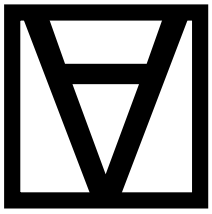
Intermediate Logic

Richard Zach

Philosophy 310
Winter Term 2015
McGill University



Contents

Preface

Formal logic has many applications both within philosophy and outside (especially in mathematics, computer science, and linguistics). This second course will introduce you to the concepts, results, and methods of formal logic necessary to understand and appreciate these applications as well as the limitations of formal logic. It will be mathematical in that you will be required to master abstract formal concepts and to prove theorems *about* logic (not just *in* logic the way you did in Phil 210); but it does not presuppose any advanced knowledge of mathematics.

We will begin by studying some basic formal concepts: sets, relations, and functions and sizes of infinite sets. We will then consider the language, semantics, and proof theory of first-order logic (FOL), and ways in which we can use first-order logic to formalize facts and reasoning abouts some domains of interest to philosophers and logicians.

In the second part of the course, we will begin to investigate the metatheory of first-order logic. We will concentrate on a few central results: the completeness theorem, which relates the proof theory and semantics of first-order logic, and the compactness theorem and Löwenheim-Skolem theorems, which concern the existence and size of first-order interpretations.

In the third part of the course, we will discuss a particular way of making precise what it means for a function to be computable, namely, when it is recursive. This will enable us to prove important results in the metatheory of logic and of formal systems formulated in first-order logic: Gödel's incompleteness theorem, the Church-Turing undecidability theorem, and Tarski's theorem about the undefinability of truth.

Week 1 (Jan 5, 7). Introduction. Sets and Relations.

Week 2 (Jan 12, 14). Functions. Enumerability.

Week 3 (Jan 19, 21). Syntax and Semantics of FOL.

Week 4 (Jan 26, 28). Structures and Theories.

Week 5 (Feb 2, 5). Sequent Calculus and Proofs in FOL.

Week 6 (Feb 9, 12). The Completeness Theorem.

Week 7 (Feb 16, 18). Compactness and Löwenheim-Skolem Theorems

Week 8 (Mar 23, 25). Recursive Functions

Week 9 (Mar 9, 11). Arithmetization of Syntax

Week 10 (Mar 16, 18). Theories and Computability

Week 11 (Mar 23, 25). Gödel's Incompleteness Theorems

Week 12 (Mar 30, Apr 1). The Undefinability of Truth.

Week 13, 14 (Apr 8, 13). Applications.

Part I

Sets, Relations, Functions

Chapter 1

Sets

1.1 Extensionality

A *set* is a collection of objects, considered as a single object. The objects making up the set are called *elements* or *members* of the set. If x is an element of a set a, we write $x \in a$; if not, we write $x \notin a$. The set which has no elements is called the *empty* set and denoted "$\emptyset$".

It does not matter how we *specify* the set, or how we *order* its elements, or indeed how *many times* we count its elements. All that matters are what its elements are. We codify this in the following principle.

Definition 1.1 (Extensionality). If A and B are sets, then $A = B$ iff every element of A is also an element of B, and vice versa.

Extensionality licenses some notation. In general, when we have some objects $a_1, \ldots, a_n$, then $\{a_1, \ldots, a_n\}$ is *the* set whose elements are $a_1, \ldots, a_n$. We emphasise the word "*the*", since extensionality tells us that there can be only *one* such set. Indeed, extensionality also licenses the following:

$$\{a, a, b\} = \{a, b\} = \{b, a\}.$$

This delivers on the point that, when we consider sets, we don't care about the order of their elements, or how many times they are specified.

Example 1.2. Whenever you have a bunch of objects, you can collect them together in a set. The set of Richard's siblings, for instance, is a set that contains one person, and we could write it as $S = \{\text{Ruth}\}$. The set of positive integers less than 4 is $\{1, 2, 3\}$, but it can also be written as $\{3, 2, 1\}$ or even as $\{1, 2, 1, 2, 3\}$. These are all the same set, by extensionality. For every element of $\{1, 2, 3\}$ is also an element of $\{3, 2, 1\}$ (and of $\{1, 2, 1, 2, 3\}$), and vice versa.

Frequently we'll specify a set by some property that its elements share. We'll use the following shorthand notation for that: $\{x : \phi(x)\}$, where the

$\phi(x)$ stands for the property that x has to have in order to be counted among the elements of the set.

Example 1.3. In our example, we could have specified S also as

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

Example 1.4. A number is called *perfect* iff it is equal to the sum of its proper divisors (i.e., numbers that evenly divide it but aren't identical to the number). For instance, 6 is perfect because its proper divisors are 1, 2, and 3, and $6 = 1 + 2 + 3$. In fact, 6 is the only positive integer less than 10 that is perfect. So, using extensionality, we can say:

$$\{6\} = \{x : x \text{ is perfect and } 0 \leq x \leq 10\}$$

We read the notation on the right as "the set of x's such that x is perfect and $0 \leq x \leq 10$". The identity here confirms that, when we consider sets, we don't care about how they are specified. And, more generally, extensionality guarantees that there is always only one set of x's such that $\phi(x)$. So, extensionality justifies calling $\{x : \phi(x)\}$ *the* set of x's such that $\phi(x)$.

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

1.2 Subsets and Power Sets

We will often want to compare sets. And one obvious kind of comparison one might make is as follows: *everything in one set is in the other too*. This situation is sufficiently important for us to introduce some new notation.

Definition 1.5 (Subset). If every element of a set A is also an element of B, then we say that A is a *subset* of B, and write $A \subseteq B$. If A is not a subset of B we write $A \not\subseteq B$. If $A \subseteq B$ but $A \neq B$, we write $A \subsetneq B$ and say that A is a *proper subset* of B.

Example 1.6. Every set is a subset of itself, and $\emptyset$ is a subset of every set. The set of even numbers is a subset of the set of natural numbers. Also, $\{a, b\} \subseteq \{a, b, c\}$. But $\{a, b, e\}$ is not a subset of $\{a, b, c\}$.

Example 1.7. The number 2 is an element of the set of integers, whereas the set of even numbers is a subset of the set of integers. However, a set may happen to *both* be an element and a subset of some other set, e.g., $\{0\} \in \{0, \{0\}\}$ and also $\{0\} \subseteq \{0, \{0\}\}$.

Extensionality gives a criterion of identity for sets: $A = B$ iff every element of A is also an element of B and vice versa. The definition of "subset" defines $A \subseteq B$ precisely as the first half of this criterion: every element of A is also an element of B. Of course the definition also applies if we switch A and B: that is, $B \subseteq A$ iff every element of B is also an element of A. And that, in turn, is exactly the "vice versa" part of extensionality. In other words, extensionality entails that sets are equal iff they are subsets of one another.

Proposition 1.8. *$A = B$ iff both $A \subseteq B$ and $B \subseteq A$.*

Now is also a good opportunity to introduce some further bits of helpful notation. In defining when A is a subset of B we said that "every element of A is ...," and filled the "..." with "an element of B". But this is such a common *shape* of expression that it will be helpful to introduce some formal notation for it.

Definition 1.9. $(\forall x \in A)\phi$ abbreviates $\forall x(x \in A \to \phi)$. Similarly, $(\exists x \in A)\phi$ abbreviates $\exists x(x \in A \wedge \phi)$.

Using this notation, we can say that $A \subseteq B$ iff $(\forall x \in A)x \in B$.

Now we move on to considering a certain kind of set: the set of all subsets of a given set.

Definition 1.10 (Power Set). The set consisting of all subsets of a set A is called the *power set of* A, written $\wp(A)$.

$$\wp(A) = \{B : B \subseteq A\}$$

Example 1.11. What are all the possible subsets of $\{a, b, c\}$? They are: $\emptyset$, $\{a\}$, $\{b\}$, $\{c\}$, $\{a, b\}$, $\{a, c\}$, $\{b, c\}$, $\{a, b, c\}$. The set of all these subsets is $\wp(\{a, b, c\})$:

$$\wp(\{a, b, c\}) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$$

1.3 Some Important Sets

Example 1.12. We will mostly be dealing with sets whose elements are mathematical objects. Four such sets are important enough to have specific names:

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\}$$
the set of natural numbers

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$
the set of integers

$$\mathbb{Q} = \{m/n : m, n \in \mathbb{Z} \text{ and } n \neq 0\}$$
the set of rationals

$$\mathbb{R} = (-\infty, \infty)$$
the set of real numbers (the continuum)

These are all *infinite* sets, that is, they each have infinitely many elements.

As we move through these sets, we are adding *more* numbers to our stock. Indeed, it should be clear that $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$: after all, every natural number is an integer; every integer is a rational; and every rational is a real. Equally, it should be clear that $\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q}$, since -1 is an integer but not a natural number, and $1/2$ is rational but not integer. It is less obvious that $\mathbb{Q} \subsetneq \mathbb{R}$, i.e., that there are some real numbers which are not rational.

We'll sometimes also use the set of positive integers $\mathbb{Z}^+ = \{1, 2, 3, \ldots\}$ and the set containing just the first two natural numbers $\mathbb{B} = \{0, 1\}$.

Example 1.13 (Strings). Another interesting example is the set A^* of *finite strings* over an alphabet A: any finite sequence of elements of A is a string over A. We include the *empty string* Λ among the strings over A, for every alphabet A. For instance,

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

If $x = x_1 \ldots x_n \in A^*$is a string consisting of n "letters" from A, then we say *length* of the string is n and write $\text{len}(x) = n$.

Example 1.14 (Infinite sequences). For any set A we may also consider the set A^ω of infinite sequences of elements of A. An infinite sequence $a_1 a_2 a_3 a_4 \ldots$ consists of a one-way infinite list of objects, each one of which is an element of A.

1.4 Unions and Intersections

In section 1.1, we introduced definitions of sets by abstraction, i.e., definitions of the form $\{x : \phi(x)\}$. Here, we invoke some property ϕ, and this property

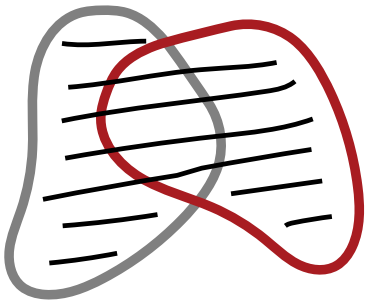


Figure 1.1: The union $A \cup B$ of two sets is set of elements of A together with those of B.

can mention sets we've already defined. So for instance, if A and B are sets, the set $\{x : x \in A \vee x \in B\}$ consists of all those objects which are elements of either A or B, i.e., it's the set that combines the elements of A and B. We can visualize this as in Figure 1.1, where the highlighted area indicates the elements of the two sets A and B together.

This operation on sets—combining them—is very useful and common, and so we give it a formal name and a symbol.

Definition 1.15 (Union). The *union* of two sets A and B, written $A \cup B$, is the set of all things which are elements of A, B, or both.

$$A \cup B = \{x : x \in A \vee x \in B\}$$

Example 1.16. Since the multiplicity of elements doesn't matter, the union of two sets which have an element in common contains that element only once, e.g., $\{a, b, c\} \cup \{a, 0, 1\} = \{a, b, c, 0, 1\}$.

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

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

We can also consider a "dual" operation to union. This is the operation that forms the set of all elements that are elements of A and are also elements of B. This operation is called *intersection*, and can be depicted as in Figure 1.2.

Definition 1.17 (Intersection). The *intersection* of two sets A and B, written $A \cap B$, is the set of all things which are elements of both A and B.

$$A \cap B = \{x : x \in A \wedge x \in B\}$$

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

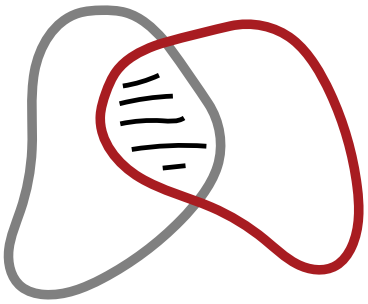


Figure 1.2: The intersection $A \cap B$ of two sets is the set of elements they have in common.

Example 1.18. If two sets have no elements in common, their intersection is empty: $\{a, b, c\} \cap \{0, 1\} = \emptyset$.

If two sets do have elements in common, their intersection is the set of all those: $\{a, b, c\} \cap \{a, b, d\} = \{a, b\}$.

The intersection of a set with one of its subsets is just the smaller set: $\{a, b, c\} \cap \{a, b\} = \{a, b\}$.

The intersection of any set with the empty set is empty: $\{a, b, c\} \cap \emptyset = \emptyset$.

We can also form the union or intersection of more than two sets. An elegant way of dealing with this in general is the following: suppose you collect all the sets you want to form the union (or intersection) of into a single set. Then we can define the union of all our original sets as the set of all objects which belong to at least one element of the set, and the intersection as the set of all objects which belong to every element of the set.

Definition 1.19. If A is a set of sets, then $\bigcup A$ is the set of elements of elements of A:

$$\begin{aligned} \bigcup A &= \{x : x \text{ belongs to an element of } A\}, \text{ i.e.,} \\ &= \{x : \text{there is a } B \in A \text{ so that } x \in B\} \end{aligned}$$

Definition 1.20. If A is a set of sets, then $\bigcap A$ is the set of objects which all elements of A have in common:

$$\begin{aligned} \bigcap A &= \{x : x \text{ belongs to every element of } A\}, \text{ i.e.,} \\ &= \{x : \text{for all } B \in A, x \in B\} \end{aligned}$$

Example 1.21. Suppose $A = \{\{a, b\}, \{a, d, e\}, \{a, d\}\}$. Then $\bigcup A = \{a, b, d, e\}$ and $\bigcap A = \{a\}$.

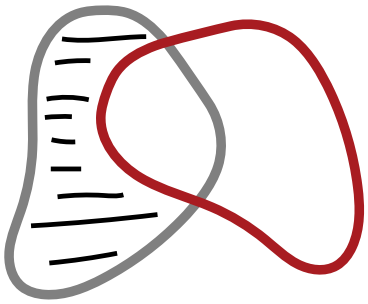


Figure 1.3: The difference $A \setminus B$ of two sets is the set of those elements of A which are not also elements of B.

We could also do the same for a sequence of sets $A_1, A_2, \ldots$

$$\bigcup_i A_i = \{x : x \text{ belongs to one of the } A_i\}$$
$$\bigcap_i A_i = \{x : x \text{ belongs to every } A_i\}.$$

When we have an *index* of sets, i.e., some set I such that we are considering A_i for each $i \in I$, we may also use these abbreviations:

$$\bigcup_{i \in I} A_i = \bigcup\{A_i : i \in I\}$$
$$\bigcap_{i \in I} A_i = \bigcap\{A_i : i \in I\}$$

Finally, we may want to think about the set of all elements in A which are not in B. We can depict this as in Figure 1.3.

Definition 1.22 (Difference). The *set difference* $A \setminus B$ is the set of all elements of A which are not also elements of B, i.e.,

$$A \setminus B = \{x : x \in A \text{ and } x \notin B\}.$$

1.5 Pairs, Tuples, Cartesian Products

It follows from extensionality that sets have no order to their elements. So if we want to represent order, we use *ordered pairs* $\langle x, y \rangle$. In an unordered pair $\{x, y\}$, the order does not matter: $\{x, y\} = \{y, x\}$. In an ordered pair, it does: if $x \neq y$, then $\langle x, y \rangle \neq \langle y, x \rangle$.

How should we think about ordered pairs in set theory? Crucially, we want to preserve the idea that ordered pairs are identical iff they share the same first element and share the same second element, i.e.:

$$\langle a, b \rangle = \langle c, d \rangle \text{ iff both } a = c \text{ and } b = d.$$

We can define ordered pairs in set theory using the Wiener-Kuratowski definition.

Definition 1.23 (Ordered pair). $\langle a, b\rangle = \{\{a\}, \{a, b\}\}$.

Having fixed a definition of an ordered pair, we can use it to define further sets. For example, sometimes we also want ordered sequences of more than two objects, e.g., *triples* $\langle x, y, z\rangle$, *quadruples* $\langle x, y, z, u\rangle$, and so on. We can think of triples as special ordered pairs, where the first element is itself an ordered pair: $\langle x, y, z\rangle$ is $\langle\langle x, y\rangle, z\rangle$. The same is true for quadruples: $\langle x, y, z, u\rangle$ is $\langle\langle\langle x, y\rangle, z\rangle, u\rangle$, and so on. In general, we talk of *ordered n-tuples* $\langle x_1, \ldots, x_n\rangle$.

Certain sets of ordered pairs, or other ordered n-tuples, will be useful.

Definition 1.24 (Cartesian product). Given sets A and B, their *Cartesian product* $A \times B$ is defined by

$$A \times B = \{\langle x, y\rangle : x \in A \text{ and } y \in B\}.$$

Example 1.25. If $A = \{0, 1\}$, and $B = \{1, a, b\}$, then their product is

$$A \times B = \{\langle 0, 1\rangle, \langle 0, a\rangle, \langle 0, b\rangle, \langle 1, 1\rangle, \langle 1, a\rangle, \langle 1, b\rangle\}.$$

Example 1.26. If A is a set, the product of A with itself, $A \times A$, is also written A^2. It is the set of *all* pairs $\langle x, y\rangle$ with $x, y \in A$. The set of all triples $\langle x, y, z\rangle$ is A^3, and so on. We can give a recursive definition:

$$A^1 = A$$
$$A^{k+1} = A^k \times A$$

Proposition 1.27. *If A has n elements and B has m elements, then $A \times B$ has $n \cdot m$ elements.*

Proof. For every element x in A, there are m elements of the form $\langle x, y\rangle \in A \times B$. Let $B_x = \{\langle x, y\rangle : y \in B\}$. Since whenever $x_1 \neq x_2$, $\langle x_1, y\rangle \neq \langle x_2, y\rangle$, $B_{x_1} \cap B_{x_2} = \emptyset$. But if $A = \{x_1, \ldots, x_n\}$, then $A \times B = B_{x_1} \cup \cdots \cup B_{x_n}$, and so has $n \cdot m$ elements.

To visualize this, arrange the elements of $A \times B$ in a grid:

$$\begin{array}{rccccc} B_{x_1} = & \{\langle x_1, y_1\rangle & \langle x_1, y_2\rangle & \ldots & \langle x_1, y_m\rangle\} \\ B_{x_2} = & \{\langle x_2, y_1\rangle & \langle x_2, y_2\rangle & \ldots & \langle x_2, y_m\rangle\} \\ \vdots & & \vdots & & \\ B_{x_n} = & \{\langle x_n, y_1\rangle & \langle x_n, y_2\rangle & \ldots & \langle x_n, y_m\rangle\} \end{array}$$

Since the x_i are all different, and the y_j are all different, no two of the pairs in this grid are the same, and there are $n \cdot m$ of them. □

Example 1.28. If A is a set, a *word* over A is any sequence of elements of A. A sequence can be thought of as an n-tuple of elements of A. For instance, if $A = \{a, b, c\}$, then the sequence "*bac*" can be thought of as the triple $\langle b, a, c\rangle$. Words, i.e., sequences of symbols, are of crucial importance in computer science. By convention, we count elements of A as sequences of length 1, and $\emptyset$ as the sequence of length 0. The set of *all* words over A then is

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

1.6 Russell's Paradox

Extensionality licenses the notation $\{x : \phi(x)\}$, for *the* set of x's such that $\phi(x)$. However, all that extensionality *really* licenses is the following thought. *If* there is a set whose members are all and only the ϕ's, *then* there is only one such set. Otherwise put: having fixed some ϕ, the set $\{x : \phi(x)\}$ is unique, *if it exists*.

But this conditional is important! Crucially, not every property lends itself to *comprehension*. That is, some properties do *not* define sets. If they all did, then we would run into outright contradictions. The most famous example of this is Russell's Paradox.

Sets may be elements of other sets—for instance, the power set of a set A is made up of sets. And so it makes sense to ask or investigate whether a set is an element of another set. Can a set be a member of itself? Nothing about the idea of a set seems to rule this out. For instance, if *all* sets form a collection of objects, one might think that they can be collected into a single set—the set of all sets. And it, being a set, would be an element of the set of all sets.

Russell's Paradox arises when we consider the property of not having itself as an element, of being *non-self-membered*. What if we suppose that there is a set of all sets that do not have themselves as an element? Does

$$R = \{x : x \notin x\}$$

exist? It turns out that we can prove that it does not.

Theorem 1.29 (Russell's Paradox). *There is no set* $R = \{x : x \notin x\}$.

Proof. If $R = \{x : x \notin x\}$ exists, then $R \in R$ iff $R \notin R$, which is a contradiction. □

Let's run through this proof more slowly. If R exists, it makes sense to ask whether $R \in R$ or not. Suppose that indeed $R \in R$. Now, R was defined as the set of all sets that are not elements of themselves. So, if $R \in R$, then R does not itself have R's defining property. But only sets that have this property are in R, hence, R cannot be an element of R, i.e., $R \notin R$. But R can't both be and not be an element of R, so we have a contradiction.

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

How do we set up a set theory which avoids falling into Russell's Paradox, i.e., which avoids making the *inconsistent* claim that $R = \{x : x \notin x\}$ exists? Well, we would need to lay down axioms which give us very precise conditions for stating when sets exist (and when they don't).

The set theory sketched in this chapter doesn't do this. It's *genuinely naïve*. It tells you only that sets obey extensionality and that, if you have some sets, you can form their union, intersection, etc. It is possible to develop set theory more rigorously than this.

Chapter 2

Relations

2.1 Relations as Sets

In section 1.3, we mentioned some important sets: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$. You will no doubt remember some interesting relations between the elements of some of these sets. For instance, each of these sets has a completely standard *order relation* on it. There is also the relation *is identical with* that every object bears to itself and to no other thing. There are many more interesting relations that we'll encounter, and even more possible relations. Before we review them, though, we will start by pointing out that we can look at relations as a special sort of set.

For this, recall two things from section 1.5. First, recall the notion of a *ordered pair*: given a and b, we can form $\langle a, b\rangle$. Importantly, the order of elements *does* matter here. So if $a \neq b$ then $\langle a, b\rangle \neq \langle b, a\rangle$. (Contrast this with unordered pairs, i.e., 2-element sets, where $\{a, b\} = \{b, a\}$.) Second, recall the notion of a *Cartesian product*: if A and B are sets, then we can form $A \times B$, the set of all pairs $\langle x, y\rangle$ with $x \in A$ and $y \in B$. In particular, $A^2 = A \times A$ is the set of all ordered pairs from A.

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

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

There is a close connection between n being less than m, and the pair $\langle n, m\rangle$ being a member of R, namely:

$$n < m \text{ iff } \langle n, m\rangle \in R.$$

Indeed, without any loss of information, we can consider the set R to *be* the $<$-relation on $\mathbb{N}$.

