(T_{mon})$, where
$$T_{mon} = T_{sgp} \cup \{\, \forall x [m(x, e) = m(e, x)] \,\}.$$
(3) A group is a monoid where every element is invertible. That is, the class of groups is $\mathbf{Mod}(T_{grp})$, where
$$T_{grp} = T_{mon} \cup \{\, \forall x \exists y [m(x, y = e)] \,\}.$$
(4) An abelian group is a group in which the binary operation is commutative. That is, the class of groups is $\mathbf{Mod}(T_{ab})$ where
$$T_{ab} = T_{grp} \cup \{\forall x \forall y [m(x, y) = m(y, x)]\}.$$

We say that the above classes of structures are *axiomatizable*, since it is possible to write down a first-order theory that completely characterizes them[2]. Certain classes, though, are *not* axiomatizable: there is no first order theory T so that the objects of interest are precisely the models of T. We will discuss non-axiomatizability, along with a number of examples, in more detail in Chapter 5, concerning the Compactness Theorem for first-order logic. For now, we simply note that even in the relatively simple world of groups, there are non-axiomatizable classes:

[2]Moreover, they are *finitely axiomatizable*, in the sense that their theories are finite

Example 1.16. (1) A group is said to be *torsion* (or *periodic*) if every element is of finite order; that is, for every element a, there is a natural number $n \geq 1$ such that $a^n = e$. The class of torsion groups is not first-order axiomatizable!

One can begin to see the obstacles to axiomatization. The most (only?) obvious way to capture the finite order property is via an expression like

$$\forall x \left(x = e \vee x^2 = e \vee x^3 = e \vee \cdots \vee x^n = e \vee \dots\right) = \forall x \Big(\bigvee_{n \geq 1} x^n = e \Big)$$

but this involves an *infinite* disjunction, which is not allowed in finitary first-order logic! For us, sentences must be finite.

You might wonder if there is a clever way around this problem—there is not, but we won't be able to show this rigorously without the Compactness Theorem, which is well in the future.

(2) Far more simply, the class of finite groups (that is, those containing only finitely many elements) is not axiomatizable!

The most obvious way to capture this would be via a sentence like

$$\exists_{\leq 1} x(x = x) \vee \exists_{\leq 2} x(x = x) \vee \exists_{\leq 3} x(x = x) \vee \cdots \vee \exists_{\leq n} x(x = x) \vee \cdots = \bigvee_{n \geq l} (\exists_{\leq n} x(x = x))$$

where $\exists_{\leq n}$ is the counting quantifier "there exist at most n..." discussed previously.

Here again, we have an infinite disjunction, which is not available to us; here again, we will need the Compactness Theorem to be sure that there is no alternative way to axiomatize the class.

1.3. **Validity.** We now consider a special class of L-formulas that are *always* true, regardless of the L-structure or valuation.

Definition 1.17. We say that a formula ϕ in L is *valid* if for any L-structure $\mathcal{M}$, $\mathcal{M} \models \phi$. In this case, we write

$$\models \phi.$$

Valid formulas are, in a strong sense, an analogue of the tautologies of propositional logic. Indeed, many of the classic examples of valid formulas arise directly from those tautologies.

Fact 1.18. (1) If Φ is a propositional tautology with propositional variables $p_1, \dots, p_n$, and $\phi_1, \dots, \phi_n$ are L-formulas, the L-formula ϕ obtained by replacing each p_i by ϕ_i in Φ is logically valid.

(2) For any L-formulas ϕ and ψ, the following formulas are valid:

$$[(\forall x \phi) \vee (\forall x \psi)] \to \forall x(\phi \vee \psi)$$

$$\exists x(\phi \wedge \psi) \to [(\exists x \phi) \wedge (\exists x \psi)]$$

(3) Let ϕ and ψ be L-formulas. If x does not occur free in ϕ, the following formula is valid:

$$(\forall x(\phi \to \psi)) \to (\phi \to \forall x \psi).$$

If x does not occur free in ψ, on the other hand, the following is valid:

$$(\forall x(\phi \to \psi)) \to (\exists x \phi \to \psi).$$

Proof. We leave most of the proof as an exercise.

(1) Although a formal proof may require an induction, the idea is relatively clear: whether $\mathcal{M} \models \phi[v]$ is determined by the truth values of the ϕ_i relative to $\mathcal{M}$ and v, in precisely the same way that the truth value of Φ is determined by those of the p_i. Notice that v induces a propositional valuation u_v on the p_i as follows:

$$u_v(p_i) = \begin{cases} 1 & \text{if } \mathcal{M} \models \phi_i[v] \\ 0 & \text{if } \mathcal{M} \not\models \phi_i[v] \end{cases}$$

Since Φ is a tautology, it is true under u_v, so we would expect $\mathcal{M} \models \phi[v]$, as well.

(2) The proof is simple, but worth writing down for practice.

(3) We prove that the second formula is valid. Let $\mathcal{M}$ be an L-structure and v a valuation. Suppose $\mathcal{M} \models \forall x(\phi \to \psi)[v]$ and $\mathcal{M} \models \exists x\phi[v]$: we wish to show that $\mathcal{M} \models \psi[v]$. The second of our assumptions means that there is some $a \in M$ with $\mathcal{M} \models \phi[v(x/a)]$. The first means that for any element of M, including a, $\mathcal{M} \models (\phi \to \psi)[v(x/a)]$; that is, whenever $\mathcal{M} \models \phi[v(x/a)]$, $\mathcal{M} \models \psi[v(x/a)]$. We have already assumed that $\mathcal{M} \models \phi[v(x/a)]$, so, indeed, $\mathcal{M} \models \psi[v(x/a)]$. Since x does not occur free in ψ, changing the value on x to a has no effect: by Fact 1.18(2), $\mathcal{M} \models \psi[v(x/a)]$ if and only if $\mathcal{M} \models \psi[v]$. So we are done.

□

We now turn to the crucial, related ideas of semantic consequence and equivalence.

Definition 1.19. Given L-formulas ϕ and ψ, we say that ψ is a *semantic consequence* of ϕ, and write

$$\phi \models \psi$$

if for any L-structure $\mathcal{M}$ and valuation v,

$$\mathcal{M} \models \phi[v] \text{ implies } \mathcal{M} \models \psi[v].$$

This is sometimes referred to as *logical consequence*, but this is ambiguous.

Definition 1.20. We say that L-formulas ϕ and ψ are *semantically equivalent* (or, sometimes, *logically equivalent*) if both

$$\phi \models \psi \text{ and } \psi \models \phi.$$

In this case, we will write $\phi \equiv \psi$.

Both of these notions should be remind you of similar ideas from propositional logic!

Example 1.21. For any formula ϕ, $\neg\forall x\phi$ is equivalent to $\exists x\neg\phi$; that is,

$$\neg\forall x\phi \equiv \exists x\neg\phi.$$

Proof. Suppose $\mathcal{M}$ and v are such that $\mathcal{M} \models (\neg\forall x\phi)[v]$. Then $\mathcal{M} \not\models (\forall x\phi)[v]$, so it is not the case that for all $a \in M$, $\mathcal{M} \models \phi[v(x/a)]$. So there must be some

$b \in M$ with the property that $\mathcal{M} \not\models \phi[v(x/b)]$. Hence $\mathcal{M} \models \neg\phi[v(x/b)]$, and thus $\mathcal{M} \models (\exists x \neg\phi)[v]$, meaning that

$$\neg\forall x\phi \models \exists x\neg\phi.$$

The other direction is similar. □

Example 1.22. Similarly to the above example, we have the following semantic equivalences:

$$\neg\exists x\phi \equiv \forall x\neg\phi$$
$$\exists x\phi \equiv \neg\forall x\neg\phi$$
$$\forall x\phi \equiv \neg\exists x\neg\phi$$
$$\forall x(\phi \wedge \psi) \equiv (\forall x\phi) \wedge (\forall x\psi)$$
$$\exists x(\phi \vee \psi) \equiv (\exists x\phi) \vee (\exists x\psi)$$

Fact 1.23. As in the propositional case, for any formulas ϕ and ψ,

(1) $\phi \models \psi$ if and only if $\models \phi \to \psi$.
(2) $\phi \equiv \psi$ if and only if $\models \phi \leftrightarrow \psi$

In the next chapter, we will develop the formal deductive system of first-order logic, including notions of provability ($\vdash \phi$) and provable equivalence ($\vdash \phi \leftrightarrow \psi$). As we go, think about the relationship between provability and validity, and between provable equivalence and semantic equivalence. Ultimately, these questions will be resolved by Gödel's Completeness Theorem in Chapter 4, which tells us that these syntactic and semantic notions match up perfectly!

1.4. **Substitutability.** We close this chapter with an important technical detail: the conditions under which a term is *substitutable* in a formula, in a sense that it can replace free occurrences of a variable without destroying the meaning of the original formula.

Notation 1.24. Given a formula ϕ, variable x, and term τ, we denote by

$$\phi_x(\tau)$$

the formula obtained by replacing every free occurrence of x in ϕ with the term τ.

In some cases, it is safe to do this; in others, not safe. So we must be careful.

Definition 1.25. We say that τ is *substitutable for x in ϕ* if no variable in τ becomes bound in $\phi_x(\tau)$.

Example 1.26. Let $L = L_{succ}$. Consider the formula

$$\phi = (x \neq 0 \to \exists y(Sy = x))$$

and term $\tau = SSy$. Then

$$\phi_x(\tau) = (SSy \neq 0 \to \exists y(Sy = SSy)).$$

Clearly the y in τ has become bound by the existential quantifier in the consequent of the conditional, so τ is not substitutable for x in ϕ.

This is a real problem, as the sense of the sentence has changed dramatically:
Before: "If x is nonzero, it is a successor."
After: "If y is not the second successor of 0, its successor and second successor are

equal."
These are *very* different statements.

The result of the substitution in the example is certainly not a special case of ϕ. If τ is substitutable for x in ϕ, on the other hand, the formula $\phi_x(\tau)$ *will be* a special case of ϕ.

Proposition 1.27. If ϕ is a formula, x a variable, and τ a term substitutable for x in ϕ, then

(1) $\models (\forall x\phi) \to \phi_x(\tau)$.
(2) $\models \phi_x(\tau) \to (\exists x\phi)$.

Proof. (1) Fix $\mathcal{M}$ and v. If $\mathcal{M} \models (\forall x\phi)[v]$, then for all $a \in M$,

$$\mathcal{M} \models \phi[v(x/a)].$$

Set $a = \tau[v]$, the value of τ under the valuation v. Then

$$\phi[v(x/a)] = \phi[v(x/\tau[v])] = \phi_x(\tau)[v].$$

So $\mathcal{M} \models \phi_x(\tau)[v]$, meaning that

$$\mathcal{M} \models ((\forall x\phi) \to \phi_x(\tau))[v].$$

(2) Similar to the above. We leave this proof as an exercise.

□

Notation 1.28. We can carry out multiple substitutions as well:

$$\phi_{x_1,\dots,x_n}(\tau_1, \dots, \tau_n)$$

denotes the result of substituting $\tau_1, \dots, \tau_n$ for *original* free occurrences of $x_1, \dots, x_n$. As we will discuss near the end of Chapter 2, this kind of substitution cannot (generally) be done iteratively—that is, $\phi_{x_1,\dots,x_n}(\tau_1, \dots, \tau_n)$ may not be the same as

$$(\dots((\phi_{x_1}(\tau_1))_{x_2}(\tau_2)\dots)_{x_n}(\tau_n).$$

In particular, there is some chance of variable confusion (if, for example, x_2 occurs in τ_1), but we can always fix this by replacing $x_1 \dots, x_n$ by completely new variables $z_1, \dots, z_n$ that do not occur in ϕ or in the terms τ_i. We will address this in detail at the end of Chapter 2.

Definition 1.29. If τ is substitutable for x in ϕ, we say that $\phi_x(\tau)$ is a *substitution instance* of ϕ. More generally, we say that a formula ϕ' is a substitution instance of ϕ if there are terms $\tau_1,\dots,\tau_n$ substitutable for $x_1, \dots, x_n$ in ϕ such that

$$\phi' = \phi_{x_1,\dots,x_n}(\tau_1, \dots, \tau_n).$$

We will return to this notion in the next chapter.

The validities in Proposition 1.27 are very important, and will form part of the axiomatics of the formal system of predicate logic, to which we turn in the next chapter.

Chapter 1 Exercises

Exercise 1.1. A *sentence* is a formula with no free variables.

(i) Find a language in which there are quantifier-free sentences.

(ii) Show that if ϕ is a sentence and $\mathcal{M}$ a structure, either $\mathcal{M} \models \phi[v]$ for all valuations v, or $\mathcal{M} \not\models \phi[v]$ for all valuations v.

Exercise 1.2. Let L be the signature of the natural numbers with successor, i.e. $L = \langle e, S, R\rangle$, with S a unary function symbol, and R a binary relation. Consider the following structures:

- $\mathcal{M}_1$: $M_1 = \mathbb{N}$, $e^{\mathcal{M}_1} = 0$, $S^{\mathcal{M}_1}(n) = n + 1$ for all $n \in \mathbb{N}$, and $R^{\mathcal{M}_1}mn$ iff $m < n$.
- $\mathcal{M}_2$: $M_2 = \mathbb{N}$, $e^{\mathcal{M}_2} = 3$, $S^{\mathcal{M}_2}(n) = 2n$ for all $n \in \mathbb{N}$, and $R^{\mathcal{M}_2}mn$ iff n is divisible by m.

(i) Compute the value of the terms $\tau = SSe$ and $\tau' = SSSx$ in each structure under the valuation $v : X \to \mathbb{N}$ with $v(y) = 1$ for all $y \in X$.

(ii) Compute the truth value of the following atomic formulas in $\mathcal{M}_1$ and $\mathcal{M}_2$ under the same valuation v as above:

$$\phi_1 : \quad SSe = x$$
$$\phi_2 : \quad Sx = y$$
$$\phi_3 : \quad RS(e)SS(x)$$

(iii) Which of the following sentences are true in $\mathcal{M}_1$, and which in $\mathcal{M}_2$?

$$\psi_1 : \quad \forall x(x \neq e \to \exists y(Sy = x))$$
$$\psi_2 : \quad \forall x \exists y(Rxy \wedge RyS(x))$$
$$\psi_3 : \quad \neg\exists x(Rxe)$$

Exercise 1.3. Consider the language of (semi)groups, $L = \langle e, m\rangle$, and the L-structure $\mathcal{M}$ with $M = \mathbb{Z}$, $e^{\mathcal{M}} = 0$, and $m^{\mathcal{M}}(p, q) = p + q$.

(i) Let ϕ be the formula $\exists x(m(x, x) = y))$. Find valuations v so that (a) $\mathcal{M} \models \phi[v]$ and (b) $\mathcal{M} \not\models \phi[v]$.

(ii) If $\mathcal{M}'$ has $M' = \mathbb{R}$, and interprets e as 0 and m as addition, is it the case that $\mathcal{M}' \models \phi$?

(iii) Is the formula $\forall x \forall y(m(x, y) = m(y, x))$ valid? If not, give an example of a L-structure in which it is false.

Exercise 1.4. Let L be a signature with two unary relation symbols, P and L. Two of the following formulas are valid, two invalid. Find the valid formulas, and justify your conclusions. For each invalid formula, find a structure (and valuation, if necessary) where the formula is falsified.

(i) $\forall x(Px \to Lx) \to \exists x(Px \wedge Lx)$

(ii) $\neg\forall x\neg(Px) \to \exists xPx$

(iii) $\forall x(Px \wedge Lx) \to (\forall x(Px) \wedge \forall x(Lx))$

(iv) $\forall x(Px \vee Lx) \to (\forall x(Px) \vee \forall x(Lx))$

Exercise 1.5. Prove that, for any propositional tautology Φ with propositional variables $p_1, p_2, \ldots, p_n$ and any formulas $\phi_1, \phi_2, \ldots, \phi_n$, the first-order formula ϕ obtained from Φ by replacing each p_i with ϕ_i is valid.

2. Formal System of First-order Logic

As we did in the case of propositional logic, we will give a complete description of a system of deduction and provability for first-order logic. Again, this description will be purely formal, defining these notions in terms of the manipulation of symbols, without regard for their potential meanings. Semantics of the kind considered in the previous chapter will not reappear until Chapter 4!

Formal language: The basic vocabulary consists of

(1) Logical symbols: We will use connectives $\neg$ and $\rightarrow$, and the quantifier $\forall$. This is enough: any formula is equivalent to one expressed with only these logical symbols (see the exercise below).
(2) Nonlogical symbols: Unless otherwise specified, we will always include the equality sign $=$. Beyond this, as discussed, we have a signature, consisting of a set of contant symbols, $\mathcal{C}$, function symbols, $\mathcal{F}$, and predicate symbols, $\mathcal{R}$, along with a countably infinite set of variable symbols, X.
(3) We generate terms and formulas from this basic vocabulary as described before.