In the same way we can construct a subset of $\mathbb{N}^2$ for any relation between numbers. Conversely, given any set of pairs of numbers $S \subseteq \mathbb{N}^2$, there is a

corresponding relation between numbers, namely, the relationship n bears to m if and only if $\langle n, m\rangle \in S$. This justifies the following definition:

Definition 2.1 (Binary relation). A *binary relation* on a set A is a subset of A^2. If $R \subseteq A^2$ is a binary relation on A and $x, y \in A$, we sometimes write Rxy (or xRy) for $\langle x, y\rangle \in R$.

Example 2.2. The set $\mathbb{N}^2$ of pairs of natural numbers can be listed in a 2-dimensional matrix like this:

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

We have put the diagonal, here, in bold, since the subset of $\mathbb{N}^2$ consisting of the pairs lying on the diagonal, i.e.,

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

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

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

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

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

is the *greater than* relation, i.e., Gnm iff $n > m$. The union of L with I, which we might call $K = L \cup I$, is the *less than or equal to* relation: Knm iff $n \leq m$. Similarly, $H = G \cup I$ is the *greater than or equal to relation.* These relations L, G, K, and H are special kinds of relations called *orders*. L and G have the property that no number bears L or G to itself (i.e., for all n, neither Lnn nor Gnn). Relations with this property are called *irreflexive*, and, if they also happen to be orders, they are called *strict orders.*

Although orders and identity are important and natural relations, it should be emphasized that according to our definition *any* subset of A^2 is a relation on A, regardless of how unnatural or contrived it seems. In particular, $\emptyset$ is a relation on any set (the *empty relation*, which no pair of elements bears), and A^2 itself is a relation on A as well (one which every pair bears), called the *universal relation*. But also something like $E = \{\langle n, m\rangle : n > 5 \text{ or } m \times n \geq 34\}$ counts as a relation.

2.2 Special Properties of Relations

Some kinds of relations turn out to be so common that they have been given special names. For instance, $\leq$ and $\subseteq$ both relate their respective domains (say, $\mathbb{N}$ in the case of $\leq$ and $\wp(A)$ in the case of $\subseteq$) in similar ways. To get at exactly how these relations are similar, and how they differ, we categorize them according to some special properties that relations can have. It turns out that (combinations of) some of these special properties are especially important: orders and equivalence relations.

Definition 2.3 (Reflexivity). A relation $R \subseteq A^2$ is *reflexive* iff, for every $x \in A$, Rxx.

Definition 2.4 (Transitivity). A relation $R \subseteq A^2$ is *transitive* iff, whenever Rxy and Ryz, then also Rxz.

Definition 2.5 (Symmetry). A relation $R \subseteq A^2$ is *symmetric* iff, whenever Rxy, then also Ryx.

Definition 2.6 (Anti-symmetry). A relation $R \subseteq A^2$ is *anti-symmetric* iff, whenever both Rxy and Ryx, then $x = y$ (or, in other words: if $x \neq y$ then either $\neg Rxy$ or $\neg Ryx$).

In a symmetric relation, Rxy and Ryx always hold together, or neither holds. In an anti-symmetric relation, the only way for Rxy and Ryx to hold together is if $x = y$. Note that this does not *require* that Rxy and Ryx holds when $x = y$, only that it isn't ruled out. So an anti-symmetric relation can be reflexive, but it is not the case that every anti-symmetric relation is reflexive. Also note that being anti-symmetric and merely not being symmetric are different conditions. In fact, a relation can be both symmetric and anti-symmetric at the same time (e.g., the identity relation is).

Definition 2.7 (Connectivity). A relation $R \subseteq A^2$ is *connected* if for all $x, y \in A$, if $x \neq y$, then either Rxy or Ryx.

Definition 2.8 (Irreflexivity). A relation $R \subseteq A^2$ is called *irreflexive* if, for all $x \in A$, not Rxx.

Definition 2.9 (Asymmetry). A relation $R \subseteq A^2$ is called *asymmetric* if for no pair $x, y \in A$ we have both Rxy and Ryx.

Note that if $A \neq \emptyset$, then no irreflexive relation on A is reflexive and every asymmetric relation on A is also anti-symmetric. However, there are $R \subseteq A^2$ that are not reflexive and also not irreflexive, and there are anti-symmetric relations that are not asymmetric.

2.3 Equivalence Relations

The identity relation on a set is reflexive, symmetric, and transitive. Relations R that have all three of these properties are very common.

Definition 2.10 (Equivalence relation). A relation $R \subseteq A^2$ that is reflexive, symmetric, and transitive is called an *equivalence relation*. Elements x and y of A are said to be *R-equivalent* if Rxy.

Equivalence relations give rise to the notion of an *equivalence class*. An equivalence relation "chunks up" the domain into different partitions. Within each partition, all the objects are related to one another; and no objects from different partitions relate to one another. Sometimes, it's helpful just to talk about these partitions *directly*. To that end, we introduce a definition:

Definition 2.11. Let $R \subseteq A^2$ be an equivalence relation. For each $x \in A$, the *equivalence class* of x in A is the set $[x]_R = \{y \in A : Rxy\}$. The *quotient* of A under R is $A/_R = \{[x]_R : x \in A\}$, i.e., the set of these equivalence classes.

The next result vindicates the definition of an equivalence class, in proving that the equivalence classes are indeed the partitions of A:

Proposition 2.12. *If $R \subseteq A^2$ is an equivalence relation, then Rxy iff $[x]_R = [y]_R$.*

Proof. For the left-to-right direction, suppose Rxy, and let $z \in [x]_R$. By definition, then, Rxz. Since R is an equivalence relation, Ryz. (Spelling this out: as Rxy and R is symmetric we have Ryx, and as Rxz and R is transitive we have Ryz.) So $z \in [y]_R$. Generalising, $[x]_R \subseteq [y]_R$. But exactly similarly, $[y]_R \subseteq [x]_R$. So $[x]_R = [y]_R$, by extensionality.

For the right-to-left direction, suppose $[x]_R = [y]_R$. Since R is reflexive, Ryy, so $y \in [y]_R$. Thus also $y \in [x]_R$ by the assumption that $[x]_R = [y]_R$. So Rxy. □

Example 2.13. A nice example of equivalence relations comes from modular arithmetic. For any a, b, and $n \in \mathbb{N}$, say that $a \equiv_n b$ iff dividing a by n gives the same remainder as dividing b by n. (Somewhat more symbolically: $a \equiv_n b$ iff, for some $k \in \mathbb{Z}$, $a - b = kn$.) Now, $\equiv_n$ is an equivalence relation, for any n. And there are exactly n distinct equivalence classes generated by $\equiv_n$; that is, $\mathbb{N}/_{\equiv_n}$ has n elements. These are: the set of numbers divisible by n without remainder, i.e., $[0]_{\equiv_n}$; the set of numbers divisible by n with remainder 1, i.e., $[1]_{\equiv_n}$; ...; and the set of numbers divisible by n with remainder $n-1$, i.e., $[n-1]_{\equiv_n}$.

2.4 Orders

Many of our comparisons involve describing some objects as being "less than", "equal to", or "greater than" other objects, in a certain respect. These involve *order* relations. But there are different kinds of order relations. For instance, some require that any two objects be comparable, others don't. Some include identity (like $\leq$) and some exclude it (like $<$). It will help us to have a taxonomy here.

Definition 2.14 (Preorder). A relation which is both reflexive and transitive is called a *preorder.*

Definition 2.15 (Partial order). A preorder which is also anti-symmetric is called a *partial order*.

Definition 2.16 (Linear order). A partial order which is also connected is called a *total order* or *linear order.*

Example 2.17. Every linear order is also a partial order, and every partial order is also a preorder, but the converses don't hold. The universal relation on A is a preorder, since it is reflexive and transitive. But, if A has more than one element, the universal relation is not anti-symmetric, and so not a partial order.

Example 2.18. Consider the *no longer than* relation $\preccurlyeq$ on $\mathbb{B}^*$: $x \preccurlyeq y$ iff $\mathrm{len}(x) \leq \mathrm{len}(y)$. This is a preorder (reflexive and transitive), and even connected, but not a partial order, since it is not anti-symmetric. For instance, $01 \preccurlyeq 10$ and $10 \preccurlyeq 01$, but $01 \neq 10$.

Example 2.19. An important partial order is the relation $\subseteq$ on a set of sets. This is not in general a linear order, since if $a \neq b$ and we consider $\wp(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$, we see that $\{a\} \nsubseteq \{b\}$ and $\{a\} \neq \{b\}$ and $\{b\} \nsubseteq \{a\}$.

Example 2.20. The relation of *divisibility without remainder* gives us a partial order which isn't a linear order. For integers n, m, we write $n \mid m$ to mean n (evenly) divides m, i.e., iff there is some integer k so that $m = kn$. On $\mathbb{N}$, this is a partial order, but not a linear order: for instance, $2 \nmid 3$ and also $3 \nmid 2$. Considered as a relation on $\mathbb{Z}$, divisibility is only a preorder since it is not anti-symmetric: $1 \mid -1$ and $-1 \mid 1$ but $1 \neq -1$.

Definition 2.21 (Strict order). A *strict order* is a relation which is irreflexive, asymmetric, and transitive.

Definition 2.22 (Strict linear order). A strict order which is also connected is called a *strict linear order.*

Example 2.23. $\leq$ is the linear order corresponding to the strict linear order $<$. $\subseteq$ is the partial order corresponding to the strict order $\subsetneq$.

Definition 2.24 (Total order). A strict order which is also connected is called a *total order*. This is also sometimes called a *strict linear order*.

Any strict order R on A can be turned into a partial order by adding the diagonal Id_A, i.e., adding all the pairs $\langle x, x\rangle$. (This is called the *reflexive closure* of R.) Conversely, starting from a partial order, one can get a strict order by removing Id_A. These next two results make this precise.

Proposition 2.25. *If R is a strict order on A, then $R^+ = R \cup \mathrm{Id}_A$ is a partial order. Moreover, if R is total, then R^+ is a linear order.*

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

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

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

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

Concerning the "moreover" clause, suppose R is a total order, i.e., that R is connected. So for all $x \neq y$, either Rxy or Ryx, i.e., either $\langle x, y\rangle \in R$ or $\langle y, x\rangle \in R$. Since $R \subseteq R^+$, this remains true of R^+, so R^+ is connected as well. □

Proposition 2.26. *If R is a partial order on X, then $R^- = R \setminus \mathrm{Id}_X$ is a strict order. Moreover, if R is linear, then R^- is total.*

Proof. This is left as an exercise. □

Example 2.27. $\leq$ is the linear order corresponding to the total order $<$. $\subseteq$ is the partial order corresponding to the strict order $\subsetneq$.

The following simple result which establishes that total orders satisfy an extensionality-like property:

Proposition 2.28. *If $<$ totally orders A, then:*

$$(\forall a, b \in A)((\forall x \in A)(x < a \leftrightarrow x < b) \rightarrow a = b)$$

Proof. Suppose $(\forall x \in A)(x < a \leftrightarrow x < b)$. If $a < b$, then $a < a$, contradicting the fact that $<$ is irreflexive; so $a \not< b$. Exactly similarly, $b \not< a$. So $a = b$, as $<$ is connected. □

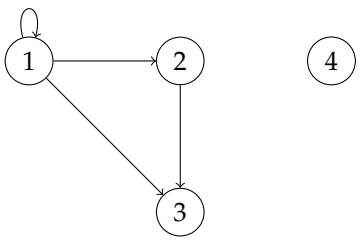
2.5 Graphs

A *graph* is a diagram in which points—called "nodes" or "vertices" (plural of "vertex")—are connected by edges. Graphs are a ubiquitous tool in discrete mathematics and in computer science. They are incredibly useful for representing, and visualizing, relationships and structures, from concrete things like networks of various kinds to abstract structures such as the possible outcomes of decisions. There are many different kinds of graphs in the literature which differ, e.g., according to whether the edges are directed or not, have labels or not, whether there can be edges from a node to the same node, multiple edges between the same nodes, etc. *Directed graphs* have a special connection to relations.

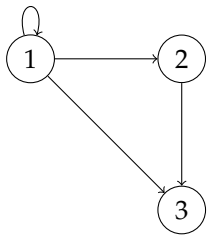
Definition 2.29 (Directed graph). A *directed graph* $G = \langle V, E \rangle$ is a set of *vertices* V and a set of *edges* $E \subseteq V^2$.

According to our definition, a graph just is a set together with a relation on that set. Of course, when talking about graphs, it's only natural to expect that they are graphically represented: we can draw a graph by connecting two vertices v_1 and v_2 by an arrow iff $\langle v_1, v_2 \rangle \in E$. The only difference between a relation by itself and a graph is that a graph specifies the set of vertices, i.e., a graph may have isolated vertices. The important point, however, is that every relation R on a set X can be seen as a directed graph $\langle X, R \rangle$, and conversely, a directed graph $\langle V, E \rangle$ can be seen as a relation $E \subseteq V^2$ with the set V explicitly specified.

Example 2.30. The graph $\langle V, E \rangle$ with $V = \{1, 2, 3, 4\}$ and $E = \{\langle 1, 1 \rangle, \langle 1, 2 \rangle, \langle 1, 3 \rangle, \langle 2, 3 \rangle\}$ looks like this:



This is a different graph than $\langle V', E\rangle$ with $V' = \{1, 2, 3\}$, which looks like this:



2.6 Operations on Relations

It is often useful to modify or combine relations. In Proposition 2.25, we considered the *union* of relations, which is just the union of two relations considered as sets of pairs. Similarly, in Proposition 2.26, we considered the relative difference of relations. Here are some other operations we can perform on relations.

Definition 2.31. Let R, S be relations, and A be any set.
The *inverse* of R is $R^{-1} = \{\langle y, x\rangle : \langle x, y\rangle \in R\}$.
The *relative product* of R and S is $(R \mid S) = \{\langle x, z\rangle : \exists y(Rxy \wedge Syz)\}$.
The *restriction* of R to A is $R{\restriction_A} = R \cap A^2$.
The *application* of R to A is $R[A] = \{y : (\exists x \in A)Rxy\}$

Example 2.32. Let $S \subseteq \mathbb{Z}^2$ be the successor relation on $\mathbb{Z}$, i.e., $S = \{\langle x, y\rangle \in \mathbb{Z}^2 : x + 1 = y\}$, so that Sxy iff $x + 1 = y$.
S^{-1} is the predecessor relation on $\mathbb{Z}$, i.e., $\{\langle x, y\rangle \in \mathbb{Z}^2 : x - 1 = y\}$.
$S \mid S$ is $\{\langle x, y\rangle \in \mathbb{Z}^2 : x + 2 = y\}$
$S{\restriction_{\mathbb{N}}}$ is the successor relation on $\mathbb{N}$.
$S[\{1, 2, 3\}]$ is $\{2, 3, 4\}$.

Definition 2.33 (Transitive closure). Let $R \subseteq A^2$ be a binary relation.
The *transitive closure* of R is $R^+ = \bigcup_{0<n\in\mathbb{N}} R^n$, where we recursively define $R^1 = R$ and $R^{n+1} = R^n \mid R$.
The *reflexive transitive closure* of R is $R^* = R^+ \cup \mathrm{Id}_A$.

Example 2.34. Take the successor relation $S \subseteq \mathbb{Z}^2$. S^2xy iff $x + 2 = y$, S^3xy iff $x + 3 = y$, etc. So S^+xy iff $x + n = y$ for some $n \geq 1$. In other words, S^+xy iff $x < y$, and S^*xy iff $x \leq y$.

Chapter 3

Functions

3.1 Basics

A *function* is a map which sends each element of a given set to a specific element in some (other) given set. For instance, the operation of adding 1 defines a function: each number n is mapped to a unique number $n + 1$.

More generally, functions may take pairs, triples, etc., as inputs and return some kind of output. Many functions are familiar to us from basic arithmetic. For instance, addition and multiplication are functions. They take in two numbers and return a third.

In this mathematical, abstract sense, a function is a *black box*: what matters is only what output is paired with what input, not the method for calculating the output.

Definition 3.1 (Function). A *function* $f\colon A \to B$ is a mapping of each element of A to an element of B.

We call A the *domain* of f and B the *codomain* of f. The elements of A are called inputs or *arguments* of f, and the element of B that is paired with an argument x by f is called the *value of* f for argument x, written $f(x)$.

The *range* $\mathrm{ran}(f)$ of f is the subset of the codomain consisting of the values of f for some argument; $\mathrm{ran}(f) = \{f(x) : x \in A\}$.

The diagram in Figure 3.1 may help to think about functions. The ellipse on the left represents the function's *domain*; the ellipse on the right represents the function's *codomain*; and an arrow points from an *argument* in the domain to the corresponding *value* in the codomain.

Example 3.2. Multiplication takes pairs of natural numbers as inputs and maps them to natural numbers as outputs, so goes from $\mathbb{N} \times \mathbb{N}$ (the domain) to $\mathbb{N}$ (the codomain). As it turns out, the range is also $\mathbb{N}$, since every $n \in \mathbb{N}$ is $n \times 1$.

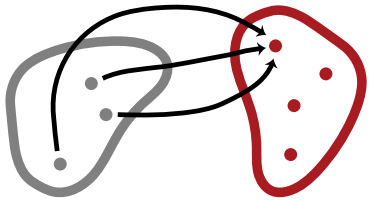


Figure 3.1: A function is a mapping of each element of one set to an element of another. An arrow points from an argument in the domain to the corresponding value in the codomain.

Example 3.3. Multiplication is a function because it pairs each input—each pair of natural numbers—with a single output: $\times : \mathbb{N}^2 \to \mathbb{N}$. By contrast, the square root operation applied to the domain $\mathbb{N}$ is not functional, since each positive integer n has two square roots: $\sqrt{n}$ and $-\sqrt{n}$. We can make it functional by only returning the positive square root: $\sqrt{\ } : \mathbb{N} \to \mathbb{R}$.

Example 3.4. The relation that pairs each student in a class with their final grade is a function—no student can get two different final grades in the same class. The relation that pairs each student in a class with their parents is not a function: students can have zero, or two, or more parents.

We can define functions by specifying in some precise way what the value of the function is for every possible argument. Different ways of doing this are by giving a formula, describing a method for computing the value, or listing the values for each argument. However functions are defined, we must make sure that for each argument we specify one, and only one, value.

Example 3.5. Let $f : \mathbb{N} \to \mathbb{N}$ be defined such that $f(x) = x + 1$. This is a definition that specifies f as a function which takes in natural numbers and outputs natural numbers. It tells us that, given a natural number x, f will output its successor $x + 1$. In this case, the codomain $\mathbb{N}$ is not the range of f, since the natural number 0 is not the successor of any natural number. The range of f is the set of all positive integers, $\mathbb{Z}^+$.

Example 3.6. Let $g : \mathbb{N} \to \mathbb{N}$ be defined such that $g(x) = x + 2 - 1$. This tells us that g is a function which takes in natural numbers and outputs natural numbers. Given a natural number n, g will output the predecessor of the successor of the successor of x, i.e., $x + 1$.

We just considered two functions, f and g, with different *definitions*. However, these are the *same function*. After all, for any natural number n, we have that $f(n) = n + 1 = n + 2 - 1 = g(n)$. Otherwise put: our definitions for f

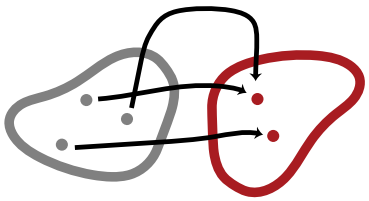


Figure 3.2: A surjective function has every element of the codomain as a value.

and g specify the same mapping by means of different equations. Implicitly, then, we are relying upon a principle of extensionality for functions,

$$\text{if } \forall x\, f(x) = g(x), \text{ then } f = g$$

provided that f and g share the same domain and codomain.

Example 3.7. We can also define functions by cases. For instance, we could define $h\colon \mathbb{N} \to \mathbb{N}$ by

$$h(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ \frac{x+1}{2} & \text{if } x \text{ is odd.} \end{cases}$$

Since every natural number is either even or odd, the output of this function will always be a natural number. Just remember that if you define a function by cases, every possible input must fall into exactly one case. In some cases, this will require a proof that the cases are exhaustive and exclusive.

3.2 Kinds of Functions

It will be useful to introduce a kind of taxonomy for some of the kinds of functions which we encounter most frequently.

To start, we might want to consider functions which have the property that every member of the codomain is a value of the function. Such functions are called surjective, and can be pictured as in Figure 3.2.

Definition 3.8 (Surjective function). A function $f\colon A \to B$ is *surjective* iff B is also the range of f, i.e., for every $y \in B$ there is at least one $x \in A$ such that $f(x) = y$, or in symbols:

$$(\forall y \in B)(\exists x \in A) f(x) = y.$$

We call such a function a surjection from A to B.

If you want to show that f is a surjection, then you need to show that every object in f's codomain is the value of $f(x)$ for some input x.

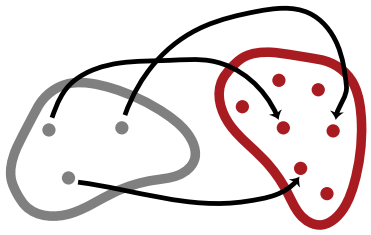


Figure 3.3: An injective function never maps two different arguments to the same value.

Note that any function *induces* a surjection. After all, given a function $f\colon A \to B$, let $f'\colon A \to \text{ran}(f)$ be defined by $f'(x) = f(x)$. Since $\text{ran}(f)$ is *defined* as $\{f(x) \in B : x \in A\}$, this function f' is guaranteed to be a surjection

Now, any function maps each possible input to a unique output. But there are also functions which never map different inputs to the same outputs. Such functions are called injective, and can be pictured as in Figure 3.3.

Definition 3.9 (Injective function). A function $f\colon A \to B$ is *injective* iff for each $y \in B$ there is at most one $x \in A$ such that $f(x) = y$. We call such a function an injection from A to B.

If you want to show that f is an injection, you need to show that for any elements x and y of f's domain, if $f(x) = f(y)$, then $x = y$.

Example 3.10. The constant function $f\colon \mathbb{N} \to \mathbb{N}$ given by $f(x) = 1$ is neither injective, nor surjective.

The identity function $f\colon \mathbb{N} \to \mathbb{N}$ given by $f(x) = x$ is both injective and surjective.

The successor function $f\colon \mathbb{N} \to \mathbb{N}$ given by $f(x) = x + 1$ is injective but not surjective.

The function $f\colon \mathbb{N} \to \mathbb{N}$ defined by:

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ \frac{x+1}{2} & \text{if } x \text{ is odd.} \end{cases}$$

is surjective, but not injective.

Often enough, we want to consider functions which are both injective and surjective. We call such functions bijective. They look like the function pictured in Figure 3.4. Bijections are also sometimes called *one-to-one correspondences*, since they uniquely pair elements of the codomain with elements of the domain.

Definition 3.11 (Bijection). A function $f\colon A \to B$ is *bijective* iff it is both surjective and injective. We call such a function a bijection from A to B (or between A and B).

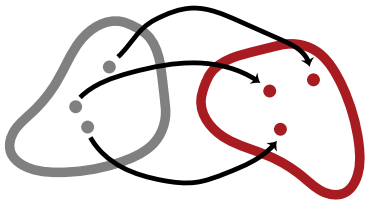


Figure 3.4: A bijective function uniquely pairs the elements of the codomain with those of the domain.

3.3 Functions as Relations

A function which maps elements of A to elements of B obviously defines a relation between A and B, namely the relation which holds between x and y iff $f(x) = y$. In fact, we might even—if we are interested in reducing the building blocks of mathematics for instance—*identify* the function f with this relation, i.e., with a set of pairs. This then raises the question: which relations define functions in this way?

Definition 3.12 (Graph of a function). Let $f : A \to B$ be a function. The *graph* of f is the relation $R_f \subseteq A \times B$ defined by

$$R_f = \{\langle x, y\rangle : f(x) = y\}.$$

The graph of a function is uniquely determined, by extensionality. Moreover, extensionality (on sets) will immediately vindicate the implicit principle of extensionality for functions, whereby if f and g share a domain and codomain then they are identical if they agree on all values.

Similarly, if a relation is "functional", then it is the graph of a function.

Proposition 3.13. *Let $R \subseteq A \times B$ be such that:*

1. *If Rxy and Rxz then $y = z$; and*
2. *for every $x \in A$ there is some $y \in B$ such that $\langle x, y\rangle \in R$.*

Then R is the graph of the function $f : A \to B$ defined by $f(x) = y$ iff Rxy.

Proof. Suppose there is a y such that Rxy. If there were another $z \neq y$ such that Rxz, the condition on R would be violated. Hence, if there is a y such that Rxy, this y is unique, and so f is well-defined. Obviously, $R_f = R$. □

Every function $f : A \to B$ has a graph, i.e., a relation on $A \times B$ defined by $f(x) = y$. On the other hand, every relation $R \subseteq A \times B$ with the properties given in Proposition 3.13 is the graph of a function $f : A \to B$. Because of this close connection between functions and their graphs, we can think of

a function simply as its graph. In other words, functions can be identified with certain relations, i.e., with certain sets of tuples. We can now consider performing similar operations on functions as we performed on relations (see section 2.6). In particular:

Definition 3.14. Let $f\colon A \to B$ be a function with $C \subseteq A$.

The *restriction* of f to C is the function $f{\restriction_C}\colon C \to B$ defined by $(f{\restriction_C})(x) = f(x)$ for all $x \in C$. In other words, $f{\restriction_C} = \{\langle x, y\rangle \in R_f : x \in C\}$.

The *application* of f to C is $f[C] = \{f(x) : x \in C\}$. We also call this the *image* of C under f.

It follows from these definitions that $\operatorname{ran}(f) = f[\operatorname{dom}(f)]$, for any function f. These notions are exactly as one would expect, given the definitions in section 2.6 and our identification of functions with relations. But two other operations—inverses and relative products—require a little more detail. We will provide that in section 3.4 and section 3.5.

3.4 Inverses of Functions

We think of functions as maps. An obvious question to ask about functions, then, is whether the mapping can be "reversed." For instance, the successor function $f(x) = x + 1$ can be reversed, in the sense that the function $g(y) = y - 1$ "undoes" what f does.

But we must be careful. Although the definition of g defines a function $\mathbb{Z} \to \mathbb{Z}$, it does not define a *function* $\mathbb{N} \to \mathbb{N}$, since $g(0) \notin \mathbb{N}$. So even in simple cases, it is not quite obvious whether a function can be reversed; it may depend on the domain and codomain.

This is made more precise by the notion of an inverse of a function.

Definition 3.15. A function $g\colon B \to A$ is an *inverse* of a function $f\colon A \to B$ if $f(g(y)) = y$ and $g(f(x)) = x$ for all $x \in A$ and $y \in B$.

If f has an inverse g, we often write f^{-1} instead of g.

Now we will determine when functions have inverses. A good candidate for an inverse of $f\colon A \to B$ is $g\colon B \to A$ "defined by"

$$g(y) = \text{"the" } x \text{ such that } f(x) = y.$$

But the scare quotes around "defined by" (and "the") suggest that this is not a definition. At least, it will not always work, with complete generality. For, in order for this definition to specify a function, there has to be one and only one x such that $f(x) = y$—the output of g has to be uniquely specified. Moreover, it has to be specified for every $y \in B$. If there are x_1 and $x_2 \in A$ with $x_1 \neq x_2$ but $f(x_1) = f(x_2)$, then $g(y)$ would not be uniquely specified for $y = f(x_1) = f(x_2)$. And if there is no x at all such that $f(x) = y$, then $g(y)$ is

not specified at all. In other words, for g to be defined, f must be both injective and surjective.

Let's go slowly. We'll divide the question into two: Given a function $f\colon A \to B$, when is there a function $g\colon B \to A$ so that $g(f(x)) = x$? Such a g "undoes" what f does, and is called a *left inverse* of f. Secondly, when is there a function $h\colon B \to A$ so that $f(h(y)) = y$? Such an h is called a *right inverse* of f—f "undoes" what h does.

Proposition 3.16. *If $f\colon A \to B$ is injective, then there is a* left inverse *$g\colon B \to A$ of f so that $g(f(x)) = x$ for all $x \in A$.*

Proof. Suppose that $f\colon A \to B$ is injective. Consider a $y \in B$. If $y \in \operatorname{ran}(f)$, there is an $x \in A$ so that $f(x) = y$. Because f is injective, there is only one such $x \in A$. Then we can define: $g(y) = x$, i.e., $g(y)$ is "the" $x \in A$ such that $f(x) = y$. If $y \notin \operatorname{ran}(f)$, we can map it to any $a \in A$. So, we can pick an $a \in A$ and define $g\colon B \to A$ by:

$$g(y) = \begin{cases} x & \text{if } f(x) = y \\ a & \text{if } y \notin \operatorname{ran}(f). \end{cases}$$

It is defined for all $y \in B$, since for each such $y \in \operatorname{ran}(f)$ there is exactly one $x \in A$ such that $f(x) = y$. By definition, if $y = f(x)$, then $g(y) = x$, i.e., $g(f(x)) = x$. □

Proposition 3.17. *If $f\colon A \to B$ is surjective, then there is a* right inverse *$h\colon B \to A$ of f so that $f(h(y)) = y$ for all $y \in B$.*

Proof. Suppose that $f\colon A \to B$ is surjective. Consider a $y \in B$. Since f is surjective, there is an $x_y \in A$ with $f(x_y) = y$. Then we can define: $h(y) = x_y$, i.e., for each $y \in B$ we choose some $x \in A$ so that $f(x) = y$; since f is surjective there is always at least one to choose from.[1] By definition, if $x = h(y)$, then $f(x) = y$, i.e., for any $y \in B$, $f(h(y)) = y$. □

By combining the ideas in the previous proof, we now get that every bijection has an inverse, i.e., there is a single function which is both a left and right inverse of f.

Proposition 3.18. *If $f\colon A \to B$ is bijective, there is a function $f^{-1}\colon B \to A$ so that for all $x \in A$, $f^{-1}(f(x)) = x$ and for all $y \in B$, $f(f^{-1}(y)) = y$.*

[1] Since f is surjective, for every $y \in B$ the set $\{x : f(x) = y\}$ is nonempty. Our definition of h requires that we choose a single x from each of these sets. That this is always possible is actually not obvious—the possibility of making these choices is simply assumed as an axiom. In other words, this proposition assumes the so-called Axiom of Choice, an issue we will gloss over. However, in many specific cases, e.g., when $A = \mathbb{N}$ or is finite, or when f is bijective, the Axiom of Choice is not required. (In the particular case when f is bijective, for each $y \in B$ the set $\{x : f(x) = y\}$ has exactly one element, so that there is no choice to make.)

Proof. Exercise. □

There is a slightly more general way to extract inverses. We saw in section 3.2 that every function f induces a surjection $f' \colon A \to \text{ran}(f)$ by letting $f'(x) = f(x)$ for all $x \in A$. Clearly, if f is injective, then f' is bijective, so that it has a unique inverse by Proposition 3.18. By a very minor abuse of notation, we sometimes call the inverse of f' simply "the inverse of f."

Proposition 3.19. *Show that if $f \colon A \to B$ has a left inverse g and a right inverse h, then $h = g$.*

Proof. Exercise. □

Proposition 3.20. *Every function f has at most one inverse.*

Proof. Suppose g and h are both inverses of f. Then in particular g is a left inverse of f and h is a right inverse. By Proposition 3.19, $g = h$. □

3.5 Composition of Functions

We saw in section 3.4 that the inverse f^{-1} of a bijection f is itself a function. Another operation on functions is composition: we can define a new function by composing two functions, f and g, i.e., by first applying f and then g. Of course, this is only possible if the ranges and domains match, i.e., the range of f must be a subset of the domain of g. This operation on functions is the analogue of the operation of relative product on relations from section 2.6.

A diagram might help to explain the idea of composition. In Figure 3.5, we depict two functions $f \colon A \to B$ and $g \colon B \to C$ and their composition $(g \circ f)$. The function $(g \circ f) \colon A \to C$ pairs each element of A with an element of C. We specify which element of C an element of A is paired with as follows: given an input $x \in A$, first apply the function f to x, which will output some $f(x) = y \in B$, then apply the function g to y, which will output some $g(f(x)) = g(y) = z \in C$.

Definition 3.21 (Composition). Let $f \colon A \to B$ and $g \colon B \to C$ be functions. The *composition* of f with g is $g \circ f \colon A \to C$, where $(g \circ f)(x) = g(f(x))$.

Example 3.22. Consider the functions $f(x) = x + 1$, and $g(x) = 2x$. Since $(g \circ f)(x) = g(f(x))$, for each input x you must first take its successor, then multiply the result by two. So their composition is given by $(g \circ f)(x) = 2(x + 1)$.

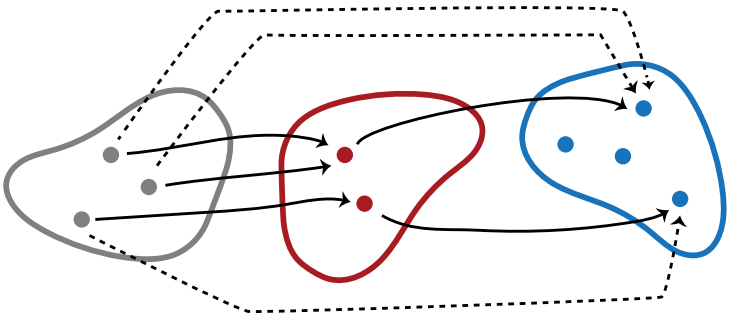


Figure 3.5: The composition $g \circ f$ of two functions f and g.

3.6 Partial Functions

It is sometimes useful to relax the definition of function so that it is not required that the output of the function is defined for all possible inputs. Such mappings are called *partial functions*.

Definition 3.23. A *partial function* $f\colon A \nrightarrow B$ is a mapping which assigns to every element of A at most one element of B. If f assigns an element of B to $x \in A$, we say $f(x)$ is *defined*, and otherwise *undefined*. If $f(x)$ is defined, we write $f(x)\downarrow$, otherwise $f(x)\uparrow$. The *domain* of a partial function f is the subset of A where it is defined, i.e., $\mathrm{dom}(f) = \{x \in A : f(x)\downarrow\}$.

Example 3.24. Every function $f\colon A \to B$ is also a partial function. Partial functions that are defined everywhere on A—i.e., what we so far have simply called a function—are also called *total* functions.

Example 3.25. The partial function $f\colon \mathbb{R} \nrightarrow \mathbb{R}$ given by $f(x) = 1/x$ is undefined for $x = 0$, and defined everywhere else.

Definition 3.26 (Graph of a partial function). Let $f\colon A \nrightarrow B$ be a partial function. The *graph* of f is the relation $R_f \subseteq A \times B$ defined by

$$R_f = \{\langle x, y\rangle : f(x) = y\}.$$

Proposition 3.27. *Suppose $R \subseteq A \times B$ has the property that whenever Rxy and Rxy' then $y = y'$. Then R is the graph of the partial function $f\colon X \nrightarrow Y$ defined by: if there is a y such that Rxy, then $f(x) = y$, otherwise $f(x)\uparrow$. If R is also* serial, *i.e., for each $x \in X$ there is a $y \in Y$ such that Rxy, then f is total.*

Proof. Suppose there is a y such that Rxy. If there were another $y' \neq y$ such that Rxy', the condition on R would be violated. Hence, if there is a y such that Rxy, that y is unique, and so f is well-defined. Obviously, $R_f = R$ and f is total if R is serial. □

Chapter 4

The Size of Sets

4.1 Introduction

When Georg Cantor developed set theory in the 1870s, one of his aims was to make palatable the idea of an infinite collection—an actual infinity, as the medievals would say. A key part of this was his treatment of the *size* of different sets. If a, b and c are all distinct, then the set $\{a, b, c\}$ is intuitively *larger* than $\{a, b\}$. But what about infinite sets? Are they all as large as each other? It turns out that they are not.

The first important idea here is that of an enumeration. We can list every finite set by listing all its elements. For some infinite sets, we can also list all their elements if we allow the list itself to be infinite. Such sets are called enumerable. Cantor's surprising result, which we will fully understand by the end of this chapter, was that some infinite sets are not enumerable.

4.2 Enumerations and Enumerable Sets

We've already given examples of sets by listing their elements. Let's discuss in more general terms how and when we can list the elements of a set, even if that set is infinite.

Definition 4.1 (Enumeration, informally). Informally, an *enumeration* of a set A is a list (possibly infinite) of elements of A such that every element of A appears on the list at some finite position. If A has an enumeration, then A is said to be *enumerable*.

A couple of points about enumerations:

1. We count as enumerations only lists which have a beginning and in which every element other than the first has a single element immediately preceding it. In other words, there are only finitely many elements between the first element of the list and any other element. In particular,

this means that every element of an enumeration has a finite position: the first element has position 1, the second position 2, etc.

2. We can have different enumerations of the same set A which differ by the order in which the elements appear: 4, 1, 25, 16, 9 enumerates the (set of the) first five square numbers just as well as 1, 4, 9, 16, 25 does.

3. Redundant enumerations are still enumerations: 1, 1, 2, 2, 3, 3, ... enumerates the same set as 1, 2, 3, ... does.

4. Order and redundancy *do* matter when we specify an enumeration: we can enumerate the positive integers beginning with 1, 2, 3, 1, ..., but the pattern is easier to see when enumerated in the standard way as 1, 2, 3, 4, ...

5. Enumerations must have a beginning: ..., 3, 2, 1 is not an enumeration of the positive integers because it has no first element. To see how this follows from the informal definition, ask yourself, "at what position in the list does the number 76 appear?"

6. The following is not an enumeration of the positive integers: 1, 3, 5, ..., 2, 4, 6, ... The problem is that the even numbers occur at places $\infty + 1$, $\infty + 2$, $\infty + 3$, rather than at finite positions.

7. The empty set is enumerable: it is enumerated by the empty list!

Proposition 4.2. *If A has an enumeration, it has an enumeration without repetitions.*

Proof. Suppose A has an enumeration $x_1, x_2, \ldots$ in which each x_i is an element of A. We can remove repetitions from an enumeration by removing repeated elements. For instance, we can turn the enumeration into a new one in which we list x_i if it is an element of A that is not among $x_1, \ldots, x_{i-1}$ or remove x_i from the list if it already appears among $x_1, \ldots, x_{i-1}$. □

The last argument shows that in order to get a good handle on enumerations and enumerable sets and to prove things about them, we need a more precise definition. The following provides it.

Definition 4.3 (Enumeration, formally). An *enumeration* of a set $A \neq \emptyset$ is any surjective function $f\colon \mathbb{Z}^+ \to A$.

Let's convince ourselves that the formal definition and the informal definition using a possibly infinite list are equivalent. First, any surjective function from $\mathbb{Z}^+$ to a set A enumerates A. Such a function determines an enumeration as defined informally above: the list $f(1), f(2), f(3), \ldots$. Since f is surjective, every element of A is guaranteed to be the value of $f(n)$ for some $n \in \mathbb{Z}^+$.

Hence, every element of A appears at some finite position in the list. Since the function may not be injective, the list may be redundant, but that is acceptable (as noted above).

On the other hand, given a list that enumerates all elements of A, we can define a surjective function $f\colon \mathbb{Z}^+ \to A$ by letting $f(n)$ be the nth element of the list, or the final element of the list if there is no nth element. The only case where this does not produce a surjective function is when A is empty, and hence the list is empty. So, every non-empty list determines a surjective function $f\colon \mathbb{Z}^+ \to A$.

Definition 4.4. A set A is enumerable iff it is empty or has an enumeration.

Example 4.5. A function enumerating the positive integers ($\mathbb{Z}^+$) is simply the identity function given by $f(n) = n$. A function enumerating the natural numbers $\mathbb{N}$ is the function $g(n) = n - 1$.

Example 4.6. The functions $f\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ and $g\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ given by

$$f(n) = 2n \text{ and}$$
$$g(n) = 2n - 1$$

enumerate the even positive integers and the odd positive integers, respectively. However, neither function is an enumeration of $\mathbb{Z}^+$, since neither is surjective.

Example 4.7. The function $f(n) = (-1)^n \lceil \frac{(n-1)}{2} \rceil$ (where $\lceil x \rceil$ denotes the *ceiling* function, which rounds x up to the nearest integer) enumerates the set of integers $\mathbb{Z}$. Notice how f generates the values of $\mathbb{Z}$ by "hopping" back and forth between positive and negative integers:

$$\begin{array}{cccccccc} f(1) & f(2) & f(3) & f(4) & f(5) & f(6) & f(7) & \dots \\ -\lceil \frac{0}{2} \rceil & \lceil \frac{1}{2} \rceil & -\lceil \frac{2}{2} \rceil & \lceil \frac{3}{2} \rceil & -\lceil \frac{4}{2} \rceil & \lceil \frac{5}{2} \rceil & -\lceil \frac{6}{2} \rceil & \dots \\ 0 & 1 & -1 & 2 & -2 & 3 & \dots & \end{array}$$

You can also think of f as defined by cases as follows:

$$f(n) = \begin{cases} 0 & \text{if } n = 1 \\ n/2 & \text{if } n \text{ is even} \\ -(n-1)/2 & \text{if } n \text{ is odd and} > 1 \end{cases}$$

Although it is perhaps more natural when listing the elements of a set to start counting from the 1st element, mathematicians like to use the natural numbers $\mathbb{N}$ for counting things. They talk about the 0th, 1st, 2nd, and so on, elements of a list. Correspondingly, we can define an enumeration as a surjective function from $\mathbb{N}$ to A. Of course, the two definitions are equivalent.

Proposition 4.8. *There is a surjection $f\colon \mathbb{Z}^+ \to A$ iff there is a surjection $g\colon \mathbb{N} \to A$.*

Proof. Given a surjection $f\colon \mathbb{Z}^+ \to A$, we can define $g(n) = f(n+1)$ for all $n \in \mathbb{N}$. It is easy to see that $g\colon \mathbb{N} \to A$ is surjective. Conversely, given a surjection $g\colon \mathbb{N} \to A$, define $f(n) = g(n-1)$. □

This gives us the following result:

Corollary 4.9. *A set A is enumerable iff it is empty or there is a surjective function $f\colon \mathbb{N} \to A$.*

We discussed above than an list of elements of a set A can be turned into a list without repetitions. This is also true for enumerations, but a bit harder to formulate and prove rigorously. Any function $f\colon \mathbb{Z}^+ \to A$ must be defined for all $n \in \mathbb{Z}^+$. If there are only finitely many elements in A then we clearly cannot have a function defined on the infinitely many elements of $\mathbb{Z}^+$ that takes as values all the elements of A but never takes the same value twice. In that case, i.e., in the case where the list without repetitions is finite, we must choose a different domain for f, one with only finitely many elements. Not having repetitions means that f must be injective. Since it is also surjective, we are looking for a bijection between some finite set $\{1, \ldots, n\}$ or $\mathbb{Z}^+$ and A.

Proposition 4.10. *If $f\colon \mathbb{Z}^+ \to A$ is surjective (i.e., an enumeration of A), there is a bijection $g\colon Z \to A$ where Z is either $\mathbb{Z}^+$ or $\{1, \ldots, n\}$ for some $n \in \mathbb{Z}^+$.*

Proof. We define the function g recursively: Let $g(1) = f(1)$. If $g(i)$ has already been defined, let $g(i+1)$ be the first value of $f(1), f(2), \ldots$ not already among $g(1), \ldots, g(i)$, if there is one. If A has just n elements, then $g(1), \ldots, g(n)$ are all defined, and so we have defined a function $g\colon \{1, \ldots, n\} \to A$. If A has infinitely many elements, then for any i there must be an element of A in the enumeration $f(1), f(2), \ldots$, which is not already among $g(1), \ldots, g(i)$. In this case we have defined a funtion $g\colon \mathbb{Z}^+ \to A$.