Axioms:

(1) **Propositional schema:** Let ϕ, ψ, and η be L-formulas. The following are axioms:
   (a) $\phi \rightarrow (\psi \rightarrow \phi)$
   (b) $[\phi \rightarrow (\psi \rightarrow \eta)] \rightarrow [(\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \eta)]$
   (c) $[(\neg\psi) \rightarrow (\neg\phi)] \rightarrow (\phi \rightarrow \psi)$
   Notice that these are precisely the axioms of propositional logic!
(2) **Quantifier scheme:** Let ϕ and ψ be L-formulas, and x a variable with no free occurrences in ϕ. Then we take
$$[\forall x(\phi \rightarrow \psi)] \rightarrow [\phi \rightarrow (\forall x\psi)]$$
   to be an axiom.

   Remark 2.1. The condition on free occurrences of x is critically important. Take $\phi = \psi = (x > 0)$ in L_{succ}. If we were to consider the quantifier schema above in this case—although we definitely shouldn't, since x occurs free in ϕ—we would have
$$\forall x(x > 0 \rightarrow x > 0) \rightarrow (x > 0 \rightarrow \forall x(x > 0)).$$
   The antecedent is true in any interpretation, but we can easily find structures in which the consequent is false. This means that this instance of the scheme would *falsifiable*, which we absolutely cannot have with any of our axioms.
(3) **Substitution scheme:** Here we take one of the validities from the end of the last chapter as an axiom scheme: if ϕ is an L-formula, x a variable, and

τ a term substitutable for x in ϕ,

$$(\forall x\phi) \to \phi_x(\tau)$$

is taken to be an axiom. Here again, the condition that τ be substitutable for x in ϕ, in the sense of Definition 1.25, is essential—substituting blindly may, as already discussed in Example 1.26, lead to disaster.
As a very important special case of this axiom scheme, we have the instance where $\tau = x$ itself:

$$\forall x\phi \to \phi$$

is included among the axioms for all ϕ and x.

(4) **Equality schema:** We include a few additional schema to ensure that the equality sign behaves as it should, and that function symbols and relations symbols are well-defined relative to the relation of equality. To handle the first problem, given any terms τ_1, τ_2, and τ_3, we have axioms
   (a) $\tau_1 = \tau_1$.
   (b) $\tau_1 = \tau_2 \leftrightarrow \tau_2 = \tau_1$.
   (c) $[\tau_1 = \tau_2 \wedge \tau_2 = \tau_3] \to \tau_1 = \tau_3$.
   (d) For any $f \in \mathcal{F}$ of arity n, and variables $x_1,\ldots,x_n,y_1,\ldots,y_n$, we have axiom

$$(x_1 = y_1 \to (x_2 = y_2 \to \ldots (x_n = y_n \to fx_1 \ldots x_n = fy_1 \ldots y_n) \ldots)).$$

   (e) For any $R \in \mathcal{R}$ of arity n, and variables $x_1,\ldots,x_n,y_1,\ldots,y_n$, we have axiom

$$(x_1 = y_1 \to (x_2 = y_2 \to \ldots (x_n = y_n \to [Rx_1 \ldots x_n \to Ry_1 \ldots y_n]) \ldots)).$$

Rules of inference: As with the propositional calculus, we would like to keep our inference rules minimal. In this case, we take

(1) **Modus Ponens (MP):** For any formulas ϕ and ψ, given ϕ and $\phi \to \psi$, we may conclude ψ.
(2) **Generalization (Gen):** From any formula ϕ, we may conclude $\forall x\phi$ (for any variable x!).

Remark 2.2. Proof theorists tend to present rules of inference like those above in the following visual form:

$$\frac{\phi,\ \phi \to \psi}{\psi}\quad \text{(MP)} \qquad\qquad \frac{\phi}{\forall x\phi}\quad \text{(Gen)}$$

Here the hypotheses are listed above the horizontal line, with the conclusion below. Especially in deductive systems with a richer supply of inference rules, this way of organizing information can be very advantageous.

Definition 2.3. A *proof* of an L-formula ϕ consists of a finite sequence of L-formulas

$$\phi_1, \phi_2, \ldots, \phi_i, \ldots, \phi_n = \phi,$$

where each formula ϕ_i, $i = 1, \ldots, n$, is either

(1) an instance of an axiom scheme of predicate logic, or
(2) follows from $\phi_1, \ldots, \phi_{i-1}$ by Modus Ponens or Generalization.

If there exists a proof of ϕ, we say that ϕ is *provable*, and write

$$\vdash_L \phi.$$

We may also say, in this case, that ϕ is a *theorem of predicate logic.*

More generally,

Definition 2.4. Given a collection of L-formulas T and an L-formula ϕ, a *proof of ϕ from T* is just as described above, but in clause (1) ϕ_i may be either an axiom of predicate logic *or an element of T*. If there is such a proof, we say that ϕ is *provable from T*, and write

$$T \vdash_L \phi.$$

Remark 2.5. When there is no chance of confusion about the language, we omit the subscript L, and simply write $\vdash \phi$ or $T \vdash \phi$ to indicate provability.

Remark 2.6. All of the theorems of propositional logic are theorems of predicate logic; that is, if ϕ is a propositional tautology in propositional variables $p_1, \ldots, p_n$, and $\chi_1, \ldots, \chi_n$ are L-formulas, then

$$\vdash \phi(p_1/\phi_1, \ldots, p_n/\chi_n).$$

That is, the formula obtained by replacing each p_i by χ_i is a theorem of predicate logic.

Example 2.7. We work in $L = L_{succ}$. The propositional formula $\phi = (p_1 \wedge p_2) \to p_1$ is a tautology. Take $\chi_1 = Sx \neq 0$ and $\chi_2 = Sx > 1$. Then

$$\vdash_L ((Sx \neq 0) \wedge (Sx > 1)) \to (Sx \neq 0).$$

Theorems obtained in this way are essentially trivial, though. Slightly less trivial are those theorems that can be proved just using the propositional fragment of predicate logic:

Definition 2.8. If a formula ϕ can be proven from a collection of formulas T using only the propositional axioms and the inference rule Modus Ponens, we say that ϕ is a *tautological consequence of T*.

Still, there are far more fish in the provable sea. Before we begin to develop a library of useful theorems, we perform a necessary sanity check: we show that nothing we prove can ever be false.

Theorem 2.9 (Soundness)**.** Any provable L-formula ϕ is valid:

$$\vdash \phi \Rightarrow \models \phi.$$

Proof. Suppose that ϕ is provable, with proof

$$\phi_1, \phi_2, \ldots, \phi_n = \phi.$$

We will show by induction that for any L-structure $\mathcal{M}$ such that $\mathcal{M} \models \phi_j$ for all $j < i$—that is, $\mathcal{M}$ believes each ϕ_j, independent of valuation—then $\mathcal{M} \models \phi_i$. For definiteness, we fix a valuation v.

To that end, suppose that $\mathcal{M}$ and v are such that $\mathcal{M} \models \phi_j[v]$ for all $j < i$.

- If ϕ_i is a propositional axiom, then clearly $\mathcal{M} \models \phi_i[v]$.

- If ϕ_i is an instance of the quantifier axiom scheme, it must be of the form
$$\forall x(\phi \to \psi) \to (\phi \to \forall x\psi.$$
for some ϕ and ψ. It is easy to verify that any such formula is valid.
- If ϕ_i is an instance of the substitution axiom scheme, it must be of the form
$$\forall x\psi \to \psi_x(\tau)$$
for some L-formula ψ and term τ. We proved that any such formula is valid in Proposition 1.27(1).
- If ϕ_i is an equality axiom, it is clearly valid: unless we are willing to tolerate a descent in madness, we will make our structures interpret $=$ in a sensible way.
- If ϕ_i follows by Modus Ponens from earlier formulas, say ϕ_j and ϕ_k, $j, k < i$, then without loss of generality, $\phi_k = \phi_j \to \phi_i$. By the induction hypothesis,
$$\mathcal{M} \models \phi_j[v] \text{ and } \mathcal{M} \models (\phi_j \to \phi_i)[v]$$
By the definition of the satisfaction relation,
$$\mathcal{M} \models \phi_i[v]$$
and we are done.
- If ϕ_i follows by Generalization from some earlier formula ϕ_j, then $\phi_i = \forall x\phi_j$. By the induction hypothesis, $\mathcal{M} \models \phi_j$, so in particular we have
$$\mathcal{M} \models \phi_j[v(x/m)]$$
for any $m \in M$, meaning that $\mathcal{M} \models \forall x\phi[v]$. So we are done.

□

Remark 2.10. This theorem is often referred to as the Correctness Theorem, instead, as it reassures us of the correctness of the deductive system: if we can prove it, it must be true.

There is an easy extension of the Soundness Theorem, as stated above:

Theorem 2.11. Let T be a collection of L-formulas and ϕ an L-formula. Then
$$T \vdash \phi \Rightarrow T \models \phi.$$

Ultimately, what we really want is, as in the case of propositional logic, the converse of this result: completeness of the deductive system. This will have to wait until Chapter 4.

We close with a series of useful facts, derivable in the deductive system described above, that help us drastically shorten our formal proofs. As they themselves are established via formal proofs, this will also allow us to get a little practice *using* our deductive system.

We begin with a few propositional tricks, whose proofs are left as an exercise. In each case, we indicate the abbreviation we will use when referring to these derived inference rules in the the course of our proofs.

Lemma 2.12. Let ϕ, ψ, and χ be formulas. Then we have:

(1) (**Contra**, short for "contrapositive") If $\vdash \phi \to \psi$, then $\vdash \neg\psi \to \neg\phi$.

(2) (**Comp**$_\to$, short for "composition of implication") If $\vdash \phi \to \psi$ and $\vdash \psi \to \chi$, then $\vdash \phi \to \chi$.

Lemma 2.13 (∀**Rule).** If $\vdash \phi \to \psi$ and x does not occur free in ϕ, then $\vdash \phi \to \forall x\psi$.

Proof. We proceed via a formal deduction, assuming, in line (1), that $\vdash \phi \to \psi$:

(1)	$\vdash \phi \to \psi$	Assump
(2)	$\vdash \forall x(\phi \to \psi)$	Gen (1)
(3)	$\forall x(\phi \to \psi) \to (\phi \to \forall x\psi)$	Axiom
(4)	$\vdash \phi \to \forall x\psi$	MP (2), (3)

This completes the proof of the lemma. □

Lemma 2.14 (∃**Rule).** If $\vdash \phi \to \psi$ and x has no free occurrences in ψ, then $\vdash (\exists x\phi) \to \phi$.

Proof. We give a formal proof here as well:

(1)	$\vdash \phi \to \psi$	Assump.
(2)	$\vdash (\phi \to \psi) \to (\neg\psi \to \neg\phi)$	Taut(ology)
(3)	$\vdash \neg\psi \to \neg\phi$	MP (1), (2)
(4)	$\vdash \neg\phi \to \forall x\neg\phi$	∀Rule (3)
(5)	$\vdash (\neg\psi \to \forall x\neg\phi) \to (\neg\forall x\neg\phi \to \psi)$	Taut
(6)	$\vdash \neg\forall x\neg\phi \to \psi$	MP (4), (5)

Rewriting the antecedent of the last line, we have $\exists x\phi \to \psi$. □

The next rule is gives the converse of the substitution axiom:

Lemma 2.15 (**Unsub).** If ϕ is a formula, x a variable, and τ a term substitutable for x in ϕ, $\vdash \phi_x(\tau) \to \exists x\phi$.

Proof.

(1)	$\vdash \forall x\neg\phi \to \neg\phi_x(\tau)$	Axiom
(2)	$\vdash \neg\neg(\forall x\neg\phi) \to \forall x\neg\phi$	Taut
(3)	$\vdash \neg\neg(\forall x\neg\phi) \to \neg\phi_x(\tau)$	Comp$_\to$ (1), (2)
(4)	$\vdash \neg\exists x\phi \to \neg\phi_x(\tau)$	Rewriting (3)
(5)	$\vdash \phi_x(\tau) \to \exists x\phi$	Contra (4)

□

Lemma 2.16. Let ϕ and ψ be formulas.

(1) (∃**Dist**) If $\vdash \phi \to \psi$, then $\vdash \exists x\phi \to \exists x\psi$.
(2) (∀**Dist**) If $\vdash \phi \to \psi$, then $\vdash \forall x\phi \to \forall x\psi$.

Recall that a formula ϕ' is said to be a substitution instance of a formula ϕ if it is of the form

$$\phi_{x_1,\ldots,x_n}(\tau_1, \ldots, \tau_n)$$

where each term τ_i is substitutable for x_i in ϕ.

Lemma 2.17. Let ϕ' be a substitution instance of ϕ. Then

$$\vdash \phi \Rightarrow \vdash \phi'.$$

Proof. We proceed by implicit induction on the number of substitutions, n.

Case $n = 1$: Then $\phi' = \phi_x(\tau)$. Then

(1)	$\vdash \phi$	Assump.
(2)	$\vdash \forall x \phi$	Gen (1)
(3)	$\vdash \forall x \phi \to \phi_x(\tau)$	Axiom
(4)	$\vdash \phi_x(\tau)$	MP (2), (3)

Larger n: We would like to simply proceed using the base case above, obtaining

$$\phi_{x_1,\dots,x_n}(\tau_1, \dots, \tau_n)$$

by iterative substitution, as

$$\phi_{x_1}(\tau_1) \rightsquigarrow (\phi_{x_1}(\tau_1))_{x_2}(\tau_2) \rightsquigarrow \cdots \rightsquigarrow (\dots((\phi_{x_1}(\tau_n))_{x_2}(\tau_2)\dots)_{x_n}(\tau_n)$$

but, as already mentioned, this iterative process may lead to a different end result: τ_2 might contain variables in τ_1, and so on.

We fix this through a two-stage process. First, we choose variables $z_1, \dots, z_n$ that do not appear in ϕ or $\tau_1, \dots, \tau_n$—because we have allowed ourselves infinitely many variables, this is possible. We rename the variables $x_1, \dots, x_n$ to $z_1, \dots, z_n$. Using the base case, we have inferences

$$\begin{aligned}
&\vdash \phi \\
&\vdash \phi_{x_1}(z_1) \\
&\vdash \phi_{x_1,x_2}(z_1, z_2) \\
&\vdots \\
&\vdash \phi_{x_1,\dots,x_n}(z_1, \dots, z_n)
\end{aligned}$$

Set $\psi = \phi_{x_1,\dots,x_n}(z_1, \dots, z_n)$. We have ensured that it is now safe to substitute, iteratively, the terms τ_i into ψ. In particular, we have

$$\begin{aligned}
&\vdash \psi \\
&\vdash \psi_{z_1}(\tau_1) \\
&\vdash \psi_{x_1,x_2}(\tau_1, \tau_2) \\
&\vdots \\
&\vdash \psi_{x_1,\dots,x_n}(\tau_1, \dots, \tau_n)
\end{aligned}$$

That is, $\vdash \phi'$. □

As an aside, we consider the following natural question: under what conditions does the converse of Lemma 2.17 hold? That is, when does provability of a substitution instance of a formula ϕ imply provability of ϕ? Only in a very special case:

Theorem 2.18. Let T be a set of L-formulas, and ϕ an L-formula with variables $x_1, \dots, x_n$. Consider a language $L' \supseteq L$ containing new constant symbols $c_1, \dots, c_n$. Then

$$T \vdash_{L'} \phi_{x_1,\dots,x_n}(c_1, \dots, c_n) \text{ if and only if } T \vdash_L \phi.$$

Proof. ($\Leftarrow$) This is precisely Lemma 2.17.

($\Rightarrow$) Suppose that $T \vdash_{L'} \phi_{x_1,\dots,x_n}(c_1, \dots, c_n)$—in the interest of brevity, we denote by ϕ' the L'-formula $\phi_{x_1,\dots,x_n}(c_1, \dots, c_n)$. Then there is a proof $\phi'_1, \dots, \phi'_m = \phi'$, where each formula is in L'. We need to convert this into a proof in L. To that

end, take variables $y_1, \ldots, y_n$ that does not occur in $\phi'_1, \ldots, \phi'_m$, and replace all occurrences of c_i in the ϕ'_j by y_i, $1 \leq i \leq n$. The result is a proof

$$\phi_1, \ldots, \phi_m = \phi_{x_1,\ldots,x_n}(y_1, \ldots, y_n).$$

So $T \vdash_L \phi_{x_1,\ldots,x_n}(y_1, \ldots, y_n)$. Since ϕ is a substitution instance of this formula, $T \vdash_L \phi$. □

Recall from an earlier exercise that we define:

Definition 2.19. Given a formula ϕ with free variables $x_1, \ldots, x_n$, the *closure* of ϕ is the formula

$$\forall x_1 \ldots \forall x_n \phi.$$

Lemma 2.20. For any set of formulas T and formula ϕ, if ϕ' is the closure of ϕ we have:

$$T \vdash \phi \text{ if and only if } T \vdash \phi'.$$

Proof. We have already proved the "if" direction, by induction. The "only if" direction follows by iterated Generalization. □

This sequence of lemmas leads to a very important conclusion: the Deduction Theorem for first-order logic. This resembles the analogous result for propositional logic, and the proof is strikingly similar—there are just a few more details to attend to.