The function g is surjective, since any element of A is among $f(1), f(2), \ldots$ (since f is surjective) and so will eventually be a value of $g(i)$ for some i. It is also injective, since if there were $j < i$ such that $g(j) = g(i)$, then $g(i)$ would already be among $g(1), \ldots, g(i-1)$, contrary to how we defined g. □

Corollary 4.11. *A set A is enumerable iff it is empty or there is a bijection $f\colon N \to A$ where either $N = \mathbb{N}$ or $N = \{0, \ldots, n\}$ for some $n \in \mathbb{N}$.*

Proof. A is enumerable iff A is empty or there is a surjective $f\colon \mathbb{Z}^+ \to A$. By Proposition 4.10, the latter holds iff there is a bijective function $f\colon Z \to A$ where $Z = \mathbb{Z}^+$ or $Z = \{1, \ldots, n\}$ for some $n \in \mathbb{Z}^+$. By the same argument as in the proof of Proposition 4.8, that in turn is the case iff there is a bijection $g\colon N \to A$ where either $N = \mathbb{N}$ or $N = \{0, \ldots, n-1\}$. □

4.3 Cantor's Zig-Zag Method

We've already considered some "easy" enumerations. Now we will consider something a bit harder. Consider the set of pairs of natural numbers, which we defined in section 1.5 thus:

$$\mathbb{N} \times \mathbb{N} = \{\langle n, m\rangle : n, m \in \mathbb{N}\}$$

We can organize these ordered pairs into an *array*, like so:

	0	**1**	**2**	**3**	...
0	$\langle 0,0\rangle$	$\langle 0,1\rangle$	$\langle 0,2\rangle$	$\langle 0,3\rangle$	...
1	$\langle 1,0\rangle$	$\langle 1,1\rangle$	$\langle 1,2\rangle$	$\langle 1,3\rangle$	...
2	$\langle 2,0\rangle$	$\langle 2,1\rangle$	$\langle 2,2\rangle$	$\langle 2,3\rangle$	...
3	$\langle 3,0\rangle$	$\langle 3,1\rangle$	$\langle 3,2\rangle$	$\langle 3,3\rangle$	...
⋮	⋮	⋮	⋮	⋮	⋱

Clearly, every ordered pair in $\mathbb{N} \times \mathbb{N}$ will appear exactly once in the array. In particular, $\langle n, m\rangle$ will appear in the nth row and mth column. But how do we organize the elements of such an array into a "one-dimensional" list? The pattern in the array below demonstrates one way to do this (although of course there are many other options):

	0	**1**	**2**	**3**	**4**	...
0	0	1	3	6	10	...
1	2	4	7	11	...	...
2	5	8	12	...	...	...
3	9	13	...	...	...	...
4	14	...	...	...	...	...
⋮	⋮	⋮	⋮	⋮	...	⋱

This pattern is called *Cantor's zig-zag method*. It enumerates $\mathbb{N} \times \mathbb{N}$ as follows:

$$\langle 0,0\rangle, \langle 0,1\rangle, \langle 1,0\rangle, \langle 0,2\rangle, \langle 1,1\rangle, \langle 2,0\rangle, \langle 0,3\rangle, \langle 1,2\rangle, \langle 2,1\rangle, \langle 3,0\rangle, \ldots$$

And this establishes the following:

Proposition 4.12. $\mathbb{N} \times \mathbb{N}$ *is enumerable.*

Proof. Let $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ take each $k \in \mathbb{N}$ to the tuple $\langle n, m\rangle \in \mathbb{N} \times \mathbb{N}$ such that k is the value of the nth row and mth column in Cantor's zig-zag array. □

This technique also generalises rather nicely. For example, we can use it to enumerate the set of ordered triples of natural numbers, i.e.:

$$\mathbb{N} \times \mathbb{N} \times \mathbb{N} = \{\langle n, m, k\rangle : n, m, k \in \mathbb{N}\}$$

We think of $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ as the Cartesian product of $\mathbb{N} \times \mathbb{N}$ with $\mathbb{N}$, that is,

$$\mathbb{N}^3 = (\mathbb{N} \times \mathbb{N}) \times \mathbb{N} = \{\langle\langle n, m\rangle, k\rangle : n, m, k \in \mathbb{N}\}$$

and thus we can enumerate $\mathbb{N}^3$ with an array by labelling one axis with the enumeration of $\mathbb{N}$, and the other axis with the enumeration of $\mathbb{N}^2$:

	0	**1**	**2**	**3**	...
$\langle \mathbf{0}, \mathbf{0}\rangle$	$\langle 0,0,0\rangle$	$\langle 0,0,1\rangle$	$\langle 0,0,2\rangle$	$\langle 0,0,3\rangle$	...
$\langle \mathbf{0}, \mathbf{1}\rangle$	$\langle 0,1,0\rangle$	$\langle 0,1,1\rangle$	$\langle 0,1,2\rangle$	$\langle 0,1,3\rangle$	...
$\langle \mathbf{1}, \mathbf{0}\rangle$	$\langle 1,0,0\rangle$	$\langle 1,0,1\rangle$	$\langle 1,0,2\rangle$	$\langle 1,0,3\rangle$	...
$\langle \mathbf{0}, \mathbf{2}\rangle$	$\langle 0,2,0\rangle$	$\langle 0,2,1\rangle$	$\langle 0,2,2\rangle$	$\langle 0,2,3\rangle$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Thus, by using a method like Cantor's zig-zag method, we may similarly obtain an enumeration of $\mathbb{N}^3$. And we can keep going, obtaining enumerations of $\mathbb{N}^n$ for any natural number n. So, we have:

Proposition 4.13. *$\mathbb{N}^n$ is enumerable, for every $n \in \mathbb{N}$.*

4.4 Pairing Functions and Codes

Cantor's zig-zag method makes the enumerability of $\mathbb{N}^n$ visually evident. But let us focus on our array depicting $\mathbb{N}^2$. Following the zig-zag line in the array and counting the places, we can check that $\langle 1,2\rangle$ is associated with the number 7. However, it would be nice if we could compute this more directly. That is, it would be nice to have to hand the *inverse* of the zig-zag enumeration, $g\colon \mathbb{N}^2 \to \mathbb{N}$, such that

$$g(\langle 0,0\rangle) = 0,\ g(\langle 0,1\rangle) = 1,\ g(\langle 1,0\rangle) = 2,\ \ldots, g(\langle 1,2\rangle) = 7,\ \ldots$$

This would enable us to calculate exactly where $\langle n, m\rangle$ will occur in our enumeration.

In fact, we can define g directly by making two observations. First: if the nth row and mth column contains value v, then the $(n+1)$st row and $(m-1)$st column contains value $v+1$. Second: the first row of our enumeration consists of the triangular numbers, starting with 0, 1, 3, 6, etc. The kth triangular number is the sum of the natural numbers $< k$, which can be computed as $k(k+1)/2$. Putting these two observations together, consider this function:

$$g(n, m) = \frac{(n+m+1)(n+m)}{2} + n$$

We often just write $g(n, m)$ rather that $g(\langle n, m\rangle)$, since it is easier on the eyes. This tells you first to determine the $(n+m)^{\text{th}}$ triangle number, and then add

n to it. And it populates the array in exactly the way we would like. So in particular, the pair $\langle 1, 2\rangle$ is sent to $\frac{4\times 3}{2} + 1 = 7$.

This function g is the *inverse* of an enumeration of a set of pairs. Such functions are called *pairing functions*.

Definition 4.14 (Pairing function). A function $f\colon A \times B \to \mathbb{N}$ is an arithmetical *pairing function* if f is injective. We also say that f *encodes* $A \times B$, and that $f(x, y)$ is the *code* for $\langle x, y\rangle$.

We can use pairing functions to encode, e.g., pairs of natural numbers; or, in other words, we can represent each *pair* of elements using a *single* number. Using the inverse of the pairing function, we can *decode* the number, i.e., find out which pair it represents.

4.5 An Alternative Pairing Function

There are other enumerations of $\mathbb{N}^2$ that make it easier to figure out what their inverses are. Here is one. Instead of visualizing the enumeration in an array, start with the list of positive integers associated with (initially) empty spaces. Imagine filling these spaces successively with pairs $\langle n, m\rangle$ as follows. Starting with the pairs that have 0 in the first place (i.e., pairs $\langle 0, m\rangle$), put the first (i.e., $\langle 0, 0\rangle$) in the first empty place, then skip an empty space, put the second (i.e., $\langle 0, 2\rangle$) in the next empty place, skip one again, and so forth. The (incomplete) beginning of our enumeration now looks like this

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	...
$\langle 0,1\rangle$		$\langle 0,2\rangle$		$\langle 0,3\rangle$		$\langle 0,4\rangle$		$\langle 0,5\rangle$		...

Repeat this with pairs $\langle 1, m\rangle$ for the place that still remain empty, again skipping every other empty place:

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	...
$\langle 0,0\rangle$	$\langle 1,0\rangle$	$\langle 0,1\rangle$		$\langle 0,2\rangle$	$\langle 1,1\rangle$	$\langle 0,3\rangle$		$\langle 0,4\rangle$	$\langle 1,2\rangle$	...

Enter pairs $\langle 2, m\rangle$, $\langle 2, m\rangle$, etc., in the same way. Our completed enumeration thus starts like this:

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	...
$\langle 0,0\rangle$	$\langle 1,0\rangle$	$\langle 0,1\rangle$	$\langle 2,0\rangle$	$\langle 0,2\rangle$	$\langle 1,1\rangle$	$\langle 0,3\rangle$	$\langle 3,0\rangle$	$\langle 0,4\rangle$	$\langle 1,2\rangle$	...

If we number the cells in the array above according to this enumeration, we will not find a neat zig-zag line, but this arrangement:

	0	**1**	**2**	**3**	**4**	**5**	...
0	1	3	5	7	9	11	...
1	2	6	10	14	18	...	...
2	4	12	20	28	...	...	...
3	8	24	40	...	...	...	...
4	16	48	...	...	...	...	...
5	32	...	...	...	...	...	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

We can see that the pairs in row 0 are in the odd numbered places of our enumeration, i.e., pair $\langle 0, m\rangle$ is in place $2m+1$; pairs in the second row, $\langle 1, m\rangle$, are in places whose number is the double of an odd number, specifically, $2 \cdot (2m+1)$; pairs in the third row, $\langle 2, m\rangle$, are in places whose number is four times an odd number, $4 \cdot (2m+1)$; and so on. The factors of $(2m+1)$ for each row, 1, 2, 4, 8, ..., are exactly the powers of 2: $1 = 2^0$, $2 = 2^1$, $4 = 2^2$, $8 = 2^3$, ... In fact, the relevant exponent is always the first member of the pair in question. Thus, for pair $\langle n, m\rangle$ the factor is 2^n. This gives us the general formula: $2^n \cdot (2m+1)$. However, this is a mapping of pairs to *positive* integers, i.e., $\langle 0, 0\rangle$ has position 1. If we want to begin at position 0 we must subtract 1 from the result. This gives us:

Example 4.15. The function $h\colon \mathbb{N}^2 \to \mathbb{N}$ given by

$$h(n, m) = 2^n(2m+1) - 1$$

is a pairing function for the set of pairs of natural numbers $\mathbb{N}^2$.

Accordingly, in our second enumeration of $\mathbb{N}^2$, the pair $\langle 0, 0\rangle$ has code $h(0, 0) = 2^0(2 \cdot 0 + 1) - 1 = 0$; $\langle 1, 2\rangle$ has code $2^1 \cdot (2 \cdot 2 + 1) - 1 = 2 \cdot 5 - 1 = 9$; $\langle 2, 6\rangle$ has code $2^2 \cdot (2 \cdot 6 + 1) - 1 = 51$.

Sometimes it is enough to encode pairs of natural numbers $\mathbb{N}^2$ without requiring that the encoding is surjective. Such encodings have inverses that are only partial functions.

Example 4.16. The function $j\colon \mathbb{N}^2 \to \mathbb{N}^+$ given by

$$j(n, m) = 2^n 3^m$$

is an injective function $\mathbb{N}^2 \to \mathbb{N}$.

4.6 Non-enumerable Sets

Some sets, such as the set $\mathbb{Z}^+$ of positive integers, are infinite. So far we've seen examples of infinite sets which were all enumerable. However, there are also infinite sets which do not have this property. Such sets are called *non-enumerable*.

First of all, it is perhaps already surprising that there are non-enumerable sets. For any enumerable set A there is a surjective function $f\colon \mathbb{Z}^+ \to A$. If a set is non-enumerable there is no such function. That is, no function mapping the infinitely many elements of $\mathbb{Z}^+$ to A can exhaust all of A. So there are "more" elements of A than the infinitely many positive integers.

How would one prove that a set is non-enumerable? You have to show that no such surjective function can exist. Equivalently, you have to show that the elements of A cannot be enumerated in a one way infinite list. The best way to do this is to show that every list of elements of A must leave at least one element out; or that no function $f\colon \mathbb{Z}^+ \to A$ can be surjective. We can do this using Cantor's *diagonal method*. Given a list of elements of A, say, x_1, x_2, ..., we construct another element of A which, by its construction, cannot possibly be on that list.

Our first example is the set $\mathbb{B}^\omega$ of all infinite, non-gappy sequences of 0's and 1's.

Theorem 4.17. *$\mathbb{B}^\omega$ is non-enumerable.*

Proof. Suppose, by way of contradiction, that $\mathbb{B}^\omega$ is enumerable, i.e., suppose that there is a list s_1, s_2, s_3, s_4, ... of all elements of $\mathbb{B}^\omega$. Each of these s_i is itself an infinite sequence of 0's and 1's. Let's call the j-th element of the i-th sequence in this list $s_i(j)$. Then the i-th sequence s_i is

$$s_i(1), s_i(2), s_i(3), \ldots$$

We may arrange this list, and the elements of each sequence s_i in it, in an array:

	1	2	3	4	...
1	$\mathbf{s_1(1)}$	$s_1(2)$	$s_1(3)$	$s_1(4)$	...
2	$s_2(1)$	$\mathbf{s_2(2)}$	$s_2(3)$	$s_2(4)$	...
3	$s_3(1)$	$s_3(2)$	$\mathbf{s_3(3)}$	$s_3(4)$	...
4	$s_4(1)$	$s_4(2)$	$s_4(3)$	$\mathbf{s_4(4)}$	...
⋮	⋮	⋮	⋮	⋮	⋱

The labels down the side give the number of the sequence in the list $s_1, s_2, \ldots$; the numbers across the top label the elements of the individual sequences. For instance, $s_1(1)$ is a name for whatever number, a 0 or a 1, is the first element in the sequence s_1, and so on.

Now we construct an infinite sequence, $\bar{s}$, of 0's and 1's which cannot possibly be on this list. The definition of $\bar{s}$ will depend on the list s_1, s_2, ... Any infinite list of infinite sequences of 0's and 1's gives rise to an infinite sequence $\bar{s}$ which is guaranteed to not appear on the list.

To define $\bar{s}$, we specify what all its elements are, i.e., we specify $\bar{s}(n)$ for all $n \in \mathbb{Z}^+$. We do this by reading down the diagonal of the array above (hence the name "diagonal method") and then changing every 1 to a 0 and every 0 to a 1. More abstractly, we define $\bar{s}(n)$ to be 0 or 1 according to whether the n-th element of the diagonal, $s_n(n)$, is 1 or 0.

$$\bar{s}(n) = \begin{cases} 1 & \text{if } s_n(n) = 0 \\ 0 & \text{if } s_n(n) = 1. \end{cases}$$

If you like formulas better than definitions by cases, you could also define $\bar{s}(n) = 1 - s_n(n)$.

Clearly $\bar{s}$ is an infinite sequence of 0's and 1's, since it is just the mirror sequence to the sequence of 0's and 1's that appear on the diagonal of our array. So $\bar{s}$ is an element of $\mathbb{B}^\omega$. But it cannot be on the list s_1, s_2, ... Why not?

It can't be the first sequence in the list, s_1, because it differs from s_1 in the first element. Whatever $s_1(1)$ is, we defined $\bar{s}(1)$ to be the opposite. It can't be the second sequence in the list, because $\bar{s}$ differs from s_2 in the second element: if $s_2(2)$ is 0, $\bar{s}(2)$ is 1, and vice versa. And so on.

More precisely: if $\bar{s}$ were on the list, there would be some k so that $\bar{s} = s_k$. Two sequences are identical iff they agree at every place, i.e., for any n, $\bar{s}(n) = s_k(n)$. So in particular, taking $n = k$ as a special case, $\bar{s}(k) = s_k(k)$ would have to hold. $s_k(k)$ is either 0 or 1. If it is 0 then $\bar{s}(k)$ must be 1—that's how we defined $\bar{s}$. But if $s_k(k) = 1$ then, again because of the way we defined $\bar{s}$, $\bar{s}(k) = 0$. In either case $\bar{s}(k) \neq s_k(k)$.

We started by assuming that there is a list of elements of $\mathbb{B}^\omega$, s_1, s_2, ... From this list we constructed a sequence $\bar{s}$ which we proved cannot be on the list. But it definitely is a sequence of 0's and 1's if all the s_i are sequences of 0's and 1's, i.e., $\bar{s} \in \mathbb{B}^\omega$. This shows in particular that there can be no list of *all* elements of $\mathbb{B}^\omega$, since for any such list we could also construct a sequence $\bar{s}$ guaranteed to not be on the list, so the assumption that there is a list of all sequences in $\mathbb{B}^\omega$ leads to a contradiction. □

This proof method is called "diagonalization" because it uses the diagonal of the array to define $\bar{s}$. Diagonalization need not involve the presence of an array: we can show that sets are not enumerable by using a similar idea even when no array and no actual diagonal is involved.

Theorem 4.18. *$\wp(\mathbb{Z}^+)$ is not enumerable.*

Proof. We proceed in the same way, by showing that for every list of subsets of $\mathbb{Z}^+$ there is a subset of $\mathbb{Z}^+$ which cannot be on the list. Suppose the following is a given list of subsets of $\mathbb{Z}^+$:

$$Z_1, Z_2, Z_3, \ldots$$

We now define a set $\overline{Z}$ such that for any $n \in \mathbb{Z}^+$, $n \in \overline{Z}$ iff $n \notin Z_n$:

$$\overline{Z} = \{n \in \mathbb{Z}^+ : n \notin Z_n\}$$ □

$\overline{Z}$ is clearly a set of positive integers, since by assumption each Z_n is, and thus $\overline{Z} \in \wp(\mathbb{Z}^+)$. But $\overline{Z}$ cannot be on the list. To show this, we'll establish that for each $k \in \mathbb{Z}^+$, $\overline{Z} \neq Z_k$.

So let $k \in \mathbb{Z}^+$ be arbitrary. We've defined $\overline{Z}$ so that for any $n \in \mathbb{Z}^+$, $n \in \overline{Z}$ iff $n \notin Z_n$. In particular, taking $n = k$, $k \in \overline{Z}$ iff $k \notin Z_k$. But this shows that $\overline{Z} \neq Z_k$, since k is an element of one but not the other, and so $\overline{Z}$ and Z_k have different elements. Since k was arbitrary, $\overline{Z}$ is not on the list $Z_1, Z_2, \ldots$

The preceding proof did not mention a diagonal, but you can think of it as involving a diagonal if you picture it this way: Imagine the sets $Z_1, Z_2, \ldots$, written in an array, where each element $j \in Z_i$ is listed in the j-th column. Say the first four sets on that list are $\{1, 2, 3, \ldots\}$, $\{2, 4, 6, \ldots\}$, $\{1, 2, 5\}$, and $\{3, 4, 5, \ldots\}$. Then the array would begin with

$$\begin{array}{lccccccc}
Z_1 = \{ & \mathbf{1}, & 2, & 3, & 4, & 5, & 6, & \ldots\} \\
Z_2 = \{ & & \mathbf{2}, & & 4, & & 6, & \ldots\} \\
Z_3 = \{ & 1, & 2, & & & 5 & & \} \\
Z_4 = \{ & & & 3, & \mathbf{4}, & 5, & 6, & \ldots\} \\
\vdots & & & & & \ddots & &
\end{array}$$

Then $\overline{Z}$ is the set obtained by going down the diagonal, leaving out any numbers that appear along the diagonal and include those j where the array has a gap in the j-th row/column. In the above case, we would leave out 1 and 2, include 3, leave out 4, etc.

4.7 Reduction

We showed $\wp(\mathbb{Z}^+)$ to be non-enumerable by a diagonalization argument. We already had a proof that $\mathbb{B}^\omega$, the set of all infinite sequences of 0s and 1s, is non-enumerable. Here's another way we can prove that $\wp(\mathbb{Z}^+)$ is non-enumerable: Show that *if* $\wp(\mathbb{Z}^+)$ *is enumerable then* $\mathbb{B}^\omega$ *is also enumerable*. Since we know $\mathbb{B}^\omega$ is not enumerable, $\wp(\mathbb{Z}^+)$ can't be either. This is called *reducing* one problem to another—in this case, we reduce the problem of enumerating $\mathbb{B}^\omega$ to the problem of enumerating $\wp(\mathbb{Z}^+)$. A solution to the latter—an enumeration of $\wp(\mathbb{Z}^+)$—would yield a solution to the former—an enumeration of $\mathbb{B}^\omega$.

How do we reduce the problem of enumerating a set B to that of enumerating a set A? We provide a way of turning an enumeration of A into an enumeration of B. The easiest way to do that is to define a surjective function $f\colon A \to B$. If $x_1, x_2, \ldots$ enumerates A, then $f(x_1), f(x_2), \ldots$ would enumerate B. In our case, we are looking for a surjective function $f\colon \wp(\mathbb{Z}^+) \to \mathbb{B}^\omega$.

Proof of Theorem 4.18 by reduction. Suppose that $\wp(\mathbb{Z}^+)$ were enumerable, and thus that there is an enumeration of it, $Z_1, Z_2, Z_3, \ldots$

Define the function $f\colon \wp(\mathbb{Z}^+) \to \mathbb{B}^\omega$ by letting $f(Z)$ be the sequence s_k such that $s_k(n) = 1$ iff $n \in Z$, and $s_k(n) = 0$ otherwise. This clearly defines a function, since whenever $Z \subseteq \mathbb{Z}^+$, any $n \in \mathbb{Z}^+$ either is an element of Z or isn't. For instance, the set $2\mathbb{Z}^+ = \{2, 4, 6, \ldots\}$ of positive even numbers gets mapped to the sequence $010101\ldots$, the empty set gets mapped to $0000\ldots$ and the set $\mathbb{Z}^+$ itself to $1111\ldots$.

It also is surjective: Every sequence of 0s and 1s corresponds to some set of positive integers, namely the one which has as its members those integers corresponding to the places where the sequence has 1s. More precisely, suppose $s \in \mathbb{B}^\omega$. Define $Z \subseteq \mathbb{Z}^+$ by:

$$Z = \{n \in \mathbb{Z}^+ : s(n) = 1\}$$

Then $f(Z) = s$, as can be verified by consulting the definition of f.

Now consider the list

$$f(Z_1), f(Z_2), f(Z_3), \ldots$$

Since f is surjective, every member of $\mathbb{B}^\omega$ must appear as a value of f for some argument, and so must appear on the list. This list must therefore enumerate all of $\mathbb{B}^\omega$.

So if $\wp(\mathbb{Z}^+)$ were enumerable, $\mathbb{B}^\omega$ would be enumerable. But $\mathbb{B}^\omega$ is non-enumerable (Theorem 4.17). Hence $\wp(\mathbb{Z}^+)$ is non-enumerable. □

It is easy to be confused about the direction the reduction goes in. For instance, a surjective function $g\colon \mathbb{B}^\omega \to B$ does *not* establish that B is non-enumerable. (Consider $g\colon \mathbb{B}^\omega \to \mathbb{B}$ defined by $g(s) = s(1)$, the function that maps a sequence of 0's and 1's to its first element. It is surjective, because some sequences start with 0 and some start with 1. But $\mathbb{B}$ is finite.) Note also that the function f must be surjective, or otherwise the argument does not go through: $f(x_1), f(x_2), \ldots$ would then not be guaranteed to include all the elements of B. For instance,

$$h(n) = \underbrace{000\ldots0}_{n\ 0\text{'s}}$$

defines a function $h\colon \mathbb{Z}^+ \to \mathbb{B}^\omega$, but $\mathbb{Z}^+$ is enumerable.

4.8 Equinumerosity

We have an intuitive notion of "size" of sets, which works fine for finite sets. But what about infinite sets? If we want to come up with a formal way of comparing the sizes of two sets of *any* size, it is a good idea to start by defining when sets are the same size. Here is Frege:

> If a waiter wants to be sure that he has laid exactly as many knives as plates on the table, he does not need to count either of them, if he simply lays a knife to the right of each plate, so that every knife on the table lies to the right of some plate. The plates and knives are thus uniquely correlated to each other, and indeed through that same spatial relationship. (Frege, 1884, §70)

The insight of this passage can be brought out through a formal definition:

Definition 4.19. A is *equinumerous* with B, written $A \approx B$, iff there is a bijection $f\colon A \to B$.