Theorem 2.21 (Deduction Theorem, DT)**.** Let T be a set of formulas, and ϕ a closed formula—that is, ϕ contains no free variables—and ψ an arbitrary formula. Then

$$T \vdash \phi \to \psi \text{ if and only if } T, \phi \vdash \psi.$$

Proof. We can *almost* use the same proof as the Deduction Theorem for propositional logic: there is just one additional case to consider in the proof of the direction

$$T, \phi \vdash \psi \Rightarrow T \vdash \phi \to \psi.$$

As in the earlier proof, we suppose that $T, \phi \vdash \psi$—that is, there is a proof $\phi_1, \ldots, \phi_n = \psi$—and prove by induction that $T \vdash \phi \to \phi_i$ for all $1 \leq i \leq n$. Suppose that this is the case for all $i < j$. Beyond the cases addressed in the earlier proof, we must consider the possibility that ϕ_j arises from an earlier formula by the new inference rule, Generalization. That is,

$$\phi_j = \forall x \phi_i$$

for some $i < j$. By the induction hypothesis, $T \vdash \phi \to \phi_i$. Since ϕ is closed, x certainly does not occur free, so we can use the ∀Rule to infer that $T \vdash \phi \to \forall x \phi_i$; that is, $T \vdash \phi \to \phi_j$. This completes the proof of the (new, more complicated) induction step. □

Remark 2.22. It cannot have escaped your notice that the formal language we have developed, which uses only the logical symbols $\to$, $\neg$, and $\forall$, is not the most natural way of expressing ideas—consider, for example, the nested conditionals in the equality axioms. It is an easy exercise to show, by induction on complexity of formulas, that any formula built from the full complement of connectives and quantifiers is provably equivalent to one built just using this restricted set.

The cost of allowing more logical symbols in the language is the need for additional rules of inference.

Chapter 2 Exercises

Exercise 2.1. Given a formula ϕ, we define its *(universal) closure* ϕ' to be the formula obtained from ϕ by adding a universal quantifier $\forall x$ for each variable x free in ϕ. For example, if ϕ is $\exists y(Rxy)$, ϕ' is $\forall x \exists y(Rxy)$. Prove that $\vdash \phi'$ if and only if $\vdash \phi$. (Hint: each direction will require an induction on the number of free variables in ϕ).

Exercise 2.2. Prove each of the following deduction rules, which describe, roughly, the distributivity of quantifiers over implication.

($\forall$Dist) If $\vdash \phi \to \psi$, then $\vdash (\forall x\phi \to \forall x\psi)$.
($\exists$Dist) If $\vdash \phi \to \psi$, then $\vdash (\exists x\phi \to \exists x\psi)$.

Exercise 2.3. We say that a set of first-order formulas T is *satisfiable* if there is a structure $\mathcal{M}$ and valuation v so that $\mathcal{M} \models \phi[v]$ for all $\phi \in T$.

(i) Show that $\{\phi_1, \phi_2, \ldots, \phi_n\} \models \phi$ iff $\{\phi_1, \phi_2, \ldots, \phi_n, \neg\phi\}$ is not satisfiable.

(ii) Using correctness of the deductive system, show that if $\{\phi_1, \phi_2, \ldots, \phi_n\} \vdash \phi$ then $\{\phi_1, \phi_2, \ldots, \phi_n, \neg\phi\}$ is not satisfiable.

(iii) Let $L = L_{ord}$. Using (ii), show that it is *not the case* that

$$\{ \quad \forall x(x \leq x), \quad \forall x\forall y([x \leq y \wedge y \leq x] \to x = y), \quad \forall x\forall y\forall z([x \leq y \wedge y \leq z] \to x \leq z) \quad \}$$
$$\vdash \quad \forall x\forall y(x < y \to \exists z[x < z \wedge z < y])$$

where $x < y$ is an abbreviation for $x \leq y \wedge x \neq y$. That is, not every partial order is dense.

Exercise 2.4. Fill in the missing details of the following proofs:
(i) For any ϕ and $x, y, \vdash \exists x\forall y\phi \to \forall y\exists x\phi$.

(1)		Sub. Axiom
(2)		$\exists$Dist (1)
(3)		Quant. Axiom (2)

(ii) For any ϕ, ψ and x, $\vdash (\forall x\phi \to \exists x\psi) \to \exists x(\phi \to \psi)$.

(1)	$\vdash \psi \to (\phi \to \psi)$	
(2)	$\vdash$	$\exists$Dist (1)
(3)	$\vdash (\neg\phi) \to (\phi \to \psi)$	
(4)	$\vdash$	$\exists$Dist (3)
(5)	$\vdash$	Taut.
(6)	$\vdash \forall x(\neg\neg\phi) \to \forall x\phi$	$\forall$Dist (5)
(7)	$\vdash \neg\forall x\phi \to \neg\forall x(\neg\neg\phi)$	
(8)	$\vdash \neg\forall x\phi \to \exists x(\neg\phi)$	
(9)	$\vdash$	Comp. (4), (8)
(10)	$\vdash (\exists x\psi \to \exists x(\phi \to \psi))$ $\to [(\neg\forall x\phi \to \exists x(\phi \to \psi)) \to ((\forall x\phi \to \exists x\psi) \to \exists x(\phi \to \psi))]$	Taut
(11)		MP ($\times 2$)

(iii) For any ϕ, ψ and x, $\vdash \exists x(\phi \to \psi) \to (\forall x\phi \to \exists x\psi)$.

(1)		Sub. Axiom
(2)	$\forall x\phi \vdash \phi$	DT (1)
(3)	$\vdash \phi \to ((\phi \to \psi) \to \psi)$	
(4)	$\forall x\phi \vdash (\phi \to \psi) \to \psi$	
(5)		$\exists$Dist (4)
(6)	$\exists x(\phi \to \psi), \forall x\phi \vdash \exists x\psi$	
(7)		DT (6)
(8)		DT (7)

(iv) Let L be a language with equality and a binary relation symbol R. Show

$$\vdash \forall x\forall y(x \neq y \to Rxy) \to \forall x\forall y(Rxy \to Ryx)$$

(1)	$\vdash \forall y(x \neq y \to Rxy) \to (x \neq y \to Rxy)$	Sub. Axiom
(2)	$\vdash \forall x\forall y(x \neq y \to Rxy) \to \forall y(x \neq y \to Rxy)$	
(3)	$\vdash$	Comp. (1), (2)
(4)	$\forall x\forall y(x \neq y \to Rxy) \vdash x \neq y \to Rxy$	
(5)	$\forall x\forall y(x \neq y \to Rxy) \vdash y \neq x \to Ryx$	Instance (4)
(6)	$\vdash y = x \to (Rxy \to Ryx)$	Axioms of $=$
(7)	$\forall x\forall y(x \neq y \to Rxy) \vdash Rxy \to Ryx$	(Long) Taut.
(8)		Gen. (7)
(9)	$\forall x\forall y(x \neq y \to Rxy) \vdash \forall x\forall y(Rxy \to Ryx)$	
(10)	$\vdash$	

Exercise 2.5. Let L be the language of the natural numbers with successor, i.e. $L = \langle S, e, R\rangle$.

(i) Complete the following proof that

$$x = e \vdash x = Se$$

(1)	$x = e \vdash x = e$	
(2)	$x = e \vdash \forall x(x = e)$	
(3)		Sub. Axiom
(4)	$x = e \vdash Se = e$	MP (2), (3)
(5)	$x = e \vdash x = Se$	

(ii) Show that $\nvdash x = e \to x = Se$. (Hint: show that the closure of $x = e \to x = Se$ is not valid.)

(iii) Why do parts (i) and (ii) not contradict the Deduction Theorem for first-order logic?

Exercise 2.6. Definition: We say that a set of formulas T is *consistent* if there is no formula ϕ such that $T \vdash \phi$ and $T \vdash \neg\phi$.

(i) Show that T is inconsistent iff for any formula ψ, $T \vdash \psi$.

(ii) Show that for any set of formulas T and formula ϕ, $T \vdash \phi$ iff the set $T \cup \{\neg\phi\}$ is inconsistent.

Exercise 2.7. Definition: We say that a set of formulas T is *complete* if it is consistent and for any formula ϕ, either $T \vdash \phi$ or $T \vdash \neg\phi$.

(i) Give an example of a (nonempty!) incomplete set of formulas in the formal language of your choice.

(ii) Open-ended question: Is it true that for any set of formulas T in a given language L, there is a complete set $\bar{T} \supseteq T$?

3. Prenex Forms

Our goal now is to prove the following: any formula ϕ is provably equivalent to a formula ϕ' of the following very specific form:

$$Q_1 x_1 \dots Q_n x_n \psi$$

where

(1) ψ is quantifier-free (or, to be completely technically correct, contains only *bounded quantifiers*—we will not consider this possibility).
(2) Each Q_i is $\exists$ or $\forall$ for $1 \leq i \leq n$.
(3) The variables $x_1, \dots, x_n$ are distinct.

Definition 3.1. We say that the formula ϕ' described above is in *prenex form*. We refer to the string of quantifiers $Q_1 x_1 \dots Q_n x_n$ as the *prefix* of ϕ'; we call ψ the *matrix* of ϕ'.

Our approach will be simple, in principle: we will, very carefully, pull all of the quantifiers out of the formula. We will need to do a great deal of work, though, to describe just how this process is to be carried out, and to ensure that the result is provably equivalent to the original formula.

3.1. Rewriting Rules.

We begin with a few basic results. For example, please recall the following:

Fact 3.2. For any formula ϕ, if ϕ' is the closure of ϕ,

$$\vdash \phi \text{ if and only if } \vdash \phi'.$$

It is also very easy to prove the following:

Fact 3.3. For any permutation $i_1, \dots, i_n$ of $\{1, \dots, n\}$,

$$\vdash \forall x_1 \dots \forall x_n \phi \leftrightarrow \forall x_{i_1} \dots \forall x_{i_n}$$

$$\vdash \exists x_1 \dots \exists x_n \phi \leftrightarrow \exists x_{i_1} \dots \exists x_{i_n}$$

That is, we can swap any two (or more) universal quantifiers, and any two (or more) existential quantifiers. Important: we cannot, in general, swap an existential and a universal quantifier!

Theorem 3.4 (Equivalent Replacement)**.** Say ϕ' can be obtained from a formula ϕ by replacing subformulas $\chi_1, \ldots, \chi_n$ of ϕ by $\chi_1', \ldots, \chi_n'$. If $\vdash \chi_i \leftrightarrow \chi_i'$ for $1 \leq i \leq n$, then

$$\vdash \phi \leftrightarrow \phi'.$$

The essential idea of the theorem is that if we replace pieces of a formula by provably equivalent pieces, the result is provably equivalent to the original formula.

Example 3.5. Let ϕ be the formula $\forall x\,[\neg\neg\pi \rightarrow \exists y(x \neq y \vee \phi)]$. Take χ_1 to be $\neg\neg\pi$, and χ_2 to be $x \neq y \vee \psi$. Consider

$$\chi_1' = \pi \qquad \text{and} \qquad \chi_2' = (x = y \rightarrow \psi).$$

Clearly, $\vdash \chi_1 \leftrightarrow \chi_1'$ and $\vdash \chi_2 \leftrightarrow \chi_2'$, so the theorem implies that

$$\vdash \phi \leftrightarrow \forall x\,[\pi \rightarrow \exists y(x = y \rightarrow \psi)]\,.$$

Proof. We proceed by induction on the complexity of the formula ϕ.

Base case: Let ϕ be an atomic formula. In this case, there are only two possible subformulas: ϕ itself, and $\emptyset$, the empty formula.

- If $\chi = \phi$, then the result of the replacement is just $\phi' = \chi'$. Since $\vdash \chi \leftrightarrow \chi'$, by assumption, it is obviously the case that $\vdash \phi \leftrightarrow \phi'$.
- If $\chi = \emptyset$, then there is no replacement whatsoever, so in fact $\phi = \phi'$.

Induction steps:

- Suppose that $\phi = \neg\psi$, and that the theorem holds for ψ. That is, $\psi \leftrightarrow \psi'$. In particular, $\vdash \psi \rightarrow \psi'$, meaning that $\vdash \neg\psi' \rightarrow \neg\psi$. Since $\phi' = \neg\psi'$, we thus have

  $$\vdash \phi' \rightarrow \phi.$$

  The proof of the other direction, $\vdash \phi \rightarrow \phi'$, is similar.
- Suppose that $\phi = \psi \rightarrow \eta$, and that the proposition holds for ψ and η; that is,

  $$\vdash \psi \leftrightarrow \psi' \qquad \text{and} \qquad \vdash \eta \leftrightarrow \eta'.$$

  In particular, $\vdash \psi' \rightarrow \psi$ and $\vdash \eta' \rightarrow \eta$. So we have the following formal proof:

(1)	$\vdash \psi' \rightarrow \psi$	Assump
(2)	$\vdash \eta' \rightarrow \eta$	Assump
(3)	$\vdash (\psi' \rightarrow \psi) \rightarrow [(\psi \rightarrow \eta) \rightarrow (\psi' \rightarrow \eta)]$	Taut
(4)	$\vdash (\psi' \rightarrow \eta) \rightarrow [(\eta \rightarrow \eta') \rightarrow (\psi' \rightarrow \eta')]$	Taut
(5)	$\vdash (\psi' \rightarrow \psi) \rightarrow ((\psi \rightarrow \eta) \rightarrow [(\eta \rightarrow \eta') \rightarrow (\psi' \rightarrow \eta')])$	$\text{Comp}_{\rightarrow}(3), (4)$
(6)	$\vdash (\psi \rightarrow \eta) \rightarrow [(\eta \rightarrow \eta') \rightarrow (\psi' \rightarrow \eta')]$	MP (3), (5)
(7)	$\vdash (\eta \rightarrow \eta') \rightarrow [(\psi \rightarrow \eta) \rightarrow (\psi' \rightarrow \eta')]$	Taut... (6)
(8)	$\vdash (\psi \rightarrow \eta) \rightarrow (\psi' \rightarrow \eta')$	MP (2), (7)

  This means precisely that $\vdash \phi \leftrightarrow \phi'$. The opposite direction, $\vdash \phi' \rightarrow \phi$, is similar.

- Suppose that $\phi = \forall x\psi$, and that the theorem holds for ψ. Then $\phi' = \forall x\psi'$, and we know from the induction hypothesis that $\vdash \psi \leftrightarrow \psi'$. In particular, $\vdash \psi \to \psi'$. We therefore have

(1)	$\vdash \psi \to \psi'$	Assump
(2)	$\vdash \forall x(\psi \to \psi')$	Gen (1)
(3)	$\vdash \forall x\psi \to \forall x\psi'$	$\forall$Dist (2)

  This means precisely that $\vdash \psi \to \psi'$. Again, the opposite direction is similar.

□

Theorem 3.6. Let ϕ and ψ be formulas and x a variable.

(1) $\vdash \exists x(\neg\phi) \leftrightarrow \neg\forall x\phi$ and $\vdash \forall x(\neg\phi) \leftrightarrow \neg\exists x\phi$.
(2) If x is not free in ϕ and $*$ is one of the connectives $\vee$, $\wedge$, or $\to$,

$$\vdash \forall x(\phi * \psi) \leftrightarrow (\phi * \forall x\psi) \text{ and } \vdash \exists x(\phi * \psi) \leftrightarrow (\phi * \exists x\psi)$$

(3) If x is not free in ϕ,

$$\vdash \exists x(\psi \to \phi) \leftrightarrow (\forall x\psi \to \phi) \text{ and } \vdash \forall x(\psi \to \phi) \leftrightarrow (\exists x\psi \to \phi)$$

Proof. (1) We prove the first of the two equivalences:

(1)	$\neg\forall x\phi \vdash \neg\forall x\phi$	Taut
(2)	$\neg\forall x\phi \neg\forall x\neg\neg\phi$	Equiv. Repl. (1)
(3)	$\neg\forall x\phi \vdash \exists x\neg\phi$	
(4)	$\vdash \neg\forall x\phi \to \exists x\neg\phi$	DT (3)

The converse is similar. The second provable equivalence follows by duality.

(2) Say x is not free in ϕ. We consider only the case where $*$ is $\to$.

(1)	$\vdash (\forall x\psi \to \psi) \to [(\phi \to \forall x\psi) \to (\phi \to \psi)]$	Taut.
(2)	$\vdash \forall x\psi \to \psi$	Sub
(3)	$\vdash (\phi \to \forall x\psi) \to (\phi \to \psi)$	MP (1), (2)
(4)	$\vdash (\phi \to \psi) \to \forall x(\phi \to \psi)$	$\forall$Rule
(5)	$\vdash (\phi \to \forall x\psi) \to \forall x(\phi \to \psi)$	$\text{Comp}_{\to}$ (3), (4)

It follows from one of our axiom schemes that $\vdash \forall x(\phi \to \psi) \to (\phi \to \forall x\psi)$, which completes the proof of the first equivalence. To prove the second, we proceed as follows:

(1)	$\vdash (\psi \to \exists x\psi) \to [(\phi \to \psi) \to (\phi \to \exists x\psi)]$	Taut
(2)	$\vdash \psi \to \exists x\psi$	Sub
(3)	$\vdash (\phi \to \psi) \to (\phi \to \exists x\psi)$	MP (1), (2)
(4)	$\vdash \exists x(\phi \to \psi) \to (\phi \to \exists x\psi)$	$\exists$Rule

The other direction is slightly more involved.