Proposition 4.20. *Equinumerosity is an equivalence relation.*

Proof. We must show that equinumerosity is reflexive, symmetric, and transitive. Let A, B, and C be sets.

Reflexivity. The identity map $\mathrm{Id}_A\colon A \to A$, where $\mathrm{Id}_A(x) = x$ for all $x \in A$, is a bijection. So $A \approx A$.

Symmetry. Suppose $A \approx B$, i.e., there is a bijection $f\colon A \to B$. Since f is bijective, its inverse f^{-1} exists and is also bijective. Hence, $f^{-1}\colon B \to A$ is a bijection, so $B \approx A$.

Transitivity. Suppose that $A \approx B$ and $B \approx C$, i.e., there are bijections $f\colon A \to B$ and $g\colon B \to C$. Then the composition $g \circ f\colon A \to C$ is bijective, so that $A \approx C$. □

Proposition 4.21. *If $A \approx B$, then A is enumerable if and only if B is.*

Proof. Suppose $A \approx B$, so there is some bijection $f\colon A \to B$, and suppose that A is enumerable. Then either $A = \emptyset$ or there is a surjective function $g\colon \mathbb{Z}^+ \to A$. If $A = \emptyset$, then $B = \emptyset$ also (otherwise there would be an element $y \in B$ but no $x \in A$ with $g(x) = y$). If, on the other hand, $g\colon \mathbb{Z}^+ \to A$ is surjective, then $f \circ g\colon \mathbb{Z}^+ \to B$ is surjective. To see this, let $y \in B$. Since f is surjective, there is an $x \in A$ such that $f(x) = y$. Since g is surjective, there is an $n \in \mathbb{Z}^+$ such that $g(n) = x$. Hence,

$$(f \circ g)(n) = f(g(n)) = f(x) = y$$

and thus $f \circ g$ is surjective. We have that $f \circ g$ is an enumeration of B, and so B is enumerable.

If B is enumerable, we obtain that A is enumerable by repeating the argument with the bijection $f^{-1}\colon B \to A$ instead of f. □

4.9 Sets of Different Sizes, and Cantor's Theorem

We have offered a precise statement of the idea that two sets have the same size. We can also offer a precise statement of the idea that one set is smaller than another. Our definition of "is smaller than (or equinumerous)" will require, instead of a bijection between the sets, an injection from the first set to the second. If such a function exists, the size of the first set is less than or equal to the size of the second. Intuitively, an injection from one set to another guarantees that the range of the function has at least as many elements as the domain, since no two elements of the domain map to the same element of the range.

Definition 4.22. A is *no larger than* B, written $A \preceq B$, iff there is an injection $f : A \to B$.

It is clear that this is a reflexive and transitive relation, but that it is not symmetric (this is left as an exercise). We can also introduce a notion, which states that one set is (strictly) smaller than another.

Definition 4.23. A is *smaller than* B, written $A \prec B$, iff there is an injection $f : A \to B$ but no bijection $g : A \to B$, i.e., $A \preceq B$ and $A \not\approx B$.

It is clear that this relation is irreflexive and transitive. (This is left as an exercise.) Using this notation, we can say that a set A is enumerable iff $A \preceq \mathbb{N}$, and that A is non-enumerable iff $\mathbb{N} \prec A$. This allows us to restate Theorem 4.18 as the observation that $\mathbb{Z}^+ \prec \wp(\mathbb{Z}^+)$. In fact, Cantor (1892) proved that this last point is *perfectly general*:

Theorem 4.24 (Cantor). $A \prec \wp(A)$, *for any set* A.

Proof. The map $f(x) = \{x\}$ is an injection $f : A \to \wp(A)$, since if $x \neq y$, then also $\{x\} \neq \{y\}$ by extensionality, and so $f(x) \neq f(y)$. So we have that $A \preceq \wp(A)$.

We will now show that there cannot be a surjective function $g : A \to \wp(A)$, let alone a bijective one, and hence that $A \not\approx \wp(A)$. For suppose that $g : A \to \wp(A)$. Since g is total, every $x \in A$ is mapped to a subset $g(x) \subseteq A$. We can show that g cannot be surjective. To do this, we define a subset $\overline{A} \subseteq A$ which by definition cannot be in the range of g. Let

$$\overline{A} = \{x \in A : x \notin g(x)\}.$$

Since $g(x)$ is defined for all $x \in A$, $\overline{A}$ is clearly a well-defined subset of A. But, it cannot be in the range of g. Let $x \in A$ be arbitrary, we will show that $\overline{A} \neq g(x)$. If $x \in g(x)$, then it does not satisfy $x \notin g(x)$, and so by the definition of $\overline{A}$, we have $x \notin \overline{A}$. If $x \in \overline{A}$, it must satisfy the defining property of $\overline{A}$, i.e., $x \in A$ and $x \notin g(x)$. Since x was arbitrary, this shows that for each

$x \in \overline{A}$, $x \in g(x)$ iff $x \notin \overline{A}$, and so $g(x) \neq \overline{A}$. In other words, $\overline{A}$ cannot be in the range of g, contradicting the assumption that g is surjective. □

It's instructive to compare the proof of Theorem 4.24 to that of Theorem 4.18. There we showed that for any list Z_1, Z_2, ..., of subsets of $\mathbb{Z}^+$ one can construct a set $\overline{Z}$ of numbers guaranteed not to be on the list. It was guaranteed not to be on the list because, for every $n \in \mathbb{Z}^+$, $n \in Z_n$ iff $n \notin \overline{Z}$. This way, there is always some number that is an element of one of Z_n or $\overline{Z}$ but not the other. We follow the same idea here, except the indices n are now elements of A instead of $\mathbb{Z}^+$. The set $\overline{B}$ is defined so that it is different from $g(x)$ for each $x \in A$, because $x \in g(x)$ iff $x \notin \overline{B}$. Again, there is always an element of A which is an element of one of $g(x)$ and $\overline{B}$ but not the other. And just as $\overline{Z}$ therefore cannot be on the list Z_1, Z_2, ..., $\overline{B}$ cannot be in the range of g.

The proof is also worth comparing with the proof of Russell's Paradox, Theorem 1.29. Indeed, Cantor's Theorem was the inspiration for Russell's own paradox.

4.10 The Notion of Size, and Schröder-Bernstein

Here is an intuitive thought: if A is no larger than B and B is no larger than A, then A and B are equinumerous. To be honest, if this thought were *wrong*, then we could scarcely justify the thought that our defined notion of equinumerosity has anything to do with comparisons of "sizes" between sets! Fortunately, though, the intuitive thought is correct. This is justified by the Schröder-Bernstein Theorem.

Theorem 4.25 (Schröder-Bernstein). *If $A \preceq B$ and $B \preceq A$, then $A \approx B$.*

In other words, if there is an injection from A to B, and an injection from B to A, then there is a bijection from A to B.

This result, however, is really rather *difficult* to prove. Indeed, although Cantor stated the result, others proved it.[1] For now, you can (and must) take it on trust.

Fortunately, Schröder-Bernstein is *correct*, and it vindicates our thinking of the relations we defined, i.e., $A \approx B$ and $A \preceq B$, as having something to do with "size". Moreover, Schröder-Bernstein is very *useful*. It can be difficult to think of a bijection between two equinumerous sets. The Schröder-Bernstein Theorem allows us to break the comparison down into cases so we only have to think of an injection from the first to the second, and vice-versa.

[1] For more on the history, see e.g., Potter (2004, pp. 165–6).

Part II

First-Order Logic

Chapter 5

Syntax and Semantics

5.1 Introduction

In order to develop the theory and metatheory of first-order logic, we must first define the syntax and semantics of its expressions. The expressions of first-order logic are terms and formulas. Terms are formed from variables, constant symbols, and function symbols. Formulas, in turn, are formed from predicate symbols together with terms (these form the smallest, "atomic" formulas), and then from atomic formulas we can form more complex ones using logical connectives and quantifiers. There are many different ways to set down the formation rules; we give just one possible one. Other systems will chose different symbols, will select different sets of connectives as primitive, will use parentheses differently (or even not at all, as in the case of so-called Polish notation). What all approaches have in common, though, is that the formation rules define the set of terms and formulas *inductively*. If done properly, every expression can result essentially in only one way according to the formation rules. The inductive definition resulting in expressions that are *uniquely readable* means we can give meanings to these expressions using the same method—inductive definition.

Giving the meaning of expressions is the domain of semantics. The central concept in semantics is that of satisfaction in a structure. A structure gives meaning to the building blocks of the language: a domain is a non-empty set of objects. The quantifiers are interpreted as ranging over this domain, constant symbols are assigned elements in the domain, function symbols are assigned functions from the domain to itself, and predicate symbols are assigned relations on the domain. The domain together with assignments to the basic vocabulary constitutes a structure. Variables may appear in formulas, and in order to give a semantics, we also have to assign elements of the domain to them—this is a variable assignment. The satisfaction relation, finally, brings these together. A formula may be satisfied in a structure $\mathfrak{M}$ relative to a variable assignment s, written as $\mathfrak{M}, s \vDash \varphi$. This relation is also defined by in-

duction on the structure of φ, using the truth tables for the logical connectives to define, say, satisfaction of $\varphi \wedge \psi$ in terms of satisfaction (or not) of φ and ψ. It then turns out that the variable assignment is irrelevant if the formula φ is a sentence, i.e., has no free variables, and so we can talk of sentences being simply satisfied (or not) in structures.

On the basis of the satisfaction relation $\mathfrak{M} \vDash \varphi$ for sentences we can then define the basic semantic notions of validity, entailment, and satisfiability. A sentence is valid, $\vDash \varphi$, if every structure satisfies it. It is entailed by a set of sentences, $\Gamma \vDash \varphi$, if every structure that satisfies all the sentences in Γ also satisfies φ. And a set of sentences is satisfiable if some structure satisfies all sentences in it at the same time. Because formulas are inductively defined, and satisfaction is in turn defined by induction on the structure of formulas, we can use induction to prove properties of our semantics and to relate the semantic notions defined.

5.2 First-Order Languages

Expressions of first-order logic are built up from a basic vocabulary containing *variables*, *constant symbols*, *predicate symbols* and sometimes *function symbols*. From them, together with logical connectives, quantifiers, and punctuation symbols such as parentheses and commas, *terms* and *formulas* are formed.

Informally, predicate symbols are names for properties and relations, constant symbols are names for individual objects, and function symbols are names for mappings. These, except for the identity predicate $=$, are the *non-logical symbols* and together make up a language. Any first-order language $\mathcal{L}$ is determined by its non-logical symbols. In the most general case, $\mathcal{L}$ contains infinitely many symbols of each kind.

In the general case, we make use of the following symbols in first-order logic:

1. Logical symbols
    a) Logical connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional), $\forall$ (universal quantifier), $\exists$ (existential quantifier).
    b) The propositional constant for falsity $\bot$.
    c) The two-place identity predicate $=$.
    d) A denumerable set of variables: v_0, v_1, v_2, ...
2. Non-logical symbols, making up the *standard language* of first-order logic
    a) A denumerable set of n-place predicate symbols for each $n > 0$: A^n_0, A^n_1, A^n_2, ...
    b) A denumerable set of constant symbols: c_0, c_1, c_2, ...

c) A denumerable set of n-place function symbols for each $n > 0$: f^n_0, f^n_1, f^n_2, ...

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

Most of our definitions and results will be formulated for the full standard language of first-order logic. However, depending on the application, we may also restrict the language to only a few predicate symbols, constant symbols, and function symbols.

Example 5.1. The language $\mathcal{L}_A$ of arithmetic contains a single two-place predicate symbol $<$, a single constant symbol 0, one one-place function symbol $'$, and two two-place function symbols $+$ and $\times$.

Example 5.2. The language of set theory $\mathcal{L}_Z$ contains only the single two-place predicate symbol $\in$.

Example 5.3. The language of orders $\mathcal{L}_\leq$ contains only the two-place predicate symbol $\leq$.

Again, these are conventions: officially, these are just aliases, e.g., $<$, $\in$, and $\leq$ are aliases for A^2_0, 0 for c_0, $'$ for f^1_0, $+$ for f^2_0, $\times$ for f^2_1.

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

A defined symbol is not officially part of the language, but is introduced as an informal abbreviation: it allows us to abbreviate formulas which would, if we only used primitive symbols, get quite long. This is obviously an advantage. The bigger advantage, however, is that proofs become shorter. If a symbol is primitive, it has to be treated separately in proofs. The more primitive symbols, therefore, the longer our proofs.

You may be familiar with different terminology and symbols than the ones we use above. Logic texts (and teachers) commonly use $\sim$, $\neg$, or ! for "negation", $\wedge$, $\cdot$, or & for "conjunction". Commonly used symbols for the "conditional" or "implication" are $\rightarrow$, $\Rightarrow$, and $\supset$. Symbols for "biconditional," "bi-implication," or "(material) equivalence" are $\leftrightarrow$, $\Leftrightarrow$, and $\equiv$. The $\bot$ symbol is variously called "falsity," "falsum,", "absurdity," or "bottom." The $\top$ symbol is variously called "truth," "verum," or "top."

It is conventional to use lower case letters (e.g., a, b, c) from the beginning of the Latin alphabet for constant symbols (sometimes called names), and lower case letters from the end (e.g., x, y, z) for variables. Quantifiers combine with variables, e.g., x; notational variations include $\forall x$, $(\forall x)$, (x), Πx, $\bigwedge_x$ for the universal quantifier and $\exists x$, $(\exists x)$, (Ex), Σx, $\bigvee_x$ for the existential quantifier.

We might treat all the propositional operators and both quantifiers as primitive symbols of the language. We might instead choose a smaller stock of

primitive symbols and treat the other logical operators as defined. "Truth functionally complete" sets of Boolean operators include $\{\neg, \vee\}$, $\{\neg, \wedge\}$, and $\{\neg, \rightarrow\}$—these can be combined with either quantifier for an expressively complete first-order language.

You may be familiar with two other logical operators: the Sheffer stroke $|$ (named after Henry Sheffer), and Peirce's arrow $\downarrow$, also known as Quine's dagger. When given their usual readings of "nand" and "nor" (respectively), these operators are truth functionally complete by themselves.

5.3 Terms and Formulas

Once a first-order language $\mathcal{L}$ is given, we can define expressions built up from the basic vocabulary of $\mathcal{L}$. These include in particular *terms* and *formulas*.

Definition 5.4 (Terms). The set of *terms* $\mathrm{Trm}(\mathcal{L})$ of $\mathcal{L}$ is defined inductively by:

1. Every variable is a term.
2. Every constant symbol of $\mathcal{L}$ is a term.
3. If f is an n-place function symbol and $t_1, \ldots, t_n$ are terms, then $f(t_1, \ldots, t_n)$ is a term.
4. Nothing else is a term.

A term containing no variables is a *closed term*.

The constant symbols appear in our specification of the language and the terms as a separate category of symbols, but they could instead have been included as zero-place function symbols. We could then do without the second clause in the definition of terms. We just have to understand $f(t_1, \ldots, t_n)$ as just f by itself if $n = 0$.

Definition 5.5 (Formulas). The set of *formulas* $\mathrm{Frm}(\mathcal{L})$ of the language $\mathcal{L}$ is defined inductively as follows:

1. $\bot$ is an atomic formula.
2. If R is an n-place predicate symbol of $\mathcal{L}$ and $t_1, \ldots, t_n$ are terms of $\mathcal{L}$, then $R(t_1, \ldots, t_n)$ is an atomic formula.
3. If t_1 and t_2 are terms of $\mathcal{L}$, then $=(t_1, t_2)$ is an atomic formula.
4. If φ is a formula, then $\neg\varphi$ is formula.
5. If φ and ψ are formulas, then $(\varphi \wedge \psi)$ is a formula.
6. If φ and ψ are formulas, then $(\varphi \vee \psi)$ is a formula.

7. If φ and ψ are formulas, then $(\varphi \to \psi)$ is a formula.
8. If φ is a formula and x is a variable, then $\forall x\, \varphi$ is a formula.
9. If φ is a formula and x is a variable, then $\exists x\, \varphi$ is a formula.
10. Nothing else is a formula.

The definitions of the set of terms and that of formulas are *inductive definitions*. Essentially, we construct the set of formulas in infinitely many stages. In the initial stage, we pronounce all atomic formulas to be formulas; this corresponds to the first few cases of the definition, i.e., the cases for $\bot$, $R(t_1, \ldots, t_n)$ and $=(t_1, t_2)$. "Atomic formula" thus means any formula of this form.

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

By convention, we write $=$ between its arguments and leave out the parentheses: $t_1 = t_2$ is an abbreviation for $=(t_1, t_2)$. Moreover, $\neg =(t_1, t_2)$ is abbreviated as $t_1 \neq t_2$. When writing a formula $(\psi * \chi)$ constructed from ψ, χ using a two-place connective $*$, we will often leave out the outermost pair of parentheses and write simply $\psi * \chi$.

Some logic texts require that the variable x must occur in φ in order for $\exists x\, \varphi$ and $\forall x\, \varphi$ to count as formulas. Nothing bad happens if you don't require this, and it makes things easier.

Definition 5.6. Formulas constructed using the defined operators are to be understood as follows:

1. $\top$ abbreviates $\neg \bot$.
2. $\varphi \leftrightarrow \psi$ abbreviates $(\varphi \to \psi) \wedge (\psi \to \varphi)$.

If we work in a language for a specific application, we will often write two-place predicate symbols and function symbols between the respective terms, e.g., $t_1 < t_2$ and $(t_1 + t_2)$ in the language of arithmetic and $t_1 \in t_2$ in the language of set theory. The successor function in the language of arithmetic is even written conventionally *after* its argument: t'. Officially, however, these are just conventional abbreviations for $A^2_0(t_1, t_2)$, $f^2_0(t_1, t_2)$, $A^2_0(t_1, t_2)$ and $f^1_0(t)$, respectively.

Definition 5.7 (Syntactic identity). The symbol $\equiv$ expresses syntactic identity between strings of symbols, i.e., $\varphi \equiv \psi$ iff φ and ψ are strings of symbols of the same length and which contain the same symbol in each place.

The $\equiv$ symbol may be flanked by strings obtained by concatenation, e.g., $\varphi \equiv (\psi \vee \chi)$ means: the string of symbols φ is the same string as the one obtained by concatenating an opening parenthesis, the string ψ, the $\vee$ symbol, the string χ, and a closing parenthesis, in this order. If this is the case, then we know that the first symbol of φ is an opening parenthesis, φ contains ψ as a substring (starting at the second symbol), that substring is followed by $\vee$, etc.

As terms and formulas are built up from basic elements via inductive definitions, we can use the following induction principles to prove things about them.

Lemma 5.8 (*Principle of induction on terms*). *Let $\mathcal{L}$ be a first-order language. If some property P holds in all of the following cases, then $P(t)$ for every $t \in \mathrm{Trm}(\mathcal{L})$.*

1. *$P(v)$ for every variable v,*
2. *$P(a)$ for every constant symbol a of $\mathcal{L}$,*
3. *If $t_1, \ldots, t_n \in \mathrm{Trm}(\mathcal{L})$, f is an n-place function symbol of $\mathcal{L}$, and $P(t_1), \ldots, P(t_n)$, then $P(f(t_1, \ldots, t_n))$.*

Lemma 5.9 (*Principle of induction on formulas*). *Let $\mathcal{L}$ be a first-order language. If some property P holds for all the atomic formulas and is such that*

1. *φ is an atomic formula.*
2. *it holds for $\neg\varphi$ whenever it holds for φ;*
3. *it holds for $(\varphi \wedge \psi)$ whenever it holds for φ and ψ;*
4. *it holds for $(\varphi \vee \psi)$ whenever it holds for φ and ψ;*
5. *it holds for $(\varphi \to \psi)$ whenever it holds for φ and ψ;*
6. *it holds for $\exists x\varphi$ whenever it holds for φ;*
7. *it holds for $\forall x\varphi$ whenever it holds for φ;*

then P holds for all formulas $\varphi \in \mathrm{Frm}(\mathcal{L})$.

5.4 Unique Readability

The way we defined formulas guarantees that every formula has a *unique reading*, i.e., there is essentially only one way of constructing it according to our formation rules for formulas and only one way of "interpreting" it. If this were not so, we would have ambiguous formulas, i.e., formulas that have more than one reading or intepretation—and that is clearly something we want to avoid. But more importantly, without this property, most of the definitions and proofs we are going to give will not go through.

Perhaps the best way to make this clear is to see what would happen if we had given bad rules for forming formulas that would not guarantee unique readability. For instance, we could have forgotten the parentheses in the formation rules for connectives, e.g., we might have allowed this:

If φ and ψ are formulas, then so is $\varphi \to \psi$.

Starting from an atomic formula θ, this would allow us to form $\theta \to \theta$. From this, together with θ, we would get $\theta \to \theta \to \theta$. But there are two ways to do this:

1. We take θ to be φ and $\theta \to \theta$ to be ψ.
2. We take φ to be $\theta \to \theta$ and ψ is θ.

Correspondingly, there are two ways to "read" the formula $\theta \to \theta \to \theta$. It is of the form $\psi \to \chi$ where ψ is θ and χ is $\theta \to \theta$, but *it is also* of the form $\psi \to \chi$ with ψ being $\theta \to \theta$ and χ being θ.

If this happens, our definitions will not always work. For instance, when we define the main operator of a formula, we say: in a formula of the form $\psi \to \chi$, the main operator is the indicated occurrence of $\to$. But if we can match the formula $\theta \to \theta \to \theta$ with $\psi \to \chi$ in the two different ways mentioned above, then in one case we get the first occurrence of $\to$ as the main operator, and in the second case the second occurrence. But we intend the main operator to be a *function* of the formula, i.e., every formula must have exactly one main operator occurrence.

Lemma 5.10. *The number of left and right parentheses in a formula φ are equal.*

Proof. We prove this by induction on the way φ is constructed. This requires two things: (a) We have to prove first that all atomic formulas have the property in question (the induction basis). (b) Then we have to prove that when we construct new formulas out of given formulas, the new formulas have the property provided the old ones do.

Let $l(\varphi)$ be the number of left parentheses, and $r(\varphi)$ the number of right parentheses in φ, and $l(t)$ and $r(t)$ similarly the number of left and right parentheses in a term t.

1. $\varphi \equiv \bot$: φ has 0 left and 0 right parentheses.
2. $\varphi \equiv R(t_1, \ldots, t_n)$: $l(\varphi) = 1 + l(t_1) + \cdots + l(t_n) = 1 + r(t_1) + \cdots + r(t_n) = r(\varphi)$. Here we make use of the fact, left as an exercise, that $l(t) = r(t)$ for any term t.
3. $\varphi \equiv t_1 = t_2$: $l(\varphi) = l(t_1) + l(t_2) = r(t_1) + r(t_2) = r(\varphi)$.
4. $\varphi \equiv \neg\psi$: By induction hypothesis, $l(\psi) = r(\psi)$. Thus $l(\varphi) = l(\psi) = r(\psi) = r(\varphi)$.