(1)	$\vdash \phi \to (\psi \to \phi)$	Taut
(2)	$\vdash \exists x\phi \to \exists x(\psi \to \phi)$	$Sub, \exists$Dist
(3)	$\vdash (\phi \to \psi) \to \exists x(\phi \to \psi)$	Unsub
(4)	$\neg\phi \to (\phi \to \psi)$	Taut
(5)	$\neg\phi \to \exists x(\phi \to \psi)$	$\text{Comp}_\to(3), (4)$
(6)	$\vdash (\neg\phi \to \exists x(\phi \to \psi))$ $\to [(\exists x\phi \to \exists x(\phi \to \psi)) \to ((\phi \to \exists x\psi) \to \exists x(\phi \to \psi))]$	Taut
(7)	$\vdash (\phi \to \exists x\psi) \to \exists x(\phi \to \psi)$	MP $\times$ 2 (2), (5), (6)

(3) For the first equivalence:

(1)	$\vdash (\forall x\psi \to \phi) \leftrightarrow (\neg\phi \to \neg\forall x\phi)$	Taut
(2)	$\vdash (\neg\phi \to \neg\forall x\neg\neg\psi) \leftrightarrow (\neg\phi \to \neg\forall x\phi)$	Equiv. Repl.
(3)	$\vdash (\forall x\psi \to \phi) \leftrightarrow (\neg\phi \to \exists x\neg\psi)$	$\text{Comp}_\to$ (1), (2), Simpl.
(4)	$\vdash (\neg\phi \to \exists x\neg\psi) \leftrightarrow \exists x(\neg\phi \to \neg\psi)$	Part (2) of the theorem
(5)	$\vdash \exists x(\neg\phi \to \neg\psi) \leftrightarrow \exists x(\psi \to \phi)$	Equiv. Repl.
(6)	$\vdash (\forall x\psi \to \phi) \leftrightarrow \exists x(\psi \to \phi)$	$\text{Comp}_\to$

We leave the proof of the second equivalence as an exercise.

□

Another important tool will be the renaming of bound variables, which is an extremely common practice in mathematics—as long as we do not create any confusion by renaming variables, we will not change the essential meaning of the expression. We need to be careful, though:

Example 3.7. We consider possible renamings of the variable y in the formula $\phi = \forall x\exists y(y = Sx)$ in the language of the natural numbers with successor. If we rename y to z, we have $\forall x\exists z(z = Sx)$, which is clearly equivalent to ϕ. If we rename y to x, on the other hand, we have $\forall x\exists x(x = Sx)$, which is nonsense at best, and a contradiction at worst. As ϕ is certainly not a contradiction, this second renaming is problematic: the new formula we obtain is fundamentally different than the original.

Definition 3.8. Let ϕ be a formula. We say that ϕ' is a *variant* of ϕ if it arises from ϕ by replacing one or more subformulas of ϕ of the form $Qx\psi$ (Q either $\forall$ or $\exists$) by

$$Qy\psi_x(y)$$

where y is not free in ψ.

Remark 3.9. This condition blocks the problematic renaming in the preceding example: in the formula $\forall x\exists y(y = Sx)$, renaming y to x would involve the subformula $\exists y(y = Sx)$—since x occurs free, we cannot replace y by x. If we did, we would no longer have a variant of the original formula.

As an easy consequence of Equivalent Replacement (Theorem 3.4 above),

Corollary 3.10. If ϕ' is a variant of a formula ϕ, it is provably equivalent; that is,

$$\vdash \phi \leftrightarrow \phi'.$$

Proof. By the Equivalent Replacement Theorem, we need only verify that the subformulas $Qx\psi$ and $Qy\psi_x(y)$ are provably equivalent, assuming x is not free in ψ. Take Q to be $\forall$.

$$\begin{array}{lll} (1) & \vdash \forall x\psi \to \psi_x(y) & \text{Sub} \\ (2) & \vdash \forall x\psi \to \forall y\psi_x(y) & \forall\text{Rule} \end{array}$$

We are able to apply the $\forall$Rule, since x is not free in ψ.

Going in the other direction, notice that x is not free (trivially) in $\psi_x(y) = \chi$, so is substitutable for y in χ. So, as above, $\vdash \forall y\chi \to \forall x\chi_y(x)$. That is,

$$\vdash \forall y\psi_x(y) \to \forall x\psi.$$

□

Remark 3.11. The best way to avoid renaming conflicts is simply to use a fresh variable, which does not occur anywhere in the formula in question!

3.2. **Computing Prenex Forms.** We are now prepared to show the following:

Theorem 3.12. For every formula ϕ, it is possible to build a prenex formula ϕ' with $\vdash \phi \leftrightarrow \phi'$.

We give the algorithm, and leave the details of the proof as an exercise (see Exercise 3.1 below). Some sources may carry out the steps in a different order than we use here—that is just fine.

Algorithm: Our input is a formula ϕ.

(1) Remove unnecessary quantifiers: in any subformula $Qx\psi$, delete Qx if there is no free occurrence of x in ψ.

(2) "Standardize the variable apart." Here we rename any variables that occur both free and bound in ϕ. One can begin renaming these variables from the right or the left: we choose the right.

Example 3.13. Consider the formula

$$[\forall x(\phi(x) \to \phi(x))] \wedge [\exists x\psi(x)] \wedge [\exists z\phi(z)] \wedge [\exists z(\psi(z) \to \chi(x))].$$

Consider the variable x. We leave its rightmost occurrence (free, in the last conjunct) as is, then move to the right, systematically renaming other instances of x:

$$[\forall x_2(\phi_x(x_2) \to \phi_x(x_2))] \wedge [\exists x_1\psi(x_1)] \wedge [\exists z\phi(z)] \wedge [\exists z(\psi(z) \to \chi(x))].$$

We still need to fix z. It appears bound in two distinct subformulas, namely the last two conjuncts. We leave the rightmost as is, but rename z in the other:

$$[\forall x_2(\phi_x(x_2) \to \phi_x(x_2))] \wedge [\exists x_1\psi(x_1)] \wedge [\exists z_1\phi(z_1)] \wedge [\exists z(\psi(z) \to \chi(x))].$$

The outcome of this step should be a formula in which no variable appears in multiple quantifiers, and no variable occurs both free and bound.

(3) Get rid of all $\leftrightarrow$s. Replace any subformula of the form $\phi \leftrightarrow \psi$ with $\phi \to \psi \wedge \psi \to \phi$. This is essential, as the rules for pulling quantifiers out of conditionals are clear-cut (see Theorem 3.6), but the situation is hopelessly messy for biconditionals.

(4) Move all of the negations inward, until they apply only to atomic formulas. For this we use, recursively, the familiar equivalences that we have already worked out:

$$\begin{array}{ccc} \neg(\forall x\phi) & \rightsquigarrow & \exists x\neg\phi \\ \neg(\exists x\phi) & \rightsquigarrow & \forall x\neg\phi \\ \neg(\phi\to\psi) & \rightsquigarrow & \phi\wedge\neg\phi \\ \neg(\phi\vee\psi) & \rightsquigarrow & \neg\phi\wedge\neg\psi \\ \neg(\phi\wedge\psi) & \rightsquigarrow & \neg\phi\vee\neg\psi \\ \neg\neg\phi & \rightsquigarrow & \phi \end{array}$$

(5) Move all of the quantifiers to the left, until you hit the quantifier prefix, using some of our tricks from Theorem 3.6: for Q either $\forall$ or $\exists$,

$$\begin{array}{ccc} (Qx\phi)\vee\psi & \rightsquigarrow & Qx(\phi\vee\psi) \\ (Qx\phi)\wedge\psi & \rightsquigarrow & Qx(\phi\wedge\psi) \\ (Qx\phi)\to\psi & \rightsquigarrow & \bar{Q}x(\phi\to\psi) \\ \phi\to Qx\phi & \rightsquigarrow & Qx(\phi\to\psi) \end{array}$$

Here $\bar{Q}$ denotes the quantifier dual to Q; that is,

$$\bar{Q}=\begin{cases}\forall & \text{if } Q=\exists\\ \exists & \text{if } Q=\forall\end{cases}$$

On a more technical note, if you are interested in minimizing the number of variables used, there are a few additional tricks:

$$\begin{array}{ccc} (\exists x\phi)\vee(\exists y\psi) & \rightsquigarrow & \exists x(\phi\vee\psi_y(x)) \\ (\forall x\phi)\wedge(\forall x\psi) & \rightsquigarrow & \forall x(\phi\wedge\psi_y(x)) \end{array}$$

This point will not be emphasized here.

After carrying out steps (1)-(5), you should be left with a formula in prenex form!

Example 3.14. (1) Let P, R, and S be unary predicate symbols, and let

$$\phi=\forall x[(\exists yPy\wedge\forall y\neg Sy)\to\neg(\exists yPy\wedge Rx)].$$

Running the algorithm, we have

$$\begin{aligned} &\forall x[(\exists yPy\wedge\forall y\neg Sy)\to\neg(\exists yPy\wedge Rx)]\\ &\rightsquigarrow \forall x[\neg(\exists yPy\wedge\forall y\neg Sy)\vee\neg(\exists yPy\wedge Rx)]\\ &\rightsquigarrow \forall x[(\forall y\neg Py\vee\exists ySy)\vee(\forall y\neg Py\vee\neg Rx)]\\ &\rightsquigarrow \forall x[(\forall y_2\neg Py_2\vee\exists y_1Sy_1)\vee(\forall y\neg Py\vee\neg Rx)]\\ &\rightsquigarrow \forall x[\forall y_2\exists y_1(\neg Py_2\vee\exists y_1Sy_1)\vee\forall y(\neg Py\vee\neg Rx)]\\ &\rightsquigarrow \forall x\forall y_2\exists y_1\forall y[(\neg Py_2\vee Sy_1)\vee(\neg Py\vee\neg Rx)]\\ &\rightsquigarrow \forall x\forall y_2\exists y_1\forall y[\neg Py_2\vee Sy_1\vee\neg Py\vee\neg Rx] \end{aligned}$$

The final line is a prenex form of ϕ. Note that we could have pulled the quantifiers for the y-variables out in a different order, giving (nominally) different prenex forms of ϕ—there is no rule, on the other hand, that would allow us to pull any of them through the quantifier $\forall x$.

(2) If F, G, and H are ternary, binary, and unary predicate symbols, respectively,

$$\begin{aligned}
&\neg(\forall x\exists y F(u,x,y) \to \exists x(\neg\forall y G(y,v) \to H(x))\\
&\rightsquigarrow \forall x\exists y F(u,x,y) \wedge \neg\exists x(\neg\forall y G(y,v) \to H(x))\\
&\rightsquigarrow \forall x\exists y F(u,x,y) \wedge \forall x\neg(\neg\forall y G(y,v) \to H(x))\\
&\rightsquigarrow \forall x\exists y F(u,x,y) \wedge \forall x(\forall y\neg G(y,v) \wedge \neg H(x))\\
&\rightsquigarrow \forall x_1\exists y F(u,x_1,y) \wedge \forall x(\forall y\neg G(y,v) \wedge \neg H(x))\\
&\rightsquigarrow \forall x_1\exists y_1 F(u,x_1,y_1) \wedge \forall x(\forall y\neg G(y,v) \wedge \neg H(x))\\
&\rightsquigarrow \forall x_1\exists y_1 F(u,x_1,y_1) \wedge \forall x\forall y(\neg G(y,v) \wedge \neg H(x))\\
&\rightsquigarrow \forall x_1\exists y_1\forall x\forall y[F(u,x_1,y_1) \wedge \neg G(y,v) \wedge \neg H(x)]
\end{aligned}$$

(3) Let P, R, and S be binary predicate symbols. Then as an application of the algorithm, we have the rewriting

$$\begin{aligned}
&\forall y(\exists x Pxy \to \exists u Ryu) \to \forall x Sxy\\
&\rightsquigarrow \forall y(\exists x_1 Px_1 y \to \exists u Ryu) \to \forall x Sxy\\
&\rightsquigarrow \forall y_1(\exists x_1 Px_1 y_1 \to \exists u Ry_1 u) \to \forall x Sxy\\
&\rightsquigarrow \forall x[\forall y_1(\exists x_1 Px_1 y_1 \to \exists u Ry_1 u) \to Sxy]\\
&\rightsquigarrow \forall x\exists y_1[(\exists x_1 Px_1 y_1 \to \exists u Ry_1 u) \to Sxy]\\
&\rightsquigarrow \forall x\exists y_1[\forall x_1(Px_1 y_1 \to \exists u Ry_1 u) \to Sxy]\\
&\rightsquigarrow \forall x\exists y_1[\forall x_1\exists u(Px_1 y_1 \to Ry_1 u) \to Sxy]\\
&\rightsquigarrow \forall x\exists y_1\forall x_1[\exists u(Px_1 y_1 \to Ry_1 u) \to Sxy]\\
&\rightsquigarrow \forall x\exists y_1\forall x_1\exists u[(Px_1 y_1 \to Ry_1 u) \to Sxy]
\end{aligned}$$

We will consider further examples in the exercises. In any case, this is the algorithm for reducing a formula to prenex form, which ultimately serves as a proof of Theorem 3.12.

3.3. **Skolem and Herbrand Normal Forms.** We now consider a few more rewriting algorithms that build on the prenex algorithm just discussed:

Skolemization	**Herbrandization**
• Eliminates $\exists$s	• Eliminates $\forall$s
• Preserves satisfiability	• Preserves validity

In each of these algorithms—which are dual to eachother—we simplify the quantifier prefixes of formulas in a language L by passing to a larger language L'. Model theorists Skolemize[3]; proof theorists Herbrandize[4]. Both have their uses, but Skolemization is rather more intuitive, and more closely connected to tricks we will need in the next chapter for the proof of the Completeness Theorem—and I am a model theorist!—so we will devote most of our energy to that variant.

Definition 3.15. We say that a formula is in *Skolem normal form* if it is in prefix normal form and contains only universal quantifiers.

[3]Thoralf Skolem, 1887-1963.

[4]Jacques Herbrand, 1908-1931. An extremely short, extremely productive mathematical career. Made essential contributions to proof theory, in particular.

Skolemization, which we now describe in detail, is an algorithm that takes as its input a formula ϕ in prenex normal form, and outputs a formula ϕ_S in Skolem normal form (that is, with no existential quantifiers).

The idea of the algorithm is that we can expand our language by a family of new constant and function symbols that provide witnesses to all existential statements: if we have a statement of the form "for all x, there is y such that...," we will incorporate a function symbol that gives us, for each x, a y with the desired property. To be precise, we proceed by cases:

Example 3.16 (Simplest case)**.** Consider a formula $\exists\phi(x)$ in language L, where ϕ is quantifier-free. We will introduce a new constant symbol (or, if you prefer, a new 0-ary function symbol) c_ϕ to serve as a witness for $\phi(x)$. Set $L' = L \cup \{c_phi\}$.

Remark 3.17. Notice that if $\mathcal{M}$ is an L-structure with $\mathcal{M} \models_L \exists x\phi(x)$, we can expand M to an L'-structure $\mathcal{M}'$, interpreting c_ϕ as an element of M witnessing the truth of the existential. So $M' \models_{L'} \phi(c_\phi)$.

This works in the other direction as well: given an L'-structure $\mathcal{M}'$ with $\mathcal{M}' \models \phi(c_\phi)$, the model $\mathcal{M} = \mathcal{M} \restriction L$ satisfies $\mathcal{M} \models_L \exists x\phi(x)$ since, in particular, $\mathcal{M} \models_L \phi[c_\phi^{\mathcal{M}'}/x]$.

In short, $\exists x\phi(x)$ is satisfiable if and only if $\phi(c_\phi)$ is satisfiable; that is, Skolemization *preserves* (the "only if" direction) and *reflects* (the "if" direction) satisfiability.

We will often omit the ϕ in the subscript, or choose a more manageable abbreviation, particularly when there are multiple variable to be replaced in the same formula. For example:

Example 3.18 (Existential prefix)**.** Consider a formula

$$\exists x_1 \exists x_2 \dots \exists x_n \phi(x_1, x_2, \dots, x_n).$$

We expand the language with a new constant symbol for each of the existentially quantified variables: $L' = L \cup \{c_{x_1}, c_{x_2}, \dots, c_{x_n}\}$. As before, this allows us to consider satisfiability in L'-structures of $\phi(c_{x_1}, c_{x_2}, \dots, c_{x_n})$, rather than satisfiability of $\exists x_1 \exists x_2 \dots \exists x_n \phi(x_1, x_2, \dots, x_n)$ in L-structures.