5. $\varphi \equiv (\psi * \chi)$: By induction hypothesis, $l(\psi) = r(\psi)$ and $l(\chi) = r(\chi)$. Thus $l(\varphi) = 1 + l(\psi) + l(\chi) = 1 + r(\psi) + r(\chi) = r(\varphi)$.

6. $\varphi \equiv \forall x\, \psi$: By induction hypothesis, $l(\psi) = r(\psi)$. Thus, $l(\varphi) = l(\psi) = r(\psi) = r(\varphi)$.

7. $\varphi \equiv \exists x\, \psi$: Similarly. □

Definition 5.11 (Proper prefix). A string of symbols ψ is a *proper prefix* of a string of symbols φ if concatenating ψ and a non-empty string of symbols yields φ.

Lemma 5.12. *If φ is a formula, and ψ is a proper prefix of φ, then ψ is not a formula.*

Proof. Exercise. □

Proposition 5.13. *If φ is an atomic formula, then it satisfies one, and only one of the following conditions.*

1. *$\varphi \equiv \bot$.*
2. *$\varphi \equiv R(t_1, \ldots, t_n)$ where R is an n-place predicate symbol, $t_1, \ldots, t_n$ are terms, and each of R, $t_1, \ldots, t_n$ is uniquely determined.*
3. *$\varphi \equiv t_1 = t_2$ where t_1 and t_2 are uniquely determined terms.*

Proof. Exercise. □

Proposition 5.14 (Unique Readability). *Every formula satisfies one, and only one of the following conditions.*

1. *φ is atomic.*
2. *φ is of the form $\neg\psi$.*
3. *φ is of the form $(\psi \wedge \chi)$.*
4. *φ is of the form $(\psi \vee \chi)$.*
5. *φ is of the form $(\psi \to \chi)$.*
6. *φ is of the form $\forall x\, \psi$.*
7. *φ is of the form $\exists x\, \psi$.*

Moreover, in each case ψ, or ψ and χ, are uniquely determined. This means that, e.g., there are no different pairs ψ, χ and ψ', χ' so that φ is both of the form $(\psi \to \chi)$ and $(\psi' \to \chi')$.

Proof. The formation rules require that if a formula is not atomic, it must start with an opening parenthesis (, $\neg$, or a quantifier. On the other hand, every formula that starts with one of the following symbols must be atomic: a predicate symbol, a function symbol, a constant symbol, $\bot$.

So we really only have to show that if φ is of the form $(\psi * \chi)$ and also of the form $(\psi' *' \chi')$, then $\psi \equiv \psi'$, $\chi \equiv \chi'$, and $* = *'$.

So suppose both $\varphi \equiv (\psi * \chi)$ and $\varphi \equiv (\psi' *' \chi')$. Then either $\psi \equiv \psi'$ or not. If it is, clearly $* = *'$ and $\chi \equiv \chi'$, since they then are substrings of φ that begin in the same place and are of the same length. The other case is $\psi \not\equiv \psi'$. Since ψ and ψ' are both substrings of φ that begin at the same place, one must be a proper prefix of the other. But this is impossible by Lemma 5.12. □

5.5 Main operator of a Formula

It is often useful to talk about the last operator used in constructing a formula φ. This operator is called the *main operator* of φ. Intuitively, it is the "outermost" operator of φ. For example, the main operator of $\neg\varphi$ is $\neg$, the main operator of $(\varphi \vee \psi)$ is $\vee$, etc.

Definition 5.15 (Main operator). The *main operator* of a formula φ is defined as follows:

1. φ is atomic: φ has no main operator.
2. $\varphi \equiv \neg\psi$: the main operator of φ is $\neg$.
3. $\varphi \equiv (\psi \wedge \chi)$: the main operator of φ is $\wedge$.
4. $\varphi \equiv (\psi \vee \chi)$: the main operator of φ is $\vee$.
5. $\varphi \equiv (\psi \to \chi)$: the main operator of φ is $\to$.
6. $\varphi \equiv \forall x\, \psi$: the main operator of φ is $\forall$.
7. $\varphi \equiv \exists x\, \psi$: the main operator of φ is $\exists$.

In each case, we intend the specific indicated *occurrence* of the main operator in the formula. For instance, since the formula $((\theta \to \alpha) \to (\alpha \to \theta))$ is of the form $(\psi \to \chi)$ where ψ is $(\theta \to \alpha)$ and χ is $(\alpha \to \theta)$, the second occurrence of $\to$ is the main operator.

This is a *recursive* definition of a function which maps all non-atomic formulas to their main operator occurrence. Because of the way formulas are defined inductively, every formula φ satisfies one of the cases in Definition 5.15. This guarantees that for each non-atomic formula φ a main operator exists. Because each formula satisfies only one of these conditions, and because the smaller formulas from which φ is constructed are uniquely determined in each

case, the main operator occurrence of φ is unique, and so we have defined a function.

We call formulas by the names in Table 5.1 depending on which symbol their main operator is.Recall, however, that defined operators do not officially appear in formulas. They are just abbreviations, so officially they cannot be the main operator of a formula. In proofs about all formulas they therefore do not have to be treated separately.

Main operator	Type of formula	Example
none	atomic (formula)	$\bot$, $R(t_1, \ldots, t_n)$, $t_1 = t_2$
$\neg$	negation	$\neg\varphi$
$\wedge$	conjunction	$(\varphi \wedge \psi)$
$\vee$	disjunction	$(\varphi \vee \psi)$
$\rightarrow$	conditional	$(\varphi \rightarrow \psi)$
$\leftrightarrow$	biconditional	$(\varphi \leftrightarrow \psi)$
$\forall$	universal (formula)	$\forall x\, \varphi$
$\exists$	existential (formula)	$\exists x\, \varphi$

Table 5.1: Main operator and names of formulas

5.6 Subformulas

It is often useful to talk about the formulas that "make up" a given formula. We call these its *subformulas*. Any formula counts as a subformula of itself; a subformula of φ other than φ itself is a *proper subformula*.

Definition 5.16 (Immediate Subformula). If φ is a formula, the *immediate subformulas* of φ are defined inductively as follows:

1. Atomic formulas have no immediate subformulas.
2. $\varphi \equiv \neg\psi$: The only immediate subformula of φ is ψ.
3. $\varphi \equiv (\psi * \chi)$: The immediate subformulas of φ are ψ and χ ($*$ is any one of the two-place connectives).
4. $\varphi \equiv \forall x\, \psi$: The only immediate subformula of φ is ψ.
5. $\varphi \equiv \exists x\, \psi$: The only immediate subformula of φ is ψ.

Definition 5.17 (Proper Subformula). If φ is a formula, the *proper subformulas* of φ are defined recursively as follows:

1. Atomic formulas have no proper subformulas.
2. $\varphi \equiv \neg\psi$: The proper subformulas of φ are ψ together with all proper subformulas of ψ.

3. $\varphi \equiv (\psi * \chi)$: The proper subformulas of φ are ψ, χ, together with all proper subformulas of ψ and those of χ.

4. $\varphi \equiv \forall x\, \psi$: The proper subformulas of φ are ψ together with all proper subformulas of ψ.

5. $\varphi \equiv \exists x\, \psi$: The proper subformulas of φ are ψ together with all proper subformulas of ψ.

Definition 5.18 (Subformula). The subformulas of φ are φ itself together with all its proper subformulas.

Note the subtle difference in how we have defined immediate subformulas and proper subformulas. In the first case, we have directly defined the immediate subformulas of a formula φ for each possible form of φ. It is an explicit definition by cases, and the cases mirror the inductive definition of the set of formulas. In the second case, we have also mirrored the way the set of all formulas is defined, but in each case we have also included the proper subformulas of the smaller formulas ψ, χ in addition to these formulas themselves. This makes the definition *recursive*. In general, a definition of a function on an inductively defined set (in our case, formulas) is recursive if the cases in the definition of the function make use of the function itself. To be well defined, we must make sure, however, that we only ever use the values of the function for arguments that come "before" the one we are defining—in our case, when defining "proper subformula" for $(\psi * \chi)$ we only use the proper subformulas of the "earlier" formulas ψ and χ.

Proposition 5.19. *Suppose ψ is a subformula of φ and χ is a subformula of ψ. Then χ is a subformula of φ. In other words, the subformula relation is transitive.*

Proposition 5.20. *Suppose φ is a formula with n connectives and quantifiers. Then φ has at most $2n + 1$ subformulas.*

5.7 Formation Sequences

Defining formulas via an inductive definition, and the complementary technique of proving properties of formulas via induction, is an elegant and efficient approach. However, it can also be useful to consider a more bottom-up, step-by-step approach to the construction of formulas, which we do here using the notion of a *formation sequence*. To show how terms and formulas can be introduced in this way without needing to refer to their inductive definitions, we first introduce the notion of an arbitrary string of symbols drawn from some language $\mathcal{L}$.

Definition 5.21 (Strings). Suppose $\mathcal{L}$ is a first-order language. An *$\mathcal{L}$-string* is a finite sequence of symbols of $\mathcal{L}$. Where the language $\mathcal{L}$ is clearly fixed by the context, we will often refer to a $\mathcal{L}$-string as a *string* simpliciter.

Example 5.22. For any first-order language $\mathcal{L}$, all $\mathcal{L}$-formulas are $\mathcal{L}$-strings, but not conversely. For example,

$$)(v_0 \to \exists$$

is an $\mathcal{L}$-string but not an $\mathcal{L}$-formula.

Definition 5.23 (Formation sequences for terms). A finite sequence of $\mathcal{L}$-strings $\langle t_0, \ldots, t_n \rangle$ is a *formation sequence* for a term t if $t \equiv t_n$ and for all $i \leq n$, either t_i is a variable or a constant symbol, or $\mathcal{L}$ contains a k-ary function symbol f and there exist $m_0, \ldots, m_k < i$ such that $t_i \equiv f(t_{m_0}, \ldots, t_{m_k})$.

Example 5.24. The sequence

$$\langle c_0, v_0, f_0^2(c_0, v_0), f_0^1(f_0^2(c_0, v_0)) \rangle$$

is a formation sequence for the term $f_0^1(f_0^2(c_0, v_0))$, as is

$$\langle v_0, c_0, f_0^2(c_0, v_0), f_0^1(f_0^2(c_0, v_0)) \rangle.$$

Definition 5.25 (Formation sequences for formulas). A finite sequence of $\mathcal{L}$-strings $\langle \varphi_0, \ldots, \varphi_n \rangle$ is a *formation sequence* for φ if $\varphi \equiv \varphi_n$ and for all $i \leq n$, either φ_i is an atomic formula or there exist $j, k < i$ and a variable x such that one of the following holds:

1. $\varphi_i \equiv \neg \varphi_j$.
2. $\varphi_i \equiv (\varphi_j \wedge \varphi_k)$.
3. $\varphi_i \equiv (\varphi_j \vee \varphi_k)$.
4. $\varphi_i \equiv (\varphi_j \to \varphi_k)$.
5. $\varphi_i \equiv \forall x \, \varphi_j$.
6. $\varphi_i \equiv \exists x \, \varphi_j$.

Example 5.26.

$$\langle A_0^1(v_0), A_1^1(c_1), (A_1^1(c_1) \wedge A_0^1(v_0)), \exists v_0 \, (A_1^1(c_1) \wedge A_0^1(v_0)) \rangle$$

is a formation sequence of $\exists v_0 \, (A_1^1(c_1) \wedge A_0^1(v_0))$, as is

$$\langle A_0^1(v_0), A_1^1(c_1), (A_1^1(c_1) \wedge A_0^1(v_0)), A_1^1(c_1), \forall v_1 \, A_0^1(v_0), \exists v_0 \, (A_1^1(c_1) \wedge A_0^1(v_0)) \rangle.$$

As can be seen from the second example, formation sequences may contain "junk": formulas which are redundant or do not contribute to the construction.

Proposition 5.27. *Every formula φ in* $\mathrm{Frm}(\mathcal{L})$ *has a formation sequence.*

Proof. Suppose φ is atomic. Then the sequence $\langle \varphi \rangle$ is a formation sequence for φ. Now suppose that ψ and χ have formation sequences $\langle \psi_0, \ldots, \psi_n \rangle$ and $\langle \chi_0, \ldots, \chi_m \rangle$ respectively.

1. If $\varphi \equiv \neg\psi$, then $\langle \psi_0, \ldots, \psi_n, \neg\psi_n \rangle$ is a formation sequence for φ.
2. If $\varphi \equiv (\psi \wedge \chi)$, then $\langle \psi_0, \ldots, \psi_n, \chi_0, \ldots, \chi_m, (\psi_n \wedge \chi_m) \rangle$ is a formation sequence for φ.
3. If $\varphi \equiv (\psi \vee \chi)$, then $\langle \psi_0, \ldots, \psi_n, \chi_0, \ldots, \chi_m, (\psi_n \vee \chi_m) \rangle$ is a formation sequence for φ.
4. If $\varphi \equiv (\psi \to \chi)$, then $\langle \psi_0, \ldots, \psi_n, \chi_0, \ldots, \chi_m, (\psi_n \to \chi_m) \rangle$ is a formation sequence for φ.
5. If $\varphi \equiv \forall x\, \psi$, then $\langle \psi_0, \ldots, \psi_n, \forall x\, \psi_n \rangle$ is a formation sequence for φ.
6. If $\varphi \equiv \exists x\, \psi$, then $\langle \psi_0, \ldots, \psi_n, \exists x\, \psi_n \rangle$ is a formation sequence for φ.

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

We can also prove the converse. This is important because it shows that our two ways of defining formulas are equivalent: they give the same results. It also means that we can prove theorems about formulas by using ordinary induction on the length of formation sequences.

Lemma 5.28. *Suppose that $\langle \varphi_0, \ldots, \varphi_n \rangle$ is a formation sequence for φ_n, and that $k \leq n$. Then $\langle \varphi_0, \ldots, \varphi_k \rangle$ is a formation sequence for φ_k.*

Proof. Exercise. □

Theorem 5.29. $\mathrm{Frm}(\mathcal{L})$ *is the set of all expressions (strings of symbols) in the language $\mathcal{L}$ with a formation sequence.*

Proof. Let F be the set of all strings of symbols in the language $\mathcal{L}$ that have a formation sequence. We have seen in Proposition 5.27 that $\mathrm{Frm}(\mathcal{L}) \subseteq F$, so now we prove the converse.

Suppose φ has a formation sequence $\langle \varphi_0, \ldots, \varphi_n \rangle$. We prove that $\varphi \in \mathrm{Frm}(\mathcal{L})$ by strong induction on n. Our induction hypothesis is that every string of symbols with a formation sequence of length $m < n$ is in $\mathrm{Frm}(\mathcal{L})$. By the definition of a formation sequence, either φ_n is atomic or there must exist $j, k < n$ such that one of the following is the case:

1. $\varphi_i \equiv \neg\varphi_j$.
2. $\varphi_i \equiv (\varphi_j \wedge \varphi_k)$.
3. $\varphi_i \equiv (\varphi_j \vee \varphi_k)$.
4. $\varphi_i \equiv (\varphi_j \to \varphi_k)$.
5. $\varphi_i \equiv \forall x\, \varphi_j$.
6. $\varphi_i \equiv \exists x\, \varphi_j$.

Now we reason by cases. If φ_n is atomic then $\varphi_n \in \mathrm{Frm}(\mathcal{L}_0)$. Suppose instead that $\varphi \equiv (\varphi_j \wedge \varphi_k)$. By Lemma 5.28, $\langle \varphi_0, \ldots, \varphi_j \rangle$ and $\langle \varphi_0, \ldots, \varphi_k \rangle$ are formation sequences for φ_j and φ_k, respectively. Since these are proper initial subsequences of the formation sequence for φ, they both have length less than n. Therefore by the induction hypothesis, φ_j and φ_k are in $\mathrm{Frm}(\mathcal{L}_0)$, and by the definition of a formula, so is $(\varphi_j \wedge \varphi_k)$. The other cases follow by parallel reasoning. □

Formation sequences for terms have similar properties to those for formulas.

Proposition 5.30. $\mathrm{Trm}(\mathcal{L})$ *is the set of all expressions t in the language $\mathcal{L}$ such that there exists a (term) formation sequence fo t.*

Proof. Exercise. □

There are two types of "junk" that can appear in formation sequences: repeated elements, and elements that are irrelevant to the construction of the formation or term. We can eliminate both by looking at minimal formation sequences.

Definition 5.31 (Minimal formation sequences). A formation sequence $\langle \varphi_0, \ldots, \varphi_n \rangle$ for φ is a *minimal formation sequence* for φ if for every other formation sequence s for φ, the length of s is greater than or equal to $n + 1$.

Proposition 5.32. *The following are equivalent:*

1. *ψ is a sub-formula of φ.*
2. *ψ occurs in every formation sequence of φ.*
3. *ψ occurs in a minimal formation sequence of φ.*

Proof. Exercise. □

Historical Remarks Formation sequences were introduced by Raymond Smullyan in his textbook *First-Order Logic* (Smullyan, 1968). Additional properties of formation sequences were established by Zuckerman (1973).

5.8 Free Variables and Sentences

Definition 5.33 (Free occurrences of a variable). The *free* occurrences of a variable in a formula are defined inductively as follows:

1. φ is atomic: all variable occurrences in φ are free.
2. $\varphi \equiv \neg\psi$: the free variable occurrences of φ are exactly those of ψ.
3. $\varphi \equiv (\psi * \chi)$: the free variable occurrences of φ are those in ψ together with those in χ.
4. $\varphi \equiv \forall x\, \psi$: the free variable occurrences in φ are all of those in ψ except for occurrences of x.
5. $\varphi \equiv \exists x\, \psi$: the free variable occurrences in φ are all of those in ψ except for occurrences of x.

Definition 5.34 (Bound Variables). An occurrence of a variable in a formula φ is *bound* if it is not free.

Definition 5.35 (Scope). If $\forall x\, \psi$ is an occurrence of a subformula in a formula φ, then the corresponding occurrence of ψ in φ is called the *scope* of the corresponding occurrence of $\forall x$. Similarly for $\exists x$.

If ψ is the scope of a quantifier occurrence $\forall x$ or $\exists x$ in φ, then the free occurrences of x in ψ are bound in $\forall x\, \psi$ and $\exists x\, \psi$. We say that these occurrences are *bound by* the mentioned quantifier occurrence.

Example 5.36. Consider the following formula:

$$\exists v_0 \underbrace{A_0^2(v_0, v_1)}_{\psi}$$

ψ represents the scope of $\exists v_0$. The quantifier binds the occurrence of v_0 in ψ, but does not bind the occurrence of v_1. So v_1 is a free variable in this case.

We can now see how this might work in a more complicated formula φ:

$$\forall v_0 \underbrace{(A_0^1(v_0) \to A_0^2(v_0, v_1))}_{\psi} \to \exists v_1 \underbrace{(A_1^2(v_0, v_1) \vee \forall v_0 \overbrace{\neg A_1^1(v_0)}^{\theta})}_{\chi}$$

ψ is the scope of the first $\forall v_0$, χ is the scope of $\exists v_1$, and θ is the scope of the second $\forall v_0$. The first $\forall v_0$ binds the occurrences of v_0 in ψ, $\exists v_1$ binds the occurrence of v_1 in χ, and the second $\forall v_0$ binds the occurrence of v_0 in θ. The first occurrence of v_1 and the fourth occurrence of v_0 are free in φ. The last occurrence of v_0 is free in θ, but bound in χ and φ.

Definition 5.37 (Sentence). A formula φ is a *sentence* iff it contains no free occurrences of variables.

5.9 Substitution

Definition 5.38 (Substitution in a term). We define $s[t/x]$, the result of *substituting* t for every occurrence of x in s, recursively:

1. $s \equiv c$: $s[t/x]$ is just s.
2. $s \equiv y$: $s[t/x]$ is also just s, provided y is a variable and $y \not\equiv x$.
3. $s \equiv x$: $s[t/x]$ is t.
4. $s \equiv f(t_1, \ldots, t_n)$: $s[t/x]$ is $f(t_1[t/x], \ldots, t_n[t/x])$.

Definition 5.39. A term t is *free for* x in φ if none of the free occurrences of x in φ occur in the scope of a quantifier that binds a variable in t.

Example 5.40.

1. v_8 is free for v_1 in $\exists v_3 A_4^2(v_3, v_1)$
2. $f_1^2(v_1, v_2)$ is *not* free for v_0 in $\forall v_2 A_4^2(v_0, v_2)$

Definition 5.41 (Substitution in a formula). If φ is a formula, x is a variable, and t is a term free for x in φ, then $\varphi[t/x]$ is the result of substituting t for all free occurrences of x in φ.

1. $\varphi \equiv \bot$: $\varphi[t/x]$ is $\bot$.
2. $\varphi \equiv P(t_1, \ldots, t_n)$: $\varphi[t/x]$ is $P(t_1[t/x], \ldots, t_n[t/x])$.
3. $\varphi \equiv t_1 = t_2$: $\varphi[t/x]$ is $t_1[t/x] = t_2[t/x]$.
4. $\varphi \equiv \neg\psi$: $\varphi[t/x]$ is $\neg\psi[t/x]$.
5. $\varphi \equiv (\psi \wedge \chi)$: $\varphi[t/x]$ is $(\psi[t/x] \wedge \chi[t/x])$.
6. $\varphi \equiv (\psi \vee \chi)$: $\varphi[t/x]$ is $(\psi[t/x] \vee \chi[t/x])$.
7. $\varphi \equiv (\psi \to \chi)$: $\varphi[t/x]$ is $(\psi[t/x] \to \chi[t/x])$.
8. $\varphi \equiv \forall y\, \psi$: $\varphi[t/x]$ is $\forall y\, \psi[t/x]$, provided y is a variable other than x; otherwise $\varphi[t/x]$ is just φ.
9. $\varphi \equiv \exists y\, \psi$: $\varphi[t/x]$ is $\exists y\, \psi[t/x]$, provided y is a variable other than x; otherwise $\varphi[t/x]$ is just φ.

Note that substitution may be vacuous: If x does not occur in φ at all, then $\varphi[t/x]$ is just φ.

The restriction that t must be free for x in φ is necessary to exclude cases like the following. If $\varphi \equiv \exists y\, x < y$ and $t \equiv y$, then $\varphi[t/x]$ would be $\exists y\, y < y$. In this case the free variable y is "captured" by the quantifier $\exists y$ upon substitution, and that is undesirable. For instance, we would like it to be the case that whenever $\forall x\, \psi$ holds, so does $\psi[t/x]$. But consider $\forall x\, \exists y\, x < y$ (here ψ is $\exists y\, x < y$). It is a sentence that is true about, e.g., the natural numbers: for every number x there is a number y greater than it. If we allowed y as a possible substitution for x, we would end up with $\psi[y/x] \equiv \exists y\, y < y$, which is false. We prevent this by requiring that none of the free variables in t would end up being bound by a quantifier in φ.