Things are more difficult when an existential quantifier occurs in the scope of a universal:

Example 3.19 (Universal prefix)**.** Say we have a formula $\forall x \exists y \phi(x, y)$; that is, "for all x, there exists a y such that..." Here the value of y witnessing the truth of the existential will depend on the value of x: for every x, a potentially different y. In short, we should think of our witnesses as being supplied by a function of x! To that end, we introduce a new unary function symbol f_ϕ to the language, and identify

$$\forall x \exists y \phi(x, y) \qquad \text{with} \qquad \forall x \phi(x, f_\phi(x)).$$

In case the existential is in the scope of multiple universals, we proceed similarly: given

$$\forall x_1 \forall x_2 \dots \forall x_n \exists y \phi(x_1, x_1, \dots, x_n, y)$$

we add an n-ary function symbol f_ϕ, and replace the original formula with

$$\forall x_1 \forall x_2 \dots \forall x_n \phi(x_1, x_1, \dots, x_n, f_\phi(x_1, x_2, \dots, x_n))$$

Remark 3.20. By an argument like the one in Example 3.17, one can show satisfiability is preserved and reflected by these rewritings, as well.

In general, we may have more complicated alternations of quantifiers, in which case we will need a combination of the techniques above.

Example 3.21. (1) Consider, for example, the formula

$$\psi = \exists x \forall y \exists z \phi(x, y, z).$$

We work our way in from the left. First, we have an existential which is not in the scope of a universal: we add a constant c_x, arriving at

$$\forall y \exists z \phi(c_x, y, z).$$

The only remaining existential quantifier is in the scope of the single universal quantifier $\forall y$, meaning we must add a function symbol, f_ϕ, and pass to the formula

$$\forall y \phi(c_x, y, f_\phi(y)).$$

We have eliminated all of the existential quantifiers, and this new formula—a Skolem formula—is, by design, satisfiable if and only if ψ is satisfiable!

(2) Consider

$$\psi = \forall x_1 \exists x_2 \forall x_3 \forall x_4 \exists x_5 \phi(x_1, x_2, x_3, x_4, x_5).$$

Working our way in from the left, the rewriting proceeds as follows:

$$\begin{aligned}
&\forall x_1 \exists x_2 \forall x_3 \forall x_4 \exists x_5 \phi(x_1, x_2, x_3, x_4, x_5) \\
&\rightsquigarrow \forall x_1 \forall x_3 \forall x_4 \exists x_5 \phi(x_1, f_{\forall x_3 \forall x_4 \exists x_5 \phi}(x_1), x_3, x_4, x_5) \\
&\rightsquigarrow \forall x_1 \forall x_3 \forall x_4 \phi(x_1, f_{\forall x_3 \forall x_4 \exists x_5 \phi}(x_1), x_3, x_4, f_\phi(x_1, x_3, x_4))
\end{aligned}$$

Here again, we have removed all of the existential quantifiers, thereby obtaining a Skolem formula, in this case by adding a unary function symbol $f_{\forall x_3 \forall x_4 \exists x_5 \phi}$ and a ternary function symbol f_ϕ. Note that the subscripts are very burdensome: once we are confident in our mastery of this process, we can call the new function symbols by more manageable names, e.g. f and g. This, of course, will make no fundamental difference!

Remark 3.22. If you would prefer a more unified treatment, you could think of constants as 0-ary function symbols (in Example 3.21(1), for example, we would use a 0-ary function symbol f_ϕ in place of c_x). This allows us to handle the cases of existential and universal in the same way—by adding function symbols—but strikes me as less intuitive than the version discussed above.

Definition 3.23. We refer to the new function symbols f_ϕ as *Skolem functions*. We say that a theory (that is, a set of sentences) is *Skolem* if there is a Skolem function for any subformula of $\phi \in T$. In this case we can eliminate existentials without changing the language!

Generally, we will need to add symbols, of course. The algorithm gives:

Theorem 3.24 (Skolem Normal Form)**.** Let ϕ be a formula in language L with no free variables, and in prenex normal form. There exists a Skolem formula ϕ^S in a language $L' \supseteq L$, where L'Ł consists of finitely many new constant and function symbols, such that ϕ is satifiable if and only if ϕ^S is satisfiable.

We say that ϕ^S is a *Skolem normal form* of ϕ. *Skolemization* can refer to this rewriting process for a single formula ϕ, or for all formulas in a given theory T. In the latter case, we obtain a Skolem theory T^S.

Herbrandization is precisely dual to Skolemization, and involves eliminating universal quantifiers by expanding the vocabulary. This is far less intuitive, at least for me, as it preserves and reflects *validity* rather than *satisfiability*, but it is very easy to describe. Our goal is Herbrand normal form:

Definition 3.25. We say that a formula is in *Herbrand normal form* if it is in prenex form and does not contain any universal quantifiers.

Notes 3.26. We consider the basic cases of the Herbrand rewriting, as we did for Skolemization.

(1) A formula with universal prefix $\forall x_1 \phi(x_1)$ is rewritten as $\phi(c_1)$, c_1 a new constant symbol. If there are multiple universal quantifiers in the prefix, we add a new constant for each one.
(2) When a universal quantifier is in the scope of an existential, as in $\exists x_1 \forall x_2 \phi(x_1, x_2)$, we add a new function symbol: $\exists x_1 \phi(x_1, g_\phi(x_1))$. More generally,
$$\exists x_1 \exists x_2 \dots \exists x_n \forall y \phi(x_1, x_2, \dots, x_n, y)$$
is rewritten as
$$\exists x_1 \exists x_2 \dots \exists x_n \forall y \phi(x_1, x_2, \dots, x_n, g_\phi(x_1, x_2, \dots, x_n)).$$
(3) For more complicated alternations of quantifiers, we use a combination of the above, e.g.
$$\begin{aligned} &\forall x_1 \exists x_2 \forall x_3 \exists x_4 \phi(x_1, x_2, x_3, x_4) \\ &\rightsquigarrow \exists x_2 \forall x_3 \exists x_4 \phi(c_{x_1}, x_2, x_3, x_4) \\ &\rightsquigarrow \exists x_2 \exists x_4 \phi(c_{x_1}, x_2, g_{\exists x_4 \phi}(x_2), x_4) \end{aligned}$$

Dually to Theorem 3.24, we have:

Theorem 3.27 (Herbrand normal form)**.** If ϕ is a formula in language L, there is a formula ϕ^H in $L' \supseteq L$, $L' \setminus L$ consisting of finitely many constant and function symbols, that is in Herbrand normal form and with the property that ϕ is valid if and only if ϕ^H is valid.

An important extension of this theory reveals that universal formulas are contradictions if and only if the contradiction is fundamentally syntactic; that is, we can obtain a contradiction by the substitution of terms for the universally quantified variables. To be precise:

Theorem 3.28 (Herbrand's Theorem)**.** If $\phi(x_1, \dots, x_n)$ has x_1, …, x_n free, the formula
$$\forall x_1 \dots \forall x_n \phi(x_1, \dots, x_n)$$
is a contradiction if and only if there are terms
$$t_1^1, \dots, t_n^1, t_1^2, \dots, t_n^2, \dots, t_1^m, \dots, t_n^m$$
so that
$$\phi(t_1^1, \dots, t_n^1) \wedge \phi(t_1^2, \dots, t_n^2) \wedge \dots \wedge \phi(t_1^m, \dots, t_n^m)$$
is a contradiction.

It should be noted that this theorem is often (in fact, almost always) stated in the dual form, where *validity* of a purely *existential* statement in prenex form is witnessed by a *disjunction* of substitution instances. Modern versions of the proof of Herbrand's Theorem employ the tools of sequent calculus (specifically, cut-elimination), which is well beyond the scope of this course.

Chapter 3 Exercises

Exercise 3.1. Prove the following:

Theorem: For any formula ϕ, there is a formula ϕ' in prenex normal form such that $\vdash \phi \leftrightarrow \phi'$.

(Hint: Proof by induction on complexity of ϕ. Clear for atomic formulas. Consider induction steps: $\phi = \neg\psi$, $\phi = \psi \to \chi$, $\phi = \forall x\psi$.)

Exercise 3.2. Convert each of the following formulas into prenex normal form, where P, Q, and R are binary relation symbols, S is a ternary relation symbol, and T and U are unary relation symbols.

(i) $\exists x Rxy \leftrightarrow \forall y Pxy$.

(ii) $\forall y(\exists x Pxy \to Qyz) \wedge \exists y(\forall x Rxy \vee Qxy)$

(iii) $\forall x(\forall y(\forall z(Sxyz \wedge Ty) \to \forall x Pxz))$

(iv) $\neg(\forall x \exists y Pxy \to \exists x \exists y Rxy) \wedge \forall x(\neg\exists y Qxy)$

(v) $(\forall x \exists y Qxy \vee \exists x \forall y Rxy) \wedge \forall x(\neg\exists y Qxy)$

(vi) $\exists x(Tx \wedge \forall y(Uy \to Qxy))$

(vii) $\forall x(Tx \to \forall y([Ty \to (Ux \to Uy)] \vee \forall z Tz))$

Exercise 3.3. Use the prenex rules to prove each of the following:

(i) $\vdash \forall x \exists y(Tx \to Ty)$

(ii) $\vdash \exists x \forall y(Tx \to Ty)$

(iii) $\forall x \exists y(Tx \to Rxy)$, $\exists x Tx \vdash \exists x \exists y Rxy$

Exercise 3.4. Give Skolem and Herbrand normal forms of the following formulas:

(i) $\exists x(Tx \wedge \forall y(Uy \to Qxy))$

(ii) $\forall x(Tx \to \forall y([Ty \to (Ux \to Uy)] \vee \forall z Tz))$

(iii) $\forall x(\forall y(\forall z(Sxyz \wedge Ty) \to \forall x Pxz))$

(iv) $(\forall x \exists y Qxy \vee \exists x \forall y Rxy) \wedge \forall x(\neg\exists y Qxy)$

(v) $\forall x \forall y[Rxy \to \exists z(Rxz \wedge Rzy)]$.

Exercise 3.5. (i) Compute a Skolem normal form ϕ^S of the formula $\phi = \forall x \exists y(x < y)$ (with ϕ, obviously, in the language of ordered sets).

(ii) Show that ϕ and ϕ^S are equisatisfiable: there is an L-structure $\mathcal{M}$ such that $\mathcal{M} \models \phi$ iff there is a structure $\mathcal{M}'$ in the expanded language such that $\mathcal{M}' \models \phi^S$.

(iii) Show that it is **not the case** that $\vdash_{L'} \phi \leftrightarrow \phi^S$. That is, ϕ and ϕ^S are not logically equivalent.

4. The Completeness Theorem

We are now in a position to prove one of the most important results in first-order logic:

Theorem 4.1 (Completeness Theorem, Gödel[5])**.** Let T be an L-theory. Then the following equivalent conditions hold:

(1) For any $\phi \in L$, $T \vdash \phi$ if and only if $T \models \phi$.
(2) T is consistent if and only if it has a model.

The statement (1) reassures us that we can work syntactically or semantically, and there is no possibility of harm: while soundness tells us that we cannot prove anything that is not true, completeness gives us the converse as well. That is, anything that is true (semantically) in every model of a theory T is provable (syntactically) from T.

The proof of this theorem is a very long project. We will prove statement (2), and that (1) and (2) are in fact equivalent. We begin with this latter point:

Proof. (1)⇒(2): Assume (2).

- Say $T \vdash \phi$. By soundness, $T \models \phi$. So we are done.
- Say $T \models \phi$. Let ϕ' be the closure of ϕ. As we have seen, $T \models \phi'$ as well. This means that for any $\mathcal{M} \models T$, $\mathcal{M} \models \phi'$; that is, any model of T believes ϕ', and does *not* believe $\neg\phi'$. Thus the set $T \cup \{\neg\phi'\}$ does not have a model. By (2), then, $T \cup \{\neg\phi'\}$ is inconsistent, from which we infer that $T \vdash \phi'$. By an earlier exercise, $T \vdash \phi$ as well.

(2)⇒(1): Assume (1).

- Say T has a model. Then it is consistent, by the Soundness Theorem.
- Say T has no model. Then, vacuously, any model of T satisfies ϕ and $\neg\phi$ for any ϕ. That is, $T \models \phi$ and $T \models \neg\phi$. By (1), $T \vdash \phi$ and $T \vdash \neg\phi$. Since ϕ was arbitrary, this means that the theory T proves absolutely everything. That is, T is inconsistent.

□

We now move to the larger task: proving (2). Note that by soundness, if T has a model, it must be consistent. This means that we need only prove the converse:

$$T \text{ consistent } \Rightarrow T \text{ has a model}$$

The antecedent expresses, intuitively, that there is *no obstacle to having a model*; the consequent, that there is in fact a model, seems much stronger!

We will require a number of concepts in this proof, some of which will be familiar.

Definition 4.2. Let T be a theory.

[5]Kurt Gödel, 1906-1978. Born in Brno at Pekařská 3, near Šilingrovo náměstí. Grew up at Pellicova 8a, below Špilberk. The most influential logician of the modern era, about whom we will have much more to say.

(1) We say that T is *complete* if it is consistent and for any sentence ϕ, $T \vdash \phi$ or $T \vdash \neg\phi$.
(2) We say that T is *maximally consistent* if it has no consistent extensions: for any $\phi \notin T$, the sets
$$T \cup \{\phi\} \text{ and } T \cup \{\neg\phi\}$$
are inconsistent. That is, if we add anything to T, we will get a contradiction.
(3) We say that T is *deductively closed* if for any ϕ with $T \vdash \phi$, $\phi \in T$. That is, any consequence of T is already in T.
(4) We say that the theory T is *Henkin* if for any formula $\phi(x)$, there is a constant symbol c such that $T \vdash \exists x\phi(x) \to \phi(c)$. We think of c as a *witness* to the truth of the existential formula $\exists\phi(x)$. We could summarize this condition by saying that a theory T is Henkin if it has *enough witnesses*[6].

Example 4.3. Let L be the language of ordered sets, containing a single binary relation symbol R, and let T_{po} be the theory of partial orders, i.e.

$$T_{PO} = \{\forall xRxx, \forall x\forall y(Rxy \wedge Ryx \to x = y), \forall x\forall y\forall z([Rxy \wedge Ryz] \to Rxz)\}.$$

- This theory is consistent, since it has a model: consider, for example, the partially ordered set of natural numbers, $\langle \mathbb{N}, \leq^{\mathbb{N}} \rangle$.
- The theory T_{PO} is not complete: consider the sentence capturing density of a partial order:
$$\phi_d = \forall x\forall y(xRy \to \exists z[xRz \wedge zRy]).$$
There are dense partial orders (e.g. $\langle \mathbb{Q}, \leq^{\mathbb{Q}} \rangle$) and non-dense partial orders (e.g. $\langle \mathbb{N}, \leq^{\mathbb{N}} \rangle$), so both $T_{PO} \cup \{\phi_d\}$ and $T_{PO} \cup \{\neg\phi_d\}$. It follows that $T \nvdash \neg\phi_d$ and $T_{PO} \nvdash \phi_d$, respectively.
- From the preceding argument, it is clear that T_{PO} is not maximally consistent: we could add ϕ_d to T_{PO} and still have a consistent set.
- The theory T_{PO} is not deductively complete, of course, since the set of its consequences will be infinite, while $|T_{PO}| = 3$.

Example 4.4. Let L be the language of ordered sets, as above, and let T_{DLO} be the theory of dense linear orders without endpoints,

$$T_{DLO} = T_{PO} \cup \{\forall x\forall y(xRy \vee yRx)\} \cup \{\phi_d\} \cup \{\neg\exists x\forall y(x \neq y \to xRy), \neg\exists y\forall x(x \neq y \to yRx)\}$$

Although it is not so easy to see[7], this extension of T_PO is complete! By adding these extra conditions, we have resolved all possible outstanding questions.

We include a few additional, and useful, examples of complete theories:

Example 4.5.
- T_{DIV}, the theory of divisible abelian groups. Recall that an abelian group is *divisible* if it satisfies
$$\forall x\exists y(ny = x)$$
for any $n \in \mathbb{N}$.

[6]Can you detect any relationship between this idea and the Skolem normal forms discussed previously?

[7]The argument uses Vaught's Test: a theory is complete if it has no finite models and just one model of some infinite size. The theory T_{DLO} has no finite models, as one can easily see, and has only one countable model up to isomorphism: $\langle \mathbb{Q}, \leq^{\mathbb{Q}} \rangle$.

- T_{ACF_p}, the theory of algebraically closed fields of characteristic p.
- For any L-structure $\mathcal{M}$, we define $\mathrm{Th}_L(\mathcal{M})$ to be the set of all L-sentences ϕ such that $\mathcal{M} \models \phi$. For any $\mathcal{M}$, $\mathrm{Th}_L(\mathcal{M})$ is complete. Note: we call $\mathrm{Th}_L(\mathcal{M})$ the *complete theory of* $\mathcal{M}$.