We often use the following convention to avoid cumbersome notation: If φ is a formula which may contain the variable x free, we also write $\varphi(x)$ to indicate this. When it is clear which φ and x we have in mind, and t is a term (assumed to be free for x in $\varphi(x)$), then we write $\varphi(t)$ as short for $\varphi[t/x]$. So for instance, we might say, "we call $\varphi(t)$ an instance of $\forall x\, \varphi(x)$." By this we mean that if φ is any formula, x a variable, and t a term that's free for x in φ, then $\varphi[t/x]$ is an instance of $\forall x\, \varphi$.

5.10 Structures for First-order Languages

First-order languages are, by themselves, *uninterpreted:* the constant symbols, function symbols, and predicate symbols have no specific meaning attached to them. Meanings are given by specifying a *structure*. It specifies the *domain*, i.e., the objects which the constant symbols pick out, the function symbols operate on, and the quantifiers range over. In addition, it specifies which constant symbols pick out which objects, how a function symbol maps objects to objects, and which objects the predicate symbols apply to. Structures are the basis for *semantic* notions in logic, e.g., the notion of consequence, validity, satisfiability. They are variously called "structures," "interpretations," or "models" in the literature.

Definition 5.42 (Structures). A *structure* $\mathfrak{M}$, for a language $\mathcal{L}$ of first-order logic consists of the following elements:

1. *Domain:* a non-empty set, $|\mathfrak{M}|$
2. *Interpretation of constant symbols:* for each constant symbol c of $\mathcal{L}$, an element $c^{\mathfrak{M}} \in |\mathfrak{M}|$
3. *Interpretation of predicate symbols:* for each n-place predicate symbol R of $\mathcal{L}$ (other than $=$), an n-place relation $R^{\mathfrak{M}} \subseteq |\mathfrak{M}|^n$
4. *Interpretation of function symbols:* for each n-place function symbol f of $\mathcal{L}$, an n-place function $f^{\mathfrak{M}} \colon |\mathfrak{M}|^n \to |\mathfrak{M}|$

Example 5.43. A structure $\mathfrak{M}$ for the language of arithmetic consists of a set, an element of $|\mathfrak{M}|$, $\mathsf{o}^{\mathfrak{M}}$, as interpretation of the constant symbol o, a one-place function $\mathsf{'}^{\mathfrak{M}}\colon |\mathfrak{M}| \to |\mathfrak{M}|$, two two-place functions $+^{\mathfrak{M}}$ and $\times^{\mathfrak{M}}$, both $|\mathfrak{M}|^2 \to |\mathfrak{M}|$, and a two-place relation $<^{\mathfrak{M}} \subseteq |\mathfrak{M}|^2$.

An obvious example of such a structure is the following:

1. $|\mathfrak{N}| = \mathbb{N}$
2. $\mathsf{o}^{\mathfrak{N}} = 0$
3. $\mathsf{'}^{\mathfrak{N}}(n) = n + 1$ for all $n \in \mathbb{N}$
4. $+^{\mathfrak{N}}(n, m) = n + m$ for all $n, m \in \mathbb{N}$
5. $\times^{\mathfrak{N}}(n, m) = n \cdot m$ for all $n, m \in \mathbb{N}$
6. $<^{\mathfrak{N}} = \{\langle n, m\rangle : n \in \mathbb{N}, m \in \mathbb{N}, n < m\}$

The structure $\mathfrak{N}$ for $\mathcal{L}_A$ so defined is called the *standard model of arithmetic*, because it interprets the non-logical constants of $\mathcal{L}_A$ exactly how you would expect.

However, there are many other possible structures for $\mathcal{L}_A$. For instance, we might take as the domain the set $\mathbb{Z}$ of integers instead of $\mathbb{N}$, and define the interpretations of o, $\mathsf{'}$, $+$, $\times$, $<$ accordingly. But we can also define structures for $\mathcal{L}_A$ which have nothing even remotely to do with numbers.

Example 5.44. A structure $\mathfrak{M}$ for the language $\mathcal{L}_Z$ of set theory requires just a set and a single-two place relation. So technically, e.g., the set of people plus the relation "x is older than y" could be used as a structure for $\mathcal{L}_Z$, as well as $\mathbb{N}$ together with $n \geq m$ for $n, m \in \mathbb{N}$.

A particularly interesting structure for $\mathcal{L}_Z$ in which the elements of the domain are actually sets, and the interpretation of $\in$ actually is the relation "x is an element of y" is the structure $\mathfrak{HF}$ of *hereditarily finite sets*:

1. $|\mathfrak{HF}| = \emptyset \cup \wp(\emptyset) \cup \wp(\wp(\emptyset)) \cup \wp(\wp(\wp(\emptyset))) \cup \ldots$;
2. $\in^{\mathfrak{HF}} = \{\langle x, y\rangle : x, y \in |\mathfrak{HF}|, x \in y\}$.

The stipulations we make as to what counts as a structure impact our logic. For example, the choice to prevent empty domains ensures, given the usual account of satisfaction (or truth) for quantified sentences, that $\exists x\,(\varphi(x) \vee \neg\varphi(x))$ is valid—that is, a logical truth. And the stipulation that all constant symbols must refer to an object in the domain ensures that the existential generalization is a sound pattern of inference: $\varphi(a)$, therefore $\exists x\,\varphi(x)$. If we allowed names to refer outside the domain, or to not refer, then we would be on our way to a *free logic*, in which existential generalization requires an additional premise: $\varphi(a)$ and $\exists x\, x = a$, therefore $\exists x\,\varphi(x)$.

5.11 Covered Structures for First-order Languages

Recall that a term is *closed* if it contains no variables.

Definition 5.45 (Value of closed terms). If t is a closed term of the language $\mathcal{L}$ and $\mathfrak{M}$ is a structure for $\mathcal{L}$, the *value* $\text{Val}^{\mathfrak{M}}(t)$ is defined as follows:

1. If t is just the constant symbol c, then $\text{Val}^{\mathfrak{M}}(c) = c^{\mathfrak{M}}$.
2. If t is of the form $f(t_1, \ldots, t_n)$, then

$$\text{Val}^{\mathfrak{M}}(t) = f^{\mathfrak{M}}(\text{Val}^{\mathfrak{M}}(t_1), \ldots, \text{Val}^{\mathfrak{M}}(t_n)).$$

Definition 5.46 (Covered structure). A structure is *covered* if every element of the domain is the value of some closed term.

Example 5.47. Let $\mathcal{L}$ be the language with constant symbols zero, one, two, $\ldots$, the binary predicate symbol $<$, and the binary function symbols $+$ and $\times$. Then a structure $\mathfrak{M}$ for $\mathcal{L}$ is the one with domain $|\mathfrak{M}| = \{0, 1, 2, \ldots\}$ and assignments $\mathit{zero}^{\mathfrak{M}} = 0$, $\mathit{one}^{\mathfrak{M}} = 1$, $\mathit{two}^{\mathfrak{M}} = 2$, and so forth. For the binary relation symbol $<$, the set $<^{\mathfrak{M}}$ is the set of all pairs $\langle c_1, c_2 \rangle \in |\mathfrak{M}|^2$ such that c_1 is less than c_2: for example, $\langle 1, 3 \rangle \in <^{\mathfrak{M}}$ but $\langle 2, 2 \rangle \notin <^{\mathfrak{M}}$. For the binary function symbol $+$, define $+^{\mathfrak{M}}$ in the usual way—for example, $+^{\mathfrak{M}}(2, 3)$ maps to 5, and similarly for the binary function symbol $\times$. Hence, the value of four is just 4, and the value of $\times(\mathit{two}, +(\mathit{three}, \mathit{zero}))$ (or in infix notation, $\mathit{two} \times (\mathit{three} + \mathit{zero})$) is

$$\begin{aligned}
\text{Val}^{\mathfrak{M}}(\times(\mathit{two}, +(\mathit{three}, \mathit{zero})) &= \\
&= \times^{\mathfrak{M}}(\text{Val}^{\mathfrak{M}}(\mathit{two}), \text{Val}^{\mathfrak{M}}(+(\mathit{three}, \mathit{zero}))) \\
&= \times^{\mathfrak{M}}(\text{Val}^{\mathfrak{M}}(\mathit{two}), +^{\mathfrak{M}}(\text{Val}^{\mathfrak{M}}(\mathit{three}), \text{Val}^{\mathfrak{M}}(\mathit{zero}))) \\
&= \times^{\mathfrak{M}}(\mathit{two}^{\mathfrak{M}}, +^{\mathfrak{M}}(\mathit{three}^{\mathfrak{M}}, \mathit{zero}^{\mathfrak{M}})) \\
&= \times^{\mathfrak{M}}(2, +^{\mathfrak{M}}(3, 0)) \\
&= \times^{\mathfrak{M}}(2, 3) \\
&= 6
\end{aligned}$$

5.12 Satisfaction of a Formula in a Structure

The basic notion that relates expressions such as terms and formulas, on the one hand, and structures on the other, are those of *value* of a term and *satisfaction* of a formula. Informally, the value of a term is an element of a structure—if the term is just a constant, its value is the object assigned to the constant by the structure, and if it is built up using function symbols, the value is computed from the values of constants and the functions assigned to the functions

in the term. A formula is *satisfied* in a structure if the interpretation given to the predicates makes the formula true in the domain of the structure. This notion of satisfaction is specified inductively: the specification of the structure directly states when atomic formulas are satisfied, and we define when a complex formula is satisfied depending on the main connective or quantifier and whether or not the immediate subformulas are satisfied.

The case of the quantifiers here is a bit tricky, as the immediate subformula of a quantified formula has a free variable, and structures don't specify the values of variables. In order to deal with this difficulty, we also introduce *variable assignments* and define satisfaction not with respect to a structure alone, but with respect to a structure plus a variable assignment.

Definition 5.48 (Variable Assignment). A *variable assignment* s for a structure $\mathfrak{M}$ is a function which maps each variable to an element of $|\mathfrak{M}|$, i.e., $s\colon \mathrm{Var} \to |\mathfrak{M}|$.

A structure assigns a value to each constant symbol, and a variable assignment to each variable. But we want to use terms built up from them to also name elements of the domain. For this we define the value of terms inductively. For constant symbols and variables the value is just as the structure or the variable assignment specifies it; for more complex terms it is computed recursively using the functions the structure assigns to the function symbols.

Definition 5.49 (Value of Terms). If t is a term of the language $\mathcal{L}$, $\mathfrak{M}$ is a structure for $\mathcal{L}$, and s is a variable assignment for $\mathfrak{M}$, the *value* $\mathrm{Val}^{\mathfrak{M}}_s(t)$ is defined as follows:

1. $t \equiv c$: $\mathrm{Val}^{\mathfrak{M}}_s(t) = c^{\mathfrak{M}}$.
2. $t \equiv x$: $\mathrm{Val}^{\mathfrak{M}}_s(t) = s(x)$.
3. $t \equiv f(t_1, \ldots, t_n)$:
$$\mathrm{Val}^{\mathfrak{M}}_s(t) = f^{\mathfrak{M}}(\mathrm{Val}^{\mathfrak{M}}_s(t_1), \ldots, \mathrm{Val}^{\mathfrak{M}}_s(t_n)).$$

Definition 5.50 (x-Variant). If s is a variable assignment for a structure $\mathfrak{M}$, then any variable assignment s' for $\mathfrak{M}$ which differs from s at most in what it assigns to x is called an *x-variant* of s. If s' is an x-variant of s we write $s' \sim_x s$.

Note that an x-variant of an assignment s does not *have* to assign something different to x. In fact, every assignment counts as an x-variant of itself.

Definition 5.51. If s is a variable assignment for a structure $\mathfrak{M}$ and $m \in |\mathfrak{M}|$, then the assignment $s[m/x]$ is the variable assignment defined by

$$s[m/x](y) = \begin{cases} m & \text{if } y \equiv x \\ s(y) & \text{otherwise.} \end{cases}$$

In other words, $s[m/x]$ is the particular x-variant of s which assigns the domain element m to x, and assigns the same things to variables other than x that s does.

Definition 5.52 (Satisfaction). Satisfaction of a formula φ in a structure $\mathfrak{M}$ relative to a variable assignment s, in symbols: $\mathfrak{M}, s \vDash \varphi$, is defined recursively as follows. (We write $\mathfrak{M}, s \nvDash \varphi$ to mean "not $\mathfrak{M}, s \vDash \varphi$.")

1. $\varphi \equiv \bot$: $\mathfrak{M}, s \nvDash \varphi$.
2. $\varphi \equiv R(t_1, \ldots, t_n)$: $\mathfrak{M}, s \vDash \varphi$ iff $\langle \mathrm{Val}^{\mathfrak{M}}_s(t_1), \ldots, \mathrm{Val}^{\mathfrak{M}}_s(t_n)\rangle \in R^{\mathfrak{M}}$.
3. $\varphi \equiv t_1 = t_2$: $\mathfrak{M}, s \vDash \varphi$ iff $\mathrm{Val}^{\mathfrak{M}}_s(t_1) = \mathrm{Val}^{\mathfrak{M}}_s(t_2)$.
4. $\varphi \equiv \neg\psi$: $\mathfrak{M}, s \vDash \varphi$ iff $\mathfrak{M}, s \nvDash \psi$.
5. $\varphi \equiv (\psi \wedge \chi)$: $\mathfrak{M}, s \vDash \varphi$ iff $\mathfrak{M}, s \vDash \psi$ and $\mathfrak{M}, s \vDash \chi$.
6. $\varphi \equiv (\psi \vee \chi)$: $\mathfrak{M}, s \vDash \varphi$ iff $\mathfrak{M}, s \vDash \psi$ or $\mathfrak{M}, s \vDash \chi$ (or both).
7. $\varphi \equiv (\psi \to \chi)$: $\mathfrak{M}, s \vDash \varphi$ iff $\mathfrak{M}, s \nvDash \psi$ or $\mathfrak{M}, s \vDash \chi$ (or both).
8. $\varphi \equiv \forall x\, \psi$: $\mathfrak{M}, s \vDash \varphi$ iff for every element $m \in |\mathfrak{M}|$, $\mathfrak{M}, s[m/x] \vDash \psi$.
9. $\varphi \equiv \exists x\, \psi$: $\mathfrak{M}, s \vDash \varphi$ iff for at least one element $m \in |\mathfrak{M}|$, $\mathfrak{M}, s[m/x] \vDash \psi$.

The variable assignments are important in the last two clauses. We cannot define satisfaction of $\forall x\, \psi(x)$ by "for all $m \in |\mathfrak{M}|$, $\mathfrak{M} \vDash \psi(m)$." We cannot define satisfaction of $\exists x\, \psi(x)$ by "for at least one $m \in |\mathfrak{M}|$, $\mathfrak{M} \vDash \psi(m)$." The reason is that if $m \in |\mathfrak{M}|$, it is not a symbol of the language, and so $\psi(m)$ is not a formula (that is, $\psi[m/x]$ is undefined). We also cannot assume that we have constant symbols or terms available that name every element of $\mathfrak{M}$, since there is nothing in the definition of structures that requires it. In the standard language, the set of constant symbols is denumerable, so if $|\mathfrak{M}|$ is not enumerable there aren't even enough constant symbols to name every object.

We solve this problem by introducing variable assignments, which allow us to link variables directly with elements of the domain. Then instead of saying that, e.g., $\exists x\, \psi(x)$ is satisfied in $\mathfrak{M}$ iff for at least one $m \in |\mathfrak{M}|$, we say it is satisfied in $\mathfrak{M}$ *relative to* s iff $\psi(x)$ is satisfied relative to $s[m/x]$ for at least one $m \in |\mathfrak{M}|$.

Example 5.53. Let $\mathcal{L} = \{a, b, f, R\}$ where a and b are constant symbols, f is a two-place function symbol, and R is a two-place predicate symbol. Consider the structure $\mathfrak{M}$ defined by:

1. $|\mathfrak{M}| = \{1, 2, 3, 4\}$
2. $a^{\mathfrak{M}} = 1$

3. $b^{\mathfrak{M}} = 2$
4. $f^{\mathfrak{M}}(x, y) = x + y$ if $x + y \leq 3$ and $= 3$ otherwise.
5. $R^{\mathfrak{M}} = \{\langle 1, 1\rangle, \langle 1, 2\rangle, \langle 2, 3\rangle, \langle 2, 4\rangle\}$

The function $s(x) = 1$ that assigns $1 \in |\mathfrak{M}|$ to every variable is a variable assignment for $\mathfrak{M}$.

Then

$$\mathrm{Val}_s^{\mathfrak{M}}(f(a, b)) = f^{\mathfrak{M}}(\mathrm{Val}_s^{\mathfrak{M}}(a), \mathrm{Val}_s^{\mathfrak{M}}(b)).$$

Since a and b are constant symbols, $\mathrm{Val}_s^{\mathfrak{M}}(a) = a^{\mathfrak{M}} = 1$ and $\mathrm{Val}_s^{\mathfrak{M}}(b) = b^{\mathfrak{M}} = 2$. So

$$\mathrm{Val}_s^{\mathfrak{M}}(f(a, b)) = f^{\mathfrak{M}}(1, 2) = 1 + 2 = 3.$$

To compute the value of $f(f(a, b), a)$ we have to consider

$$\mathrm{Val}_s^{\mathfrak{M}}(f(f(a, b), a)) = f^{\mathfrak{M}}(\mathrm{Val}_s^{\mathfrak{M}}(f(a, b)), \mathrm{Val}_s^{\mathfrak{M}}(a)) = f^{\mathfrak{M}}(3, 1) = 3,$$

since $3 + 1 > 3$. Since $s(x) = 1$ and $\mathrm{Val}_s^{\mathfrak{M}}(x) = s(x)$, we also have

$$\mathrm{Val}_s^{\mathfrak{M}}(f(f(a, b), x)) = f^{\mathfrak{M}}(\mathrm{Val}_s^{\mathfrak{M}}(f(a, b)), \mathrm{Val}_s^{\mathfrak{M}}(x)) = f^{\mathfrak{M}}(3, 1) = 3,$$

An atomic formula $R(t_1, t_2)$ is satisfied if the tuple of values of its arguments, i.e., $\langle \mathrm{Val}_s^{\mathfrak{M}}(t_1), \mathrm{Val}_s^{\mathfrak{M}}(t_2)\rangle$, is an element of $R^{\mathfrak{M}}$. So, e.g., we have $\mathfrak{M}, s \vDash R(b, f(a, b))$ since $\langle \mathrm{Val}^{\mathfrak{M}}(b), \mathrm{Val}^{\mathfrak{M}}(f(a, b))\rangle = \langle 2, 3\rangle \in R^{\mathfrak{M}}$, but $\mathfrak{M}, s \nvDash R(x, f(a, b))$ since $\langle 1, 3\rangle \notin R^{\mathfrak{M}}[s]$.

To determine if a non-atomic formula φ is satisfied, you apply the clauses in the inductive definition that applies to the main connective. For instance, the main connective in $R(a, a) \to (R(b, x) \vee R(x, b))$ is the $\to$, and

$$\begin{aligned}&\mathfrak{M}, s \vDash R(a, a) \to (R(b, x) \vee R(x, b)) \text{ iff}\\ &\quad \mathfrak{M}, s \nvDash R(a, a) \text{ or } \mathfrak{M}, s \vDash R(b, x) \vee R(x, b)\end{aligned}$$

Since $\mathfrak{M}, s \vDash R(a, a)$ (because $\langle 1, 1\rangle \in R^{\mathfrak{M}}$) we can't yet determine the answer and must first figure out if $\mathfrak{M}, s \vDash R(b, x) \vee R(x, b)$:

$$\begin{aligned}&\mathfrak{M}, s \vDash R(b, x) \vee R(x, b) \text{ iff}\\ &\quad \mathfrak{M}, s \vDash R(b, x) \text{ or } \mathfrak{M}, s \vDash R(x, b)\end{aligned}$$

And this is the case, since $\mathfrak{M}, s \vDash R(x, b)$ (because $\langle 1, 2\rangle \in R^{\mathfrak{M}}$).

Recall that an x-variant of s is a variable assignment that differs from s at most in what it assigns to x. For every element of $|\mathfrak{M}|$, there is an x-variant of s:

$$s_1 = s[1/x], \qquad s_2 = s[2/x],$$
$$s_3 = s[3/x], \qquad s_4 = s[4/x].$$

So, e.g., $s_2(x) = 2$ and $s_2(y) = s(y) = 1$ for all variables y other than x. These are all the x-variants of s for the structure $\mathfrak{M}$, since $|\mathfrak{M}| = \{1, 2, 3, 4\}$. Note, in particular, that $s_1 = s$ (s is always an x-variant of itself).

To determine if an existentially quantified formula $\exists x\, \varphi(x)$ is satisfied, we have to determine if $\mathfrak{M}, s[m/x] \vDash \varphi(x)$ for at least one $m \in |\mathfrak{M}|$. So,

$$\mathfrak{M}, s \vDash \exists x\, (R(b, x) \vee R(x, b)),$$

since $\mathfrak{M}, s[1/x] \vDash R(b, x) \vee R(x, b)$ ($s[3/x]$ would also fit the bill). But,

$$\mathfrak{M}, s \nvDash \exists x\, (R(b, x) \wedge R(x, b))$$

since, whichever $m \in |\mathfrak{M}|$ we pick, $\mathfrak{M}, s[m/x] \nvDash R(b, x) \wedge R(x, b)$.

To determine if a universally quantified formula $\forall x\, \varphi(x)$ is satisfied, we have to determine if $\mathfrak{M}, s[m/x] \vDash \varphi(x)$ for all $m \in |\mathfrak{M}|$. So,

$$\mathfrak{M}, s \vDash \forall x\, (R(x, a) \to R(a, x)),$$

since $\mathfrak{M}, s[m/x] \vDash R(x, a) \to R(a, x)$ for all $m \in |\mathfrak{M}|$. For $m = 1$, we have $\mathfrak{M}, s[1/x] \vDash R(a, x)$ so the consequent is true; for $m = 2$, 3, and 4, we have $\mathfrak{M}, s[m/x] \nvDash R(x, a)$, so the antecedent is false. But,

$$\mathfrak{M}, s \nvDash \forall x\, (R(a, x) \to R(x, a))$$

since $\mathfrak{M}, s[2/x] \nvDash R(a, x) \to R(x, a)$ (because $\mathfrak{M}, s[2/x] \vDash R(a, x)$ and $\mathfrak{M}, s[2/x] \nvDash R(x, a)$).