Question 4.6. Is elementary arithmetic (in the form, say, of Peano arithmetic, Definition 6.5) complete?

This is a very big question, which will ultimately be answered, negatively, by Gödel's Incompleteness Theorems—this will have to wait until the final lecture of the course.

The property of deductive closure is nothing to worry about, really. In fact, model theorists tend to assume that all theories are deductively closed, since we can always just take the *deductive closure*:

$$T \mapsto \overline{T} = T \cup \{\phi \in \mathbf{Sent}_L \mid T \vdash \phi\}$$

It is time to prove (2) of the Completeness Theorem. Fix a consistent theory T. We will proceed according to the following outline:

(I) We expand T to a consistent Henkin theory T', through the careful addition of witnessing constants and new axioms.
(II) We show that any consistent theory, including T', can be extended to a maximally consistent theory T^*.
(III) We show that any maximally consistent Henkin theory, like T^*, has a model.

Part (I)

A little basic terminology:

Definition 4.7. (1) A language L' is an *extension* of a language L if every nonlogical symbol of L is in L'.
(2) A theory T' in L' is an *extension* of an L-theory T if for every $\phi \in L$, $T \vdash \phi$ implies $T' \vdash \phi'$.
(3) We say that T' is a *conservative extension* of T if for all $\phi \in L$,

$$T \vdash \phi \text{ if and only if } T' \vdash \phi.$$

Put another way, T' is a conservative extension of T if it cannot prove any new L-sentences.

To go from consistent theory T to consistent Henkin theory T', we proceed in stages.

- Define L_1 to be an extension of L where, for every existential formula $\exists x \phi(x)$ in L, we add a new constant symbol c_ϕ:

$$L_1 = L \cup \{c_\phi \mid \exists x \phi(x) \in L\}$$

- Define T_1 to be the extension of T in L_1 obtained by adding all axioms of the form $\exists x \phi(x) \leftrightarrow \phi(c_\phi)$. Notice that T_1 has witnesses for all existential formulas in L, but there will be new existential formulas in L_1—we will have to add witnesses for these as well. So:
- Define L_2 to be the extension of L_1 obtained by adding new constant symbols c_ϕ for all $\exists x \phi(x)$ in L_1. Let T_2 be the extension of T_1 obtained by adding axioms $\exists x \phi(x) \leftrightarrow \phi(c)$ for each $\exists x \phi(x)$ in L_1. Here we have added

witnesses for existential formulas in L_1, but will still need to worry about those in $L_2 \setminus L_1$. So the process continues.

- Given L_n and T_n, we extend to L_{n+1} and T_{n+1} in the same manner, obtaining sequences of extensions

$$L \subseteq L_1 \subseteq \cdots \subseteq L_n \subseteq L_{n+1} \subseteq \ldots$$

$$T \subseteq T_1 \subseteq \cdots \subseteq T_n \subseteq T_{n+1} \subseteq \ldots$$

- Define

$$L' = \bigcup_{n=1}^{\infty} L_n \qquad \text{and} \qquad T' = \bigcup_{n=1}^{\infty} T_n.$$

We should check that T' is Henkin, and a conservative extension of T.

To see that T' is Henkin, consider any $\exists x \phi(x)$ in L. Since $\exists x \phi(x)$ is in L, the union of the L_n, it will be contained in L_k for some $k \in \mathbb{N}$. But then, by construction, there is $c_\phi \in L_{k+1}$ with $T_{k+1} \vdash \exists x \phi(x) \leftrightarrow \phi(c_\phi)$. As $T_{k+1} \subseteq T'$, it follows that

$$T' \vdash \exists x \phi(x) \leftrightarrow \phi(c_\phi),$$

as desired. We leave the proof of conservativity as an exercise.

Part (II)

We show that any consistent theory has a maximally consistent extension. Fix T' a consistent theory in language L'. Let

$$X = \{S \subseteq L' \,|\, T' \cup S \text{ is consistent}\}.$$

The set X is partially ordered by subset inclusion, and given any chain

$$S_0 \subseteq S_1 \subseteq \cdots \subseteq S_n \subseteq \ldots$$

in X, the set

$$S = \bigcup_{n=1}^{\infty} S_n$$

is in X: if $T' \cup S$ were inconsistent, there would have to be some finite inconsistent subset of $T' \cup S$. Being finite, it must in fact be contained in $T' \cup S_n$ for some $n \in \mathbb{N}$. But this contradicts the assumption that $S_n \in X$. By Zorn's Lemma (which is, incidentally, equivalent to the *Axiom of Choice*), there is a maximal element $S^* \in X$; that is, for any $S \in X$, $S \subseteq S^*$.

Define $T^* = T' \cup S^*$. This theory is maximally consistent, as one can easily verify.

Remark 4.8. The theory T^* is also complete. To see this, suppose, to the contrary, that $T^* \nvdash \phi$ and $T^* \nvdash \neg\phi$ for some ϕ. Then $T^* \cup \{\neg\phi\}$ is consistent. That is, $T' \cup S^* \cup \{\phi\}$ is consistent, which contradicts maximality of S^*.

Remark 4.9. It is clear, too, that if T' is Henkin, the theory T^*—which has the same language—is also Henkin. Moreover, T^* is a conservative extension of T'.

Part (III)

We show that any complete Henkin theory T^* has a model. This is an incredibly clever—but very finicky—argument. I would encourage you to study it carefully, as a character-building exercise, if nothing else. The idea is that we can construct

a model of the complete Henkin theory *out of the language itself*, with the elements of the model being T^*-provable-equivalence classes of terms!

To be more precise, let V be the set of all variable-free terms in L^*, the language of T^*; that is, V consists of all terms built purely from constant and function symbols in L^*. Given terms τ_1 and τ_2 in V, define

$$\tau_1 \sim \tau_2 \qquad \text{if and only if} \qquad T^* \vdash \tau_1 = \tau_2.$$

Given a term $\tau \in V$, we denote the $\sim$-equivalence class of τ by $[\tau]_\sim$. We define a model $\mathcal{M}$ with underlying set consisting of $\sim$-equivalence classes of terms in V. That is, $M = V/\sim$.

We define the interpretations of the symbols in L^* as follows:

- For any constant symbol c, $c^{\mathcal{M}} = [c]_\sim$.
- For any function symbol f and $[\tau_1]_\sim, [\tau_2]_\sim, \ldots, [\tau_n]_\sim \in M$,
$$f^{\mathcal{M}}([\tau_1]_\sim, [\tau_2]_\sim, \ldots, [\tau_n]_\sim) = [f(\tau_1, \tau_2, \ldots, \tau_n)]_\sim.$$
- For any relation symbol R and $[\tau_1]_\sim, [\tau_2]_\sim, \ldots, [\tau_n]_\sim \in M$,
$$(\tau_1, \tau_2, \ldots, \tau_n) \in R^{\mathcal{M}} \qquad \text{if and only if} \qquad T \vdash R\tau_1\tau_2 \ldots \tau_n.$$

You should check that this structure is a model of T^*. Once that is taken care of, we are done! Why? We have shown that for any consistent theory T, we can, via a sequence of conservative extensions, obtain

$$\begin{array}{ccccc} T & \to & T' & \to & T^* \\ \text{Consistent} & & \text{Consistent, Henkin} & & \text{Complete, Henkin} \\ L & \to & L' & \to & L' \end{array}$$

By part (III), there is an L'-structure $\mathcal{M}' \models T^*$. Let $\mathcal{M}$ be the *reduct of* $\mathcal{M}$ *to* L, denoted $\mathcal{M} = \mathcal{M}' \restriction L$, which is obtained from $\mathcal{M}'$ by simply forgetting all of the interpretations of symbols in $L' \setminus L$ (see the definition below). By conservativity of the extensions $T \subseteq T' \subseteq T^*$,

$$\mathcal{M} \models T.$$

This, at long last, brings us to the end of the proof of the Completeness Theorem.

In connection with the last point, we emphasize:

Definition 4.10. Let L and L' be languages, with $L \subseteq L'$.

(1) Given an L'-structure $\mathcal{M}'$, the *reduct of* $\mathcal{M}'$ *to* L, denoted $\mathcal{M} = \mathcal{M}' \restriction L$, is an L-structure with the same underlying set as $\mathcal{M}'$, which interprets all of the symbols in L in the same way as $\mathcal{M}'$, but forgets the interpretation of all symbols in $L' \setminus L$.
(2) Given an L-structure $\mathcal{M}$, we say that an L'-structure $\mathcal{M}'$ is an *expansion of* $\mathcal{M}$ *to* L if $\mathcal{M} = \mathcal{M}' \restriction L$. That is, $\mathcal{M}'$ is just like $\mathcal{M}$ in every way, except that it also interprets the new symbols in $L' \setminus L$.

This is an exceedingly common phenomenon across mathematics: when considering a particular object, we can view it in different ways, focusing only on those aspects that are of interest in the context of a particular problem.

Example 4.11. Let L be the language of ordered sets, and let L' be the language of the natural numbers with successor. If we doing arithmetic of any kind, we want to consider the natural numbers as an L'-structure (at the very least); that is, we want to work with $\langle \mathbb{N}, 0^{\mathbb{N}}, S^{\mathbb{N}}, \leq^{\mathbb{N}} \rangle$. Sometimes, though, we are concerned only with the order, and forget about any arithmetic structure: in this case, we work with the reduct of $\langle \mathbb{N}, 0^{\mathbb{N}}, S^{\mathbb{N}}, \leq^{\mathbb{N}} \rangle$ to L, namely $\langle \mathbb{N}, \leq^{\mathbb{N}} \rangle$

We can continue in this way: for example, we can think of $\mathbb{Z}$ as a ring, or we can regard it as a group, forgetting the multiplication operation.

Chapter 4 Exercises

Exercise 4.1. Zorn's Lemma: If X is a partially ordered set and for each increasing chain

$$x_0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq \ldots$$

there is an element x_ω with $x_\omega \geq x_n$ for all $n \in \mathbb{N}$, then X contains a largest element.

(i) Show that every vector space V over a field F (use $F = \mathbb{R}$, if you want) has a basis (equivalently, a maximal linearly independent set). [Hint: Let X be the set of all linearly independent subsets of V, ordered by $\subseteq$.]

(ii) Modify the argument to show that any linearly independent subset of a vector space is contained in a basis.

Exercise 4.2. Check that the Henkin structure constructed in part (III) of the proof of the Completeness Theorem is actually a model of the theory.

Exercise 4.3. (i) Show that if T is an L-theory, and f a new function symbol, then if $T \vdash \forall x \exists y \phi(x, y)$, $T \cup \{\forall x \phi(x, f(x))\}$ is a conservative extension of T.

(ii) What, if anything, does this tell you about Skolemization and conservativity?

Exercise 4.4. Let T' in language L' be a conservative extension of T in language L.

(i) Show that if T is consistent, T' is consistent.

(ii) Show that if, moreover, T is deductively complete and $\mathrm{Th}_{L'}(\mathcal{M}') = T'$ for some L'-structure $\mathcal{M}'$, $\mathrm{Th}_L(\mathcal{M}' \restriction L) = T$.

Exercise 4.5. Show that each of the following theories is not complete:

(i) T_{grp}, the theory of groups (see Example 1.15(1)), in the language of groups.

(ii) T_{fld}, the theory of fields, in the language of rings.

(iii) T_{LO}, the theory of linear (total) orderings, in the language of ordered sets.

(iv) T_{sets}, the theory of (naive) sets, in the language consisting of the binary predicate symbol $\in$. [Hint: Counting quantifiers!]

5. The Compactness Theorem, with Applications

We now consider an important consequence of the Completeness Theorem, which allows us to construct nearly any structure we like. As a magic wand we can wave over our problems, its power is more or less commensurate with the Axiom of Choice

or Zorn's Lemma. The result in question, the Compactness Theorem, is also the work of Gödel.

Theorem 5.1. Let T be an L-theory. If for every finite subset $\Gamma \subseteq T$ there is a model $\mathcal{M}_\Gamma \models \Gamma$, then T itself has a model: $\mathcal{M} \models T$ for some $\mathcal{M}$. Equivalently, T is consistent if and only if every finite subset $\Gamma \subseteq T$ is consistent.

Proof. That the two formulations are equivalent is clear from the Completeness Theorem. It suffices, then, to prove the second formulation.

Clearly, if T contains a finite inconsistent subset Γ, it must itself be inconsistent. For the converse, suppose that T is inconsistent. Then there is some formula ϕ such that $T \vdash \phi$ and $T \vdash \neg\phi$. Since proofs are finite, by definition, there must in fact be finite subsets T_+ and T_- of T with

$$T_+ \vdash \phi \text{ and } T_- \vdash \neg\phi.$$

Notice that $\Gamma = T_+ \cup T_-$ is also finite, and

$$\Gamma \vdash \phi \text{ and } \Gamma \vdash \neg\phi.$$

This means, of course, that Γ is inconsistent. □

We will focus on the first formulation and its applications; the second is more useful in the context of proof theory. What does the Compactness Theorem really say? Given a set of sentences/rules T, which may well be infinite, if any finite subset of rules $\Gamma \subseteq T$ is satisfiable—by some particular, individual model $\mathcal{M}_\Gamma$—then there is a single model $\mathcal{M}$ that satisfies absolutely all of the rules in T *at the same time*. This is a big deal.

5.1. **Nonstandard Models.** We begin with a number of examples in which the Compactness Theorem allows us to construct strange and interesting new structures which, from the perspective of first-order logic, are indistinguishable from the simple ones we know and love.

Example 5.2. We will construct a model of the theory of the natural numbers with successor containing a nonstandard element which is not a successor of 0. As this example provides a useful template for those that follow, we divide the argument carefully into its essential components.

Consider $N = \langle \mathbb{N}, S, <, 0 \rangle$, the ordered natural numbers with successor, in the language $L = \langle S, <, 0 \rangle$. Define $T = \mathrm{Th}_L(N)$, where

$$\mathrm{Th}_L(N) = \{\phi \in \mathbf{Sent}_L \,|\, N \models \phi\}$$

is the *complete theory of* N: the exhaustive first-order description of N. So for any $M \in \mathbf{Str}_L$, if $M \models T$ then it satisfies precisely the same first order sentences as N.

Step 1: Expand the language L to $L' = \langle S, <, 0, c \rangle$. Define a new theory

$$T' = T \cup \{c > 0, c > S0, c > S^2 0, \ldots, c > S^n 0, \ldots\}$$

Step 2: Find a model of T' by compactness. In particular, we check that any finite subset of T' has a model. So, let $\Gamma \subseteq T'$ be finite. Then

$$\Gamma \subseteq T \cup \{c > 0, c > S0, \ldots, c > S^k\}$$

for some $k < \omega$. Does Γ have a model? Of course: expand N to an L'-structure M_Γ by interpreting c as the natural number $k+2$, and the rest of the symbols in the same way as before. So $M_\Gamma \models \Gamma$. By compactness, then, there is a model of T', i.e. an L'-structure $\mathfrak{N}$ such that $\mathfrak{N} \models T'$. Notice that $c^{\mathfrak{N}} > (S^{\mathfrak{N}})^n 0^{\mathfrak{N}}$ for all $n < \omega$: the interpretation of c is the desired nonstandard element.

Step 3: We really want an L-structure, though, so we take the reduct: $\bar{N} = \mathfrak{N}|L$. We still have $\bar{N} \models T$, and the element named by c in $|N'| = |N|$ is still nonstandard!

What does this model look like, by the way? There is a standard copy of $\mathbb{N}$, and the element named by c out at infinity. But we also have all successors of c and, since it's nonzero, all of its predecessors as well—none of which are standard, of course. So really we have a copy of $\mathbb{N}$ and a copy of $\mathbb{Z}$ sitting out at infinity, i.e. $\mathbb{N} \amalg \mathbb{Z}$. In fact, there might be other nonstandard elements not related to this one by successor, so a general model of T will in fact be

$$\mathbb{N} \amalg (\coprod_{i \in X} \mathbb{Z}).$$

for any index set X.

We note that while this compactness business is an amazing magic trick (it also allows us to manufacture infinitesimals, graph colorings, and so on), it's also cause for concern. What this example in particular tells us is that the expressive power of first order logic is much lower than we might like: it can't tell the difference between the standard natural numbers, $\mathbb{N}$, and the strange tentacle monsters just described. We can boost the expressive power by moving to, say, an infinitary logic $L_\kappa\lambda$, but then we lose compactness... So there's a tradeoff.

We turn now to an application with its roots in the earliest days of the calculus—both Leibniz and Newton made essential use of the notion of infinitesimal numbers; that is, numbers that are greater than 0, but smaller than any positive real number. This is particularly well-developed by Leibniz, but in any case, these strange quantities play an essential role in the concrete computations of each of the duelling creators of the differential calculus. This notion of infinitesimals was extremely controversial at the time—the philosopher George Berkeley (1685-1753) wrote a *vicious* satire on the subject, entitled "The Analyst"—and were treated as, at best, a questionable notational shorthand until they were finally given real legitimacy by the mathematical logician Abraham Robinson. We give a simplified version of his defense of infinitesimals: we will show that there is a model of the theory of the real numbers as an ordered field which contains infinitesimal objects—indistinguishable from the ordinary real numbers in the sense of first-order logic, but with infinitesimals!

Example 5.3 (Nonstandard real numbers)**.** Consider the real numbers as an ordered field,

$$\mathcal{R} = \{\mathbb{R}, 0, 1, +, \cdot, \leq\},$$

in the language of ordered fields (see Definition **??**), which we will denote by L. Let $T = \mathrm{Th}_L(\mathcal{R})$, the complete first-order theory of $\mathcal{R}$.

Step 1: Let $L' = L \cup \{c\}$, where c is a new constant symbol. Let T' be the L'-theory

$$T' = \mathrm{Th}_L(\mathcal{R}) \cup \{0 < c, c < 1, c < 1/2, c < 1/3, \ldots, c < 1/n, \ldots\},$$

where we here use terms $1/k$ as a shorthand for the appropriate formal expressions, e.g. $1/3$ stands for $(1+1+1)^{-1}$. Notice that if $\mathcal{M}' \models T'$, it will contain an element $c^{\mathcal{M}'}$ that is larger than O, but smaller than $1/n$ for any $n \in \mathbb{N}$—that is, an infinitesimal.

Step 2: We show that T' has a model, by compactness. Let $\Gamma \subset T'$ be finite. Since Γ is finite, in fact

$$\Gamma \subset T \cup \{c > 0, c < 1, c < 1/2, \ldots, c < 1/k\}$$

for some $k \in \mathbb{N}$. Take $\mathcal{M}'_\Gamma$ to be $\mathcal{R}$—the standard real numbers—with $c^{\mathcal{M}'_\Gamma} = \frac{1}{2k}$. Then, clearly, $\mathcal{M}'_\Gamma \models \Gamma$. Since this works for any finite Γ, T' is satifiable: there is some L-structure $\mathcal{M}'$ with $\mathcal{M}' \models T'$.

Step 3: Take the reduct $\mathcal{M} = \mathcal{M}'|L$. The element $c^{\mathcal{M}'} \in |\mathcal{M}'| = |\mathcal{M}|$ is still there, and still infinitesimal!

Once again, the model we have constructed is indistinguishable from the standard real numbers, $\mathcal{R}$, as far as first-order logic is concerned. And yet it contains non-standard elements.

5.2. **Colorings of graphs and maps.**

Example 5.4 (Graph colorings)**.** Recall that give a graph $G = (V, E)$, V the vertex set and E the edge relation, a *k-coloring of G* is a function $\kappa : V \to \{1, 2, \ldots, k\}$ such that adjacent vertices in G always receive different colors; that is,

$$xEy \to \kappa(x) \neq \kappa(y).$$

Let L be the signature containing only a single binary relation, $L = \langle E \rangle$, and let $T = \mathrm{Th}_L(G)$. We define an expanded language by adding an assortment of new constant symbols:

$$L' = L \cup \{c_x\}_{x \in V} \cup \{c_1, \ldots, c_k\} \cup \{\kappa\}$$

Here the new constants c_x serves as names for the vertices of G, the constants c_i, $i = 1, \ldots, k$ are names for the colors, and κ is a unary function symbol representing the coloring function. Define:

$$\begin{array}{rcl} T' = \mathrm{Th}_L(G) & \cup & \{c_x E c_y \to \kappa(c_x) \neq \kappa(c_y) \,|\, x, y \in V\} \\ & \cup & \{\kappa(c_x) = c_1 \vee \cdots \vee \kappa(c_x) = c_k \,|\, x \in V\} \end{array}.$$

If $\mathcal{M}' \models T$, $\mathcal{M}'|L$ is a copy of G with a k-coloring; that is, G is k-colorable. By compactness, this is true if and only if for any finite $\Gamma \supseteq T'$, there is a model $\mathcal{M}_\Gamma \models \Gamma$. Notice that any such Γ will be contained in $\mathrm{Th}_L(G)$ along with finitely many of the additional axioms, say those involving $c_{x_1}, \ldots, c_{x_n}$ for some $x_1, \ldots, x_n \in V$. This has a model if and only if the finite subgraph of G on the vertices $x_1, \ldots, x_n$ has a k-coloring. That is:

Fact 5.5. A graph is k-colorable if and only if all of its finite subgraphs are k-colorable.

This illustrates how the Compactness Theorem often operates as a local-to-global principle, allowing us to push properties from small pieces of a structure to the structure as a whole. It is also connected, rather surprisingly, to a very high-profile mathematical theorem:

Theorem 5.6 (Four Coloring Theorem)**.** Any planar map can be colored with four colors so that no adjoining regions receive the same color.

Map colorings of the form described in the theorem correspond precisely to graph colorings of the kind we have been considering: given a planar map, represent each region as a vertex, and draw an edge between two such vertices just in case the corresponding regions are adjacent. In this light, Fact 5.5 guarantees that is suffices to check that any finite planar graph is 4-colorable. There are, as it happens, only around 300 possible finite configurations, and all of these have been checked for 4-colorability by computer. So the theorem is proved!

Incidentally, the Four Color Theorem was the first genuinely important result to be proved almost entirely by computer (Appel and Haken, 1976).

5.3. **Inexpressibility in First-order.** While the examples in Section 4.2 above tended to emphasize the wonderful power of the compactness theorem to create objects with strange properties, they have a more unsettling side: in each case, the argument reveals that the linguistic resources of first order logic are not rich enough to describe and rule out these strange behaviors. For example, there is no first-order sentence in the language of ordered fields that says "There are no infinitesimal elements"—such a sentence would already be in the complete theory of the real numbers, which is also satisfied by our carefully constructed model *with* infinitesimals!

We close with one additional example of this phenomenon. Recall that a graph is said to be *connected* if, given any two vertices x and y, there is a path between them consisting of finitely many edges. We show that this property cannot be expressed in first order logic.

Example 5.7 (Connectedness is not first-order)**.** Let $L = \langle E \rangle$ be the language of graphs, consisting of the single binary relation symbol E. Let T be the L-theory of loop-free undirected graphs; that is,

$$T = \{\forall x(\neg Exx), \forall x \forall y(Exy \to Eyx)\}.$$

Proposition 5.8. There is no first-order theory $T' \supset T$ whose models are exactly the connected graphs.

Proof. Suppose there is such a T'. In the expanded language $L(T') \cup \{a, b\}$, with a and b new constant symbols, consider the family of sentences of the form

$$\phi_n = \neg \exists x_1 \ldots \exists x_n (a = x_1 \wedge Ex_1x_2 \wedge Ex_2x_3 \wedge \ldots \wedge Ex_{n-1}x_n \wedge x_n = b$$

for each $n \in \mathbb{N}$. That is, each ϕ_n says that there is no path of length n from the vertex named by a to the vertex named by b. Consider

$$T'' = T' \cup \{\phi_b\}_{n \in \mathbb{N}}.$$

Any finite subset Γ of T'' is contained in

$$T' \cup \{\phi_0, \phi_1, \ldots, \phi_k\}$$

for some $k \in \mathbb{N}$. Does $T' \cup \{\phi_0, \phi_1, \ldots, \phi_k\}$ have a model? Absolutely: take $\mathcal{M}_\Gamma$ to be the connected graph (hence already a model of T') with $k + 2$ nodes arranged in sequence:

$$v_1 \text{ —— } v_2 \text{ —— } v_3 \text{ —— } \ldots \text{ —— } v_{k+2}$$

and interpret $a^{\mathcal{M}_\Gamma} = x_0$ and $b^{\mathcal{M}_\Gamma} = x_{k+2}$. Clearly there is no path of length 1, 2, ..., k between the vertices named by a and b, meaning that $\mathcal{M}_\Gamma \models T' \cup \{\phi_0, \phi_1, \ldots, \phi_k\}$. So it is certainly a model of the subset $\Gamma \subseteq T' \cup \{\phi_0, \phi_1, \ldots, \phi_k\}$.

By compactness, there is a model $\mathcal{M}'' \models T''$. Since $\mathcal{M}'' \models T'$, it is connected; since it satisfies all of the ϕ_n, there are elements $a^{\mathcal{M}''}$ and $b^{\mathcal{M}''}$ that are not connected by a path of length n for any $n \in \mathbb{N}$. But this would mean $\mathcal{M}''$ is *not* connected. Contradiction. □

Recall the classes of groups which were claimed to be nonaxiomatizable in Example 1.16: you should go back and try to prove this, using the Compactness Theorem.

5.4. **Sizes of models.** We close with a final application of the Compactness Theorem, related to the existence of models of specific sizes.

First, we consider the finite case:

Fact 5.9. If T is a complete theory and has a model of finite cardinality n, every model of T is of cardinality n.

Proof. Let $\phi = \exists_{=n} x = x$; that is, "there are exactly n elements." By assumption, there is a model $\mathcal{M} \models T$ of size n, so $\mathcal{M} \models \phi$ as well. That is, $\mathcal{M} \models T \cup \{\phi\}$. Since $T \cup \{\phi\}$ is satisfiable, $T \nvdash \neg\phi$. By the Completeness Theorem, then, $T \vdash \phi$. By soundness, moreover, this implies that $T \models \phi$. Put another way, for any $\mathcal{M} \models T$, $\mathcal{M} \models \phi$; that is, any model of T contains n elements. □

This suggests a hard truth: model theory of the kind we are primarily considering here is not particularly useful in working with finite structures. There is an entire field devoted to such questions, *finite model theory*, which is very important in applications and eminently worthy of your attention: we do not, however, have the time to address it in this course. For that reason, we will restrict our attention to infinite models.

We sketch the basics of infinite cardinal numbers here. There is far, far more to say, but we leave that, too, for another course. A cardinal number is, at its core, the size of a set. For example,

$0, 1, 2, \ldots n, \ldots$	$\aleph_0$,	$\aleph_1$,
Finite cardinals/sizes of sets	"aleph null"/"aleph nought"	"aleph one"
	Smallest infinite cardinal	Next smallest cardinal
	Size of $\mathbb{N}$, $\mathbb{Q}$	*Successor* of $\aleph_0$: $\aleph_1 = \aleph_0^+$

$\aleph_2$	$\aleph_3, \aleph_4, \ldots, \aleph_n, \ldots$	$\aleph_\omega$
"aleph two"	...and so on for all $n \in \mathbb{N}$	First cardinal $> \aleph_n$, all $n \in \mathbb{N}$
Next smallest cardinal	$\aleph_{n+1} = \aleph_n^+$	Not a successor: *limit* cardinal

Then we have the successor of $\aleph_\omega$, and its successor, and so on, and the limit cardinal larger than all of them, and the process continues. There are infinitely many infinite cardinals. There are so many, in fact, that the collection of all infinite cardinals is *too large to be a set*—if we permit it to be a set, we would run into serious contradictions in the foundations of mathematics.

There is another operation that allows us to create new cardinals from existing ones, corresponding to the powerset operation.

Definition 5.10. Given a set X, the *powerset of* X is

$$\mathcal{P}(X) = \{A \mid A \subseteq X\},$$

which is often denoted, too, by 2^X. The cardinality of the powerset of X is denoted by $2^{|X|}$.

Fact 5.11. The cardinality of the $\mathbb{R}$, also refereed to as the *cardinality of the continuum*, is $2^{\aleph_0}$. In fact, one can set up an explicit bijection between subsets of $\mathbb{N}$ and (binary expansions of) real numbers.

It is natural to ask, then, whether there is any relationship between 2^{aleph_0} and $\aleph_0$, $\aleph_1$, or any of the other cardinals in the giant increasing sequence of successors and limits discussed above.

Fact 5.12 (Cantor)**.** By a clever argument, referred to as *diagonalization*, $2^{\aleph_0} > \aleph_0$. That is, the set of real numbers is uncountable.

Nearly anything else that we might like to say along these lines is independent of the axioms of set theory: it is just as consistent with the axioms that $2^{\aleph_0} = \aleph_{17}$, for example, as it is that $2^{\aleph_0} \neq \aleph_{17}$.

Example 5.13. The *continuum hypothesis (CH)* asserts that $2^{aleph_0} = \aleph_1$; that is, the cardinality of the continuum is the smallest value it can possibly be. It is, of course, independent of the axioms of set theory, in the sense of the preceding paragraph. If necessary, we can adopt it as a simplifying assumption, without fear of running into a contradiction: it is far better not to assume it, though!

We will not go into further detail here, as the above suffices to state a few last consequences of first-order compactness.

Theorem 5.14 (Downward Löwenheim-Skolem Theorem)**.** For any consistent theory T in language L, there is a model $\mathcal{M} \models T$ of cardinality less than

$$\max(|L|, \aleph_0),$$

where $|L|$, the cardinality of the language L, is the cardinality of the set of its nonlogical symbols.

Example 5.15. Let L be the language of ordered sets, consisting of a single binary predicate symbol R, and let

$$T = \{\forall x Rxx, \forall x \forall y (Rxy \wedge Ryx \to x = y), \forall x \forall y \forall z ([Rxy \wedge Ryz] \to Rxz)\};$$

that is, T is the theory of partial orders. The theory T is certainly consistent (partial orders exist!), and $|L| = 1$. Then by Downward Löwenheim-Skolem, there is a model $\mathcal{M} \models T$ with $|M| \leq \max(1, \aleph_0) = \aleph_0$. That is, there is a countable partial order. This, of course, is no surprise: consider $\mathbb{N}$ as an ordered set!

Example 5.16 (Skolem's Paradox, Baby Version)**.** Take L to be the language of ordered fields, and $T = \mathrm{Th}(\mathbb{R})$. This theory is certainly consistent, and L is countable: by Downward Löwenheim-Skolem, then, T has a countably infinite model. That is, there is a countable version of the real numbers, which is extremely surprising. As we just noted in Facts 5.11 and 5.12, $|\mathbb{R}| = 2^{\aleph_0} > \aleph_0$!

This, and the grown-up version involving models of set theory, just *seems* like a paradox. Really, it illustrates only that countability is not an absolute phenomenon.

Theorem 5.17 (Upward Löwenheim-Skolem)**.** Let T be an L-theory with an infinite model. For any cardinal $\kappa \geq |T|$, there is a model $\mathcal{M} \models T$ of size κ.

Proof. (Sketch) Add κ new constants to the language L, and add sentences to T that force these new symbols to be interpreted by distinct elements. Now, use compactness. □

You could be forgiven for thinking that this is all abstract nonsense, and this business about infinite models cannot possibly be of any practical use. My response:

Fact 5.18. If a theory T has just one model of some size $\kappa > \aleph_0$, then the theory T is decidable—there exists an algorithm to determine whether any given formula ϕ or its negation is provable from T.

So if a theory has a unique model (up to isomorphism) of some uncountable size, its set of consequences is tractable, in computational terms. This sets us up nicely, incidentally, for a discussion of decidability, computability, and, ultimately, Gödel's Incompleteness Theorems.

Chapter 5 Exercises

Exercise 5.1. Suppose that T is a theory with arbitrarily large finite models, i.e. for each $n \in \mathbb{N}$, there is a model $M \models T$ containing at least n elements.

(i) Is T complete?

(ii) Show that T has an infinite model.

Exercise 5.2. Show that for any finite partially ordered set $(X, \leq)$, there is an order $\leq^*$ on X extending $\leq$ that is linear: for any $x, y \in X$, either $x \leq^* y$ or $y \leq^* x$.

(ii) Using (i) and the Compactness Theorem, show that the same is true for any *infinite* partially ordered set $\langle X, \leq \rangle$.

(Hint: $L = L_{ord} = \langle R \rangle$, $T = \mathrm{Th}(\langle X, \leq \rangle)$, $L' = L \cup \{c_x\}_{x \in X}$, and $T' = T \cup \{\forall x \forall y (Rxy \vee Ryx)\}$.)

Exercise 5.3. Consider the theory of (loop-free, undirected) graphs,

$$T = \{\forall x(\neg Exx), \forall x \forall y(Exy \to Eyx)\}$$

in the language containing just a single binary (edge) relation E. Show that there is no extension $T' \supset T$ whose models are exactly the connected graphs. (A graph is connected if one can pass from any vertex to any other by a finite series of edges.)

(Hint: Suppose there *is* such a T'. In a language containing two new constant symbols a and b, consider the sentences of the form

$$\phi_n = \neg \exists x_1 \ldots \exists x_n (a = x_1 \wedge Ex_1x_2 \wedge \cdots \wedge Ex_{n-1}x_n \wedge x_n = b)$$

for $n \in \mathbb{N}$. Use compactness to show $T' \cup \{\phi_n\}_{n \in \mathbb{N}}$ is satisfiable...)

Exercise 5.4. Special case of Upward Löwenheim-Skolem Theorem: Prove that if a theory T has an infinite model, it has an *uncountably* infinite model. (Hint: Add lots of new constant symbols, use compactness.)

Exercise 5.5. Suppose T is a complete theory in a countable language, with an uncountably infinite model. Does T have a countably infinite model? Finite models? If so, why? If not, why not?

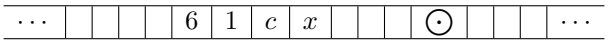
6. Computability and the Incompleteness Theorems

For the final week of the course, we will drastically change our focus. In fact, the substance of this final lecture, concerning computability theory and Gödel's incompleteness theorems, deserves a semester course of its own, so we will just barely skim the surface. The hope is to get a decent basic understanding of the concepts, and a feel for what the incompleteness theorems say—and categorically do not say—as they are clearly among the most significant results in modern mathematics.

6.1. **Computability and Decidability.**

The foundation of theoretical computer science rests on three equivalent definitions of what is means for a function (typically defined on $\mathbb{N}$) to be *computable*:

(1) Turing machines (Alan Turing, 1935): In this account, a function is computable just in case it can be computed by a very simple kind of machine, consisting of an infinite (1-dimensional) strip, divided into single-input cells, and a read/write head that moves freely in both directions along the strip. Graphically,



At each stage, the head, here denoted by $\odot$, can
   (a) Read the symbol at its current location.
   (b) Erase the symbol at its current location (if applicable).
   (c) Overwrite the contents of its current location with a new symbol.
   (d) Move to the right or left.

   It should be noted that there are many varieties of Turing machine—different configurations of cells, different symbols available, and so on. It may not be obvious, but one can compute an awful lot using Turing machines, including: successor, addition, constant functions, bounded subtraction, and so on.

(2) Recursive functions (Gödel, Herbrand, Kleene, and others): Arguably a more mathematically natural way of making sense of computability, with its roots in far older mathematics (e.g. the Fibonacci sequence), a function is computable in this view precisely when it is represented by recursive function, i.e. one which is obtained by composition and recursion from a small family of very simple functions, which everyone would agree are computable. This basic library of functions, the *primitive recursive functions*, consists of, e.g. constant functions, successor, pairing, and projection.
   This is the formalization of computability used in Gödel's proofs of the incompleteness theorems...

(3) λ-calculus (Church): The λ-calculus is a formal theory of function definition and application, which is closely linked with LISP and other functional programming languages, as well as the logical analysis of fragments of real, natural language. A function is computable just in case it is represented by a λ-term. Here again, the library of λ-terms/functions is built from a small collection of basic terms, namely the *combinators*.
This formulation makes clear the genuinely deep connection between logic, functional programming, and category theory (among other things), but, sadly, is too complicated to cover in this abbreviated discussion.

Fact 6.1 (Church-Turing Thesis)**.** If a function is computable in one of the above senses, it is computable in all of the others, as well. That is, characterizations (1), (2) and (3) are equivalent!

For our purposes, we will not need to insist on an absolutely detailed account of what is means to be computable and, in particular, we will not need to choose any particular one of the options described above. The hope is simply to satisfy you that, yes, it is possible to give a rigorous, formal definition of *computability*, just in case anyone should ever force us to do so.

Definition 6.2. Let X be a subset of $\mathbb{N}$.

(1) We say that X is *decidable* if there is an algorithm—that is, a computable function—that determines for each $n \in \mathbb{N}$ whether $n \in X$ or not in a finite amount of time.
(2) We say that X is *semidecidable* if there is an algorithm that lists all elements of X: x_0, x_1, x_2,....

Notes 6.3. (1) Terminology: decidable sets are often called *recursive*, and semidecidable sets *recursively enumerable*.
(2) We can generalize to the case of subsets X of arbitrary (countable) sets Y, i.e. we may wish to know whether a particular theory T is decidable in the set of all sentences in its language.
(3) If $X \subseteq \mathbb{N}$ is semidecidable, there is an algorithm that will confirm in finite time that elements $n \in X$ are, in fact, in X. If $n \notin X$, on the other hand, we would be left waiting forever while the elements of X are being listed—in short, the algorithm cannot tell us that $n \notin X$.
This is precisely the difference between decidable and semidecidable subsets.
(4) If X is decidable, it is semidecidable. A partial converse: If X is semidecidable and $\mathbb{N} \setminus X$, the complement of X in $\mathbb{N}$, is semidecidable, then X is decidable.

Proof. Let $n \in \mathbb{N}$. We would like to know, in finite time, whether $n \in X$ or $n \notin X$, i.e. $n \in \mathbb{N} \setminus X$. Use the listing algorithms for X and $\mathbb{N} \setminus X$,

$$X : x_0, x_1, x_2, \ldots, x_n, \ldots$$
$$\mathbb{N} \setminus X : y_0, y_1, y_2, \ldots, y_k, \ldots,$$

running them in parallel. In finite time, n will appear in one of the two lists, and we will have our answer. □

Fact 6.4. There are semidecidable sets that are not decidable. The classic example is the *halting problem* (Turing/Davis): there is no single general procedure capable

of determining for any combination of algorithm and input whether an output will be produced.

6.2. **The Incompleteness Theorems.** In what follows, we will speak a great deal about "elementary number theory," which will involve one of the following formalizations—which formalization we use will be made clear in each case.

Definition 6.5. We work in the language of arithmetic.

(1) Elementary arithmetic: Known as *Robinson arithmetic*, and denoted by Q. The axioms are the following:
   (a) $\forall x(Sx \neq 0)$
   (b) $\forall x \forall y(Sx = Sy \to x = y)$
   (c) $\forall y(y = 0 \vee \exists x(Sx = y))$
   (d) $\forall x(x + 0 = x)$
   (e) $\forall x \forall y(x + Sy = S(x + y))$
   (f) $\forall x(x \cdot 0 = 0)$
   (g) $\forall x \forall y(x \cdot Sy = (x \cdot y) + x)$

   Axioms (d) and (e) amount to a recursive definition of addition. This, it should be noted, is a very basic, weak system of axioms for arithmetic.
(2) *Primitive recursive arithmetic*, or *PRA*: Roughly, this is

$$\mathrm{Q} + (\text{primitive recursive functions})$$

   or

$$\mathrm{Q} + (\text{mathematical induction for existential formulas } \phi)$$

   That is, we add to Q an axiom scheme ensuring that if an existential formula ϕ (that is, $\phi = \exists x_1 \dots \exists x_n \phi$, with ψ quantifier-free) holds at 0 and holds at $n + 1$ whenever it holds at n, then ϕ holds for all natural numbers.
(3) *Peano arithmetic*, or *PA*: An even stronger system. Roughly,

$$\mathrm{Q} + (\textbf{recursive} \text{ functions})$$

   or

$$\mathrm{Q} + (\text{mathematical induction for } \textbf{arbitrary} \text{ formulas } \phi).$$

Remark 6.6. It is crucial to note that all of the above theories are *effectively axiomatizable*; that is, the axiom sets are either finite (as Q) or decidable (as in PRA and PA, where we must check whether a given formula is an instance of the appropriate induction axiom scheme). Both of the incompleteness theorems are concerned with effectively axiomatizable extensions of these three theories.

Theorem 6.7 (First Incompleteness Theorem)**.** The theory Q is *essentially incomplete*; that is, any consistent, effectively axiomatizable extension $T \supset Q$ is incomplete.

Remark 6.8. Recall that a theory T is incomplete if there is a formula ϕ such that neither $T \vdash \phi$ nor $T \vdash \neg\phi$. So there are things we can neither prove nor disprove using the resources of T. Without straying too far into philosophy, this result destroyed Hilbert's finitist program, which was a central organizing of mathematics in Gödel's day.

Proof. The argument is *incredibly* clever. We begin by sketching the sets, then will expand on them slightly over the next few pages. A full proof, unfortunately, is well beyond the scope of this course.

I. Gödel numbering: we can give a clear, unambiguous coding of formulas (which can talk about $\mathbb{N}$) as natural numbers:

$$\begin{array}{ccc} \phi & \mapsto & \ulcorner\phi\urcorner \in \mathbb{N} \\ \text{(formula)} & & (\text{“Gödel number of } \phi\text{”}) \end{array}$$

II. Deduction rules (e.g. MP) are computable functions of the Gödel numbers—this means that provability is decidable, in the sense of Definition 6.2.

III. For any formula $\phi(x)$, we can construct a sentence ψ such that

$$T \vdash \psi \leftrightarrow \phi(\ulcorner\psi\urcorner);$$

that is, ψ say “I have the property given by the formula ϕ.

IV. Take $\phi(x)$ to be “x is not the Gödel number of a formula provable in T.”

In more detail, now:

Step I. We begin by ordering the symbols of the vocabulary of arithmetic, assigning each one a distinct natural number:

σ	(	)	=	$\wedge$	$\vee$	$\neg$	$\forall$	$\exists$	0	S	+	x_0	x_1	…	x_n	…
$\ulcorner\sigma\urcorner$	1	2	3	4	5	6	7	8	9	10	11	12	13	…	$12+n$	…

with the countably infinite list of variables at the end. We convert formulas to natural numbers in a unique way, taking advantage of the uniqueness of prime factorizations. To be precise, given a formula $\phi = \sigma_1\sigma_2\ldots\sigma_n$, where the terms σ_i are individual symbols in the vocabulary, the Gödel number of ϕ, denoted $\ulcorner\phi\urcorner$, is defined to be

$$\ulcorner\phi\urcorner = 2^{\ulcorner\sigma_1\urcorner}3^{\ulcorner\sigma_2\urcorner}\ldots p_n^{\ulcorner\sigma_n\urcorner},$$

where p_n denotes the nth prime.

Example 6.9. Consider the formula $\phi = \neg(S0 = 0)$. Then, using the symbol numbering above,

$\neg$	(	S	0	=	0	)
6	1	10	9	3	9	2

.

This means that

$$\ulcorner\phi\urcorner = 2^6 \cdot 3^1 \cdot 5^{10} \cdot 7^9 \cdot 11^3 \cdot 13^9 \cdot 17^2.$$

As mentioned above, the uniqueness of prime factorization ensures that, for any positive integer n, there is exactly one possible sequence of symbols $\sigma_1\sigma_2\ldots\sigma_n$ that has it as its Gödel number. Since the set of well-formed formulas is clearly decidable, we have:

Fact 6.10. The set of Gödel numbers of well-formed formulas,

$$\{\ulcorner\phi\urcorner \mid \phi \text{ formula}\},$$

is decidable in $\mathbb{N}$.

Step II. In what follows, we define the Gödel number of a proof $\phi_1, \ldots, \phi_n$ to be $\ulcorner\phi_1 \wedge \cdots \wedge \phi_n\urcorner$. For every logical operation, we define a corresponding (computable) function or predicate on the natural numbers:

(1) Negation:
$$\mathrm{neg}(\ulcorner\phi\urcorner) = \ulcorner\neg\phi\urcorner.$$
(2) Implication:
$$\mathrm{impl}(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) = \ulcorner\phi \to \psi\urcorner$$
(3) Modus Ponens:
$$\mathrm{MP}(x, y, z)$$
is a ternary predicate on $\mathbb{N}$ that holds just in case z is the Gödel number of a formula that follows by Modus Ponens from formulas with Gödel numbers x and y. That is,
$$\mathrm{MP}(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner, \ulcorner\chi\urcorner)$$
holds if and only if χ follows from ϕ and ψ by Modus Ponens.
(4) Proof-ness:
$$\mathrm{Proof}_T(x, y)$$
is a binary predicate on $\mathbb{N}$ that holds just in case x is the Gödel number of a proof in T of the formula with Gödel number y.
(5) Provability:
$$\mathrm{Prov}_T(y) = \exists x \mathrm{Proof}_T(x, y)$$
is a unary predicate that holds just in case y is the Gödel number of a formula provable in T.

Although it takes some work to prove it formally, all of these predicates are decidable!

Step III. Diagonalization:

This is the subtlest part of the proof, and one of many examples of a *fixed point argument* in mathematics (the Brouwer fixed point theorem, concerning continuous maps from the unit disk to itself, is another such example). The end result of this delicate argument is the result quoted in the summary above:

Lemma 6.11. For any formula $\phi(x)$, there is a sentence ψ such that
$$T \vdash \psi \leftarrow \phi(\ulcorner\psi\urcorner).$$

This formula ψ is able to talk about itself: it says, roughly speaking, "I have the property ϕ."

Step IV. We now put all of the pieces together. Let $\phi(x)$ be $\neg\mathrm{Prov}_T(x)$, "x is not the Gödel number of a formula provable in T." By diagonalization, there is a sentence G_T, the *Gödel sentence of* T, so that
$$T \vdash G_T \leftrightarrow \neg\mathrm{Prov}_T(\ulcorner G_T\urcorner).$$

Roughly, G_T says "I am not provable in T."

Say that $T \vdash G_T$. Then $T \vdash \mathrm{Prov}_T(\ulcorner G_T\urcorner)$[8] and therefore $T \vdash \neg G_T$ from the equivalence above. But then T proves both G_T and $\neg G_T$, contradicting consistency of T.

[8]This needs some justification: that provability implies provability of provability is known as *weak representability* of the predicate $\mathrm{Prov}_T(-)$.

Suppose, on the other hand, that $T \vdash \neg G_T$. Then $T \nvdash G_T$ and, after a little more fiddling with representability, we can say that no $n \in \mathbb{N}$ is the Gödel number of a proof of G_T, i.e. that $T \vdash \neg \text{Proof}_T(n, \lceil G_T \rceil)$ for all n. By consistency[9],

$$T \nvdash \exists x \text{Proof}_T(x, \lceil G_T \rceil).$$

This implies that $T \nvdash \text{Prov}_T(\lceil G_T \rceil)$ and, by the equivalence, $T \nvdash \neg G_T$. Contradiction.

We have shown that T proves neither G_T nor $\neg G_T$, so the theory is incomplete. □

Theorem 6.12 (Second Incompleteness Theorem)**.** If T is a 1-consistent, effectively axiomatizable extension of PRA, then

$$T \nvdash \text{Con}_T,$$

where Con_T is a sentence equivalent to consistency of the theory T.

Idea: No sufficiently rich mathematical theory can prove its own consistency. While this was very damaging to Hilbert's program, it likely feels less surprising in this postmodern era. From a certain standpoint, it seems only natural a formal system cannot verify its own correctness: surely we must step outside of the system first, before we are able to evaluate questions of this nature.

Proof. The proof is considerably more technical. We here observe only that the sentence Con_T may be taken to be $\text{Prov}_T(\lceil \bot \rceil)$, where $\bot$ is any contradiction. □

Gödel's Incompleteness Theorems pose a number of important challenges to mathematics as a field, and to some extent, to working mathematicians going about their daily lives. These include:

(1) There will always be things we cannot prove in our favorite formal systems (at least if those systems contain any meaningful fragment of number theory).
(2) Whatever system we work in, we need to look *outside* the system for a proof of its consistency. This is a particular challenge for those using formal methods to test the correctness of software, e.g. for air traffic control, and for those developing automated theorem provers.

How do we meet these challenges? A few options:

(1) Transfinite induction. We may allow ourselves to carry out induction not just on the natural numbers, but on some initial segment of the infinite ordinals as well:

$$1, 2, 3, \ldots, n, \ldots, \omega, \omega + 1, \ldots, \omega \cdot 2, \ldots, \omega^2, \ldots, \omega^\omega, \ldots, \omega^{\omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}}, \ldots$$

The exponential tower of ω copies of ω above, which is denoted ϵ_0, is of surprising significance: if we allow ourselves induction up to ϵ_0 in PA, we can prove Con_{PA}.

(2) More computable functions. If we give ourselves access to more computable functions (or, equivalently, bits of second order logic), we can settle all kinds of undecidable propositions. The field of "reverse mathematics" is concerned with

[9]Actually, we need not just consistency, but 1-*consistency*: T is 1-consistent if there is no existential formula ϕ such that $T \vdash \neg\phi(n)$ for all $n \in \mathbb{N}$, but $T \vdash \exists x \phi(x)$. This is a stronger condition.

precisely this approach: given a mathematical result, the game is to determine the weakest possible fraction of second-order logic that is required to prove it.

You could be forgiven for thinking that incompleteness of the kind discussed here is unlikely to affect your working life. Certainly, the Gödel sentence is a very artificial kind of statement, which has nothing much to do with "real" mathematics. There are, though, natural mathematical examples of incompleteness: concrete, wholesome mathematical statements that can be neither proven nor disproven in PA. We give a few examples here, as a kind of encouragement to further reading.

Example 6.13 (Paris-Harrington, 1972)**.** The Paris-Harrington Conjecture is a combinatorial/coloring principle along the lines of Ramsey's Theorem. In fact, the conjecture implies $\mathrm{Con}_{\mathrm{PA}}$, so it is certainly not provable in PA.

Example 6.14 (Goodstein sequences)**.** Goodstein sequences are relatively simple to define and, at first, grow *extremely* rapidly. That they all eventually converge to 0 is provable (it is a theorem!), but not provable in PA.

E-mail address: qmlieberman@vutbr.cz

Department of Mathematics, Faculty of Mechanical Engineering, Brno University of Technology, Brno, Czech Republic