For a more complicated case, consider

$$\forall x\, (R(a, x) \to \exists y\, R(x, y)).$$

Since $\mathfrak{M}, s[3/x] \nvDash R(a, x)$ and $\mathfrak{M}, s[4/x] \nvDash R(a, x)$, the interesting cases where we have to worry about the consequent of the conditional are only $m = 1$ and $= 2$. Does $\mathfrak{M}, s[1/x] \vDash \exists y\, R(x, y)$ hold? It does if there is at least one $n \in |\mathfrak{M}|$ so that $\mathfrak{M}, s[1/x][n/y] \vDash R(x, y)$. In fact, if we take $n = 1$, we have $s[1/x][n/y] = s[1/y] = s$. Since $s(x) = 1$, $s(y) = 1$, and $\langle 1, 1\rangle \in R^{\mathfrak{M}}$, the answer is yes.

To determine if $\mathfrak{M}, s[2/x] \vDash \exists y\, R(x, y)$, we have to look at the variable assignments $s[2/x][n/y]$. Here, for $n = 1$, this assignment is $s_2 = s[2/x]$, which does not satisfy $R(x, y)$ ($s_2(x) = 2$, $s_2(y) = 1$, and $\langle 2, 1\rangle \notin R^{\mathfrak{M}}$). However,

consider $s[2/x][3/y] = s_2[3/y]$. $\mathfrak{M}, s_2[3/y] \vDash R(x,y)$ since $\langle 2,3\rangle \in R^{\mathfrak{M}}$, and so $\mathfrak{M}, s_2 \vDash \exists y\, R(x,y)$.

So, for all $n \in |\mathfrak{M}|$, either $\mathfrak{M}, s[m/x] \nvDash R(a,x)$ (if $m = 3, 4$) or $\mathfrak{M}, s[m/x] \vDash \exists y\, R(x,y)$ (if $m = 1, 2$), and so

$$\mathfrak{M}, s \vDash \forall x\, (R(a,x) \to \exists y\, R(x,y)).$$

On the other hand,

$$\mathfrak{M}, s \nvDash \exists x\, (R(a,x) \wedge \forall y\, R(x,y)).$$

We have $\mathfrak{M}, s[m/x] \vDash R(a,x)$ only for $m = 1$ and $m = 2$. But for both of these values of m, there is in turn an $n \in |\mathfrak{M}|$, namely $n = 4$, so that $\mathfrak{M}, s[m/x][n/y] \nvDash R(x,y)$ and so $\mathfrak{M}, s[m/x] \nvDash \forall y\, R(x,y)$ for $m = 1$ and $m = 2$. In sum, there is no $m \in |\mathfrak{M}|$ such that $\mathfrak{M}, s[m/x] \vDash R(a,x) \wedge \forall y\, R(x,y)$.

5.13 Variable Assignments

A variable assignment s provides a value for *every* variable—and there are infinitely many of them. This is of course not necessary. We require variable assignments to assign values to all variables simply because it makes things a lot easier. The value of a term t, and whether or not a formula φ is satisfied in a structure with respect to s, only depend on the assignments s makes to the variables in t and the free variables of φ. This is the content of the next two propositions. To make the idea of "depends on" precise, we show that any two variable assignments that agree on all the variables in t give the same value, and that φ is satisfied relative to one iff it is satisfied relative to the other if two variable assignments agree on all free variables of φ.

Proposition 5.54. *If the variables in a term t are among $x_1, \ldots, x_n$, and $s_1(x_i) = s_2(x_i)$ for $i = 1, \ldots, n$, then $\mathrm{Val}^{\mathfrak{M}}_{s_1}(t) = \mathrm{Val}^{\mathfrak{M}}_{s_2}(t)$.*

Proof. By induction on the complexity of t. For the base case, t can be a constant symbol or one of the variables $x_1, \ldots, x_n$. If $t = c$, then $\mathrm{Val}^{\mathfrak{M}}_{s_1}(t) = c^{\mathfrak{M}} = \mathrm{Val}^{\mathfrak{M}}_{s_2}(t)$. If $t = x_i$, $s_1(x_i) = s_2(x_i)$ by the hypothesis of the proposition, and so $\mathrm{Val}^{\mathfrak{M}}_{s_1}(t) = s_1(x_i) = s_2(x_i) = \mathrm{Val}^{\mathfrak{M}}_{s_2}(t)$.

For the inductive step, assume that $t = f(t_1, \ldots, t_k)$ and that the claim holds for $t_1, \ldots, t_k$. Then

$$\begin{aligned}\mathrm{Val}^{\mathfrak{M}}_{s_1}(t) &= \mathrm{Val}^{\mathfrak{M}}_{s_1}(f(t_1, \ldots, t_k)) = \\ &= f^{\mathfrak{M}}(\mathrm{Val}^{\mathfrak{M}}_{s_1}(t_1), \ldots, \mathrm{Val}^{\mathfrak{M}}_{s_1}(t_k))\end{aligned}$$

For $j = 1, \ldots, k$, the variables of t_j are among $x_1, \ldots, x_n$. By induction hypothesis, $\mathrm{Val}^{\mathfrak{M}}_{s_1}(t_j) = \mathrm{Val}^{\mathfrak{M}}_{s_2}(t_j)$. So,

$$\begin{aligned}\mathrm{Val}^{\mathfrak{M}}_{s_1}(t) &= \mathrm{Val}^{\mathfrak{M}}_{s_1}(f(t_1, \ldots, t_k)) = \\ &= f^{\mathfrak{M}}(\mathrm{Val}^{\mathfrak{M}}_{s_1}(t_1), \ldots, \mathrm{Val}^{\mathfrak{M}}_{s_1}(t_k)) = \\ &= f^{\mathfrak{M}}(\mathrm{Val}^{\mathfrak{M}}_{s_2}(t_1), \ldots, \mathrm{Val}^{\mathfrak{M}}_{s_2}(t_k)) = \\ &= \mathrm{Val}^{\mathfrak{M}}_{s_2}(f(t_1, \ldots, t_k)) = \mathrm{Val}^{\mathfrak{M}}_{s_2}(t).\end{aligned}$$ □

Proposition 5.55. *If the free variables in φ are among $x_1, \ldots, x_n$, and $s_1(x_i) = s_2(x_i)$ for $i = 1, \ldots, n$, then $\mathfrak{M}, s_1 \vDash \varphi$ iff $\mathfrak{M}, s_2 \vDash \varphi$.*

Proof. We use induction on the complexity of φ. For the base case, where φ is atomic, φ can be: $\bot$, $R(t_1, \ldots, t_k)$ for a k-place predicate R and terms $t_1, \ldots, t_k$, or $t_1 = t_2$ for terms t_1 and t_2.

1. $\varphi \equiv \bot$: both $\mathfrak{M}, s_1 \nvDash \varphi$ and $\mathfrak{M}, s_2 \nvDash \varphi$.
2. $\varphi \equiv R(t_1, \ldots, t_k)$: let $\mathfrak{M}, s_1 \vDash \varphi$. Then

   $$\langle \mathrm{Val}^{\mathfrak{M}}_{s_1}(t_1), \ldots, \mathrm{Val}^{\mathfrak{M}}_{s_1}(t_k) \rangle \in R^{\mathfrak{M}}.$$

   For $i = 1, \ldots, k$, $\mathrm{Val}^{\mathfrak{M}}_{s_1}(t_i) = \mathrm{Val}^{\mathfrak{M}}_{s_2}(t_i)$ by Proposition 5.54. So we also have $\langle \mathrm{Val}^{\mathfrak{M}}_{s_2}(t_i), \ldots, \mathrm{Val}^{\mathfrak{M}}_{s_2}(t_k) \rangle \in R^{\mathfrak{M}}$.
3. $\varphi \equiv t_1 = t_2$: suppose $\mathfrak{M}, s_1 \vDash \varphi$. Then $\mathrm{Val}^{\mathfrak{M}}_{s_1}(t_1) = \mathrm{Val}^{\mathfrak{M}}_{s_1}(t_2)$. So,

   $$\begin{aligned}\mathrm{Val}^{\mathfrak{M}}_{s_2}(t_1) &= \mathrm{Val}^{\mathfrak{M}}_{s_1}(t_1) && \text{(by Proposition 5.54)} \\ &= \mathrm{Val}^{\mathfrak{M}}_{s_1}(t_2) && \text{(since } \mathfrak{M}, s_1 \vDash t_1 = t_2\text{)} \\ &= \mathrm{Val}^{\mathfrak{M}}_{s_2}(t_2) && \text{(by Proposition 5.54),}\end{aligned}$$

   so $\mathfrak{M}, s_2 \vDash t_1 = t_2$.

Now assume $\mathfrak{M}, s_1 \vDash \psi$ iff $\mathfrak{M}, s_2 \vDash \psi$ for all formulas ψ less complex than φ. The induction step proceeds by cases determined by the main operator of φ. In each case, we only demonstrate the forward direction of the biconditional; the proof of the reverse direction is symmetrical. In all cases except those for the quantifiers, we apply the induction hypothesis to sub-formulas ψ of φ. The free variables of ψ are among those of φ. Thus, if s_1 and s_2 agree on the free variables of φ, they also agree on those of ψ, and the induction hypothesis applies to ψ.

1. $\varphi \equiv \neg\psi$: if $\mathfrak{M}, s_1 \vDash \varphi$, then $\mathfrak{M}, s_1 \nvDash \psi$, so by the induction hypothesis, $\mathfrak{M}, s_2 \nvDash \psi$, hence $\mathfrak{M}, s_2 \vDash \varphi$.

2. $\varphi \equiv \psi \wedge \chi$: if $\mathfrak{M}, s_1 \vDash \varphi$, then $\mathfrak{M}, s_1 \vDash \psi$ and $\mathfrak{M}, s_1 \vDash \chi$, so by induction hypothesis, $\mathfrak{M}, s_2 \vDash \psi$ and $\mathfrak{M}, s_2 \vDash \chi$. Hence, $\mathfrak{M}, s_2 \vDash \varphi$.

3. $\varphi \equiv \psi \vee \chi$: if $\mathfrak{M}, s_1 \vDash \varphi$, then $\mathfrak{M}, s_1 \vDash \psi$ or $\mathfrak{M}, s_1 \vDash \chi$. By induction hypothesis, $\mathfrak{M}, s_2 \vDash \psi$ or $\mathfrak{M}, s_2 \vDash \chi$, so $\mathfrak{M}, s_2 \vDash \varphi$.

4. $\varphi \equiv \psi \to \chi$: exercise.

5. $\varphi \equiv \exists x\, \psi$: if $\mathfrak{M}, s_1 \vDash \varphi$, there is an $m \in |\mathfrak{M}|$ so that $\mathfrak{M}, s_1[m/x] \vDash \psi$. Let $s_1' = s_1[m/x]$ and $s_2' = s_2[m/x]$. The free variables of ψ are among x_1, $\ldots$, x_n, and x. $s_1'(x_i) = s_2'(x_i)$, since s_1' and s_2' are x-variants of s_1 and s_2, respectively, and by hypothesis $s_1(x_i) = s_2(x_i)$. $s_1'(x) = s_2'(x) = m$ by the way we have defined s_1' and s_2'. Then the induction hypothesis applies to ψ and s_1', s_2', so $\mathfrak{M}, s_2' \vDash \psi$. Hence, since $s_2' = s_2[m/x]$, there is an $m \in |\mathfrak{M}|$ such that $\mathfrak{M}, s_2[m/x] \vDash \psi$, and so $\mathfrak{M}, s_2 \vDash \varphi$.

6. $\varphi \equiv \forall x\, \psi$: exercise.

By induction, we get that $\mathfrak{M}, s_1 \vDash \varphi$ iff $\mathfrak{M}, s_2 \vDash \varphi$ whenever the free variables in φ are among $x_1, \ldots, x_n$ and $s_1(x_i) = s_2(x_i)$ for $i = 1, \ldots, n$. □

Sentences have no free variables, so any two variable assignments assign the same things to all the (zero) free variables of any sentence. The proposition just proved then means that whether or not a sentence is satisfied in a structure relative to a variable assignment is completely independent of the assignment. We'll record this fact. It justifies the definition of satisfaction of a sentence in a structure (without mentioning a variable assignment) that follows.

Corollary 5.56. *If φ is a sentence and s a variable assignment, then $\mathfrak{M}, s \vDash \varphi$ iff $\mathfrak{M}, s' \vDash \varphi$ for every variable assignment s'.*

Proof. Let s' be any variable assignment. Since φ is a sentence, it has no free variables, and so every variable assignment s' trivially assigns the same things to all free variables of φ as does s. So the condition of Proposition 5.55 is satisfied, and we have $\mathfrak{M}, s \vDash \varphi$ iff $\mathfrak{M}, s' \vDash \varphi$. □

Definition 5.57. If φ is a sentence, we say that a structure $\mathfrak{M}$ *satisfies* φ, $\mathfrak{M} \vDash \varphi$, iff $\mathfrak{M}, s \vDash \varphi$ for all variable assignments s.

If $\mathfrak{M} \vDash \varphi$, we also simply say that φ *is true in* $\mathfrak{M}$.

Proposition 5.58. *Let $\mathfrak{M}$ be a structure, φ be a sentence, and s a variable assignment. $\mathfrak{M} \vDash \varphi$ iff $\mathfrak{M}, s \vDash \varphi$.*

Proof. Exercise. □

Proposition 5.59. *Suppose $\varphi(x)$ only contains x free, and $\mathfrak{M}$ is a structure. Then:*

1. $\mathfrak{M} \vDash \exists x\, \varphi(x)$ *iff* $\mathfrak{M}, s \vDash \varphi(x)$ *for at least one variable assignment* s.

2. $\mathfrak{M} \vDash \forall x\, \varphi(x)$ *iff* $\mathfrak{M}, s \vDash \varphi(x)$ *for all variable assignments* s.

Proof. Exercise. □

5.14 Extensionality

Extensionality, sometimes called relevance, can be expressed informally as follows: the only factors that bear upon the satisfaction of formula φ in a structure $\mathfrak{M}$ relative to a variable assignment s, are the size of the domain and the assignments made by $\mathfrak{M}$ and s to the elements of the language that actually appear in φ.

One immediate consequence of extensionality is that where two structures $\mathfrak{M}$ and $\mathfrak{M}'$ agree on all the elements of the language appearing in a sentence φ and have the same domain, $\mathfrak{M}$ and $\mathfrak{M}'$ must also agree on whether or not φ itself is true.

Proposition 5.60 (Extensionality). *Let φ be a formula, and $\mathfrak{M}_1$ and $\mathfrak{M}_2$ be structures with $|\mathfrak{M}_1| = |\mathfrak{M}_2|$, and s a variable assignment on $|\mathfrak{M}_1| = |\mathfrak{M}_2|$. If $c^{\mathfrak{M}_1} = c^{\mathfrak{M}_2}$, $R^{\mathfrak{M}_1} = R^{\mathfrak{M}_2}$, and $f^{\mathfrak{M}_1} = f^{\mathfrak{M}_2}$ for every constant symbol c, relation symbol R, and function symbol f occurring in φ, then $\mathfrak{M}_1, s \vDash \varphi$ iff $\mathfrak{M}_2, s \vDash \varphi$.*

Proof. First prove (by induction on t) that for every term, $\mathrm{Val}_s^{\mathfrak{M}_1}(t) = \mathrm{Val}_s^{\mathfrak{M}_2}(t)$. Then prove the proposition by induction on φ, making use of the claim just proved for the induction basis (where φ is atomic). □

Corollary 5.61 (Extensionality for Sentences). *Let φ be a sentence and $\mathfrak{M}_1$, $\mathfrak{M}_2$ as in Proposition 5.60. Then $\mathfrak{M}_1 \vDash \varphi$ iff $\mathfrak{M}_2 \vDash \varphi$.*

Proof. Follows from Proposition 5.60 by Corollary 5.56. □

Moreover, the value of a term, and whether or not a structure satisfies a formula, only depend on the values of its subterms.

Proposition 5.62. *Let $\mathfrak{M}$ be a structure, t and t' terms, and s a variable assignment. Then $\mathrm{Val}_s^{\mathfrak{M}}(t[t'/x]) = \mathrm{Val}^{\mathfrak{M}}_{s[\mathrm{Val}_s^{\mathfrak{M}}(t')/x]}(t)$.*

Proof. By induction on t.

1. If t is a constant, say, $t \equiv c$, then $t[t'/x] = c$, and $\mathrm{Val}_s^{\mathfrak{M}}(c) = c^{\mathfrak{M}} = \mathrm{Val}^{\mathfrak{M}}_{s[\mathrm{Val}_s^{\mathfrak{M}}(t')/x]}(c)$.

2. If t is a variable other than x, say, $t \equiv y$, then $t[t'/x] = y$, and $\mathrm{Val}_s^{\mathfrak{M}}(y) = \mathrm{Val}^{\mathfrak{M}}_{s[\mathrm{Val}_s^{\mathfrak{M}}(t')/x]}(y)$ since $s \sim_x s[\mathrm{Val}_s^{\mathfrak{M}}(t')/x]$.

3. If $t \equiv x$, then $t[t'/x] = t'$. But $\mathrm{Val}^{\mathfrak{M}}_{s[\mathrm{Val}^{\mathfrak{M}}_s(t')/x]}(x) = \mathrm{Val}^{\mathfrak{M}}_s(t')$ by definition of $s[\mathrm{Val}^{\mathfrak{M}}_s(t')/x]$.

4. If $t \equiv f(t_1, \ldots, t_n)$ then we have:

$$\begin{aligned}
&\mathrm{Val}^{\mathfrak{M}}_s(t[t'/x]) = \\
&\quad = \mathrm{Val}^{\mathfrak{M}}_s(f(t_1[t'/x], \ldots, t_n[t'/x])) \\
&\qquad \text{by definition of } t[t'/x] \\
&\quad = f^{\mathfrak{M}}(\mathrm{Val}^{\mathfrak{M}}_s(t_1[t'/x]), \ldots, \mathrm{Val}^{\mathfrak{M}}_s(t_n[t'/x])) \\
&\qquad \text{by definition of } \mathrm{Val}^{\mathfrak{M}}_s(f(\ldots)) \\
&\quad = f^{\mathfrak{M}}(\mathrm{Val}^{\mathfrak{M}}_{s[\mathrm{Val}^{\mathfrak{M}}_s(t')/x]}(t_1), \ldots, \mathrm{Val}^{\mathfrak{M}}_{s[\mathrm{Val}^{\mathfrak{M}}_s(t')/x]}(t_n)) \\
&\qquad \text{by induction hypothesis} \\
&\quad = \mathrm{Val}^{\mathfrak{M}}_{s[\mathrm{Val}^{\mathfrak{M}}_s(t')/x]}(t) \text{ by definition of } \mathrm{Val}^{\mathfrak{M}}_{s[\mathrm{Val}^{\mathfrak{M}}_s(t')/x]}(f(\ldots))
\end{aligned}$$ □

Proposition 5.63. *Let $\mathfrak{M}$ be a structure, φ a formula, t' a term, and s a variable assignment. Then $\mathfrak{M}, s \vDash \varphi[t'/x]$ iff $\mathfrak{M}, s[\mathrm{Val}^{\mathfrak{M}}_s(t')/x] \vDash \varphi$.*

Proof. Exercise. □

The point of Propositions 5.62 and 5.63 is the following. Suppose we have a term t or a formula φ and some term t', and we want to know the value of $t[t'/x]$ or whether or not $\varphi[t'/x]$ is satisfied in a structure $\mathfrak{M}$ relative to a variable assignment s. Then we can either perform the substitution first and then consider the value or satisfaction relative to $\mathfrak{M}$ and s, or we can first determine the value $m = \mathrm{Val}^{\mathfrak{M}}_s(t')$ of t' in $\mathfrak{M}$ relative to s, change the variable assignment to $s[m/x]$ and then consider the value of t in $\mathfrak{M}$ and $s[m/x]$, or whether $\mathfrak{M}, s[m/x] \vDash \varphi$. Propositions 5.62 and 5.63 guarantee that the answer will be the same, whichever way we do it.

5.15 Semantic Notions

Given the definition of structures for first-order languages, we can define some basic semantic properties of and relationships between sentences. The simplest of these is the notion of *validity* of a sentence. A sentence is valid if it is satisfied in every structure. Valid sentences are those that are satisfied regardless of how the non-logical symbols in it are interpreted. Valid sentences are therefore also called *logical truths*—they are true, i.e., satisfied, in any structure and hence their truth depends only on the logical symbols occurring in them and their syntactic structure, but not on the non-logical symbols or their interpretation.

Definition 5.64 (Validity). A sentence φ is *valid*, $\vDash \varphi$, iff $\mathfrak{M} \vDash \varphi$ for every structure $\mathfrak{M}$.

Definition 5.65 (Entailment). A set of sentences Γ *entails* a sentence φ, $\Gamma \vDash \varphi$, iff for every structure $\mathfrak{M}$ with $\mathfrak{M} \vDash \Gamma$, $\mathfrak{M} \vDash \varphi$.

Definition 5.66 (Satisfiability). A set of sentences Γ is *satisfiable* if $\mathfrak{M} \vDash \Gamma$ for some structure $\mathfrak{M}$. If Γ is not satisfiable it is called *unsatisfiable*.

Proposition 5.67. *A sentence φ is valid iff $\Gamma \vDash \varphi$ for every set of sentences Γ.*

Proof. For the forward direction, let φ be valid, and let Γ be a set of sentences. Let $\mathfrak{M}$ be a structure so that $\mathfrak{M} \vDash \Gamma$. Since φ is valid, $\mathfrak{M} \vDash \varphi$, hence $\Gamma \vDash \varphi$.

For the contrapositive of the reverse direction, let φ be invalid, so there is a structure $\mathfrak{M}$ with $\mathfrak{M} \nvDash \varphi$. When $\Gamma = \{\top\}$, since $\top$ is valid, $\mathfrak{M} \vDash \Gamma$. Hence, there is a structure $\mathfrak{M}$ so that $\mathfrak{M} \vDash \Gamma$ but $\mathfrak{M} \nvDash \varphi$, hence Γ does not entail φ. □

Proposition 5.68. *$\Gamma \vDash \varphi$ iff $\Gamma \cup \{\neg\varphi\}$ is unsatisfiable.*

Proof. For the forward direction, suppose $\Gamma \vDash \varphi$ and suppose to the contrary that there is a structure $\mathfrak{M}$ so that $\mathfrak{M} \vDash \Gamma \cup \{\neg\varphi\}$. Since $\mathfrak{M} \vDash \Gamma$ and $\Gamma \vDash \varphi$, $\mathfrak{M} \vDash \varphi$. Also, since $\mathfrak{M} \vDash \Gamma \cup \{\neg\varphi\}$, $\mathfrak{M} \vDash \neg\varphi$, so we have both $\mathfrak{M} \vDash \varphi$ and $\mathfrak{M} \nvDash \varphi$, a contradiction. Hence, there can be no such structure $\mathfrak{M}$, so $\Gamma \cup \{\neg\varphi\}$ is unsatisfiable.

For the reverse direction, suppose $\Gamma \cup \{\neg\varphi\}$ is unsatisfiable. So for every structure $\mathfrak{M}$, either $\mathfrak{M} \nvDash \Gamma$ or $\mathfrak{M} \vDash \varphi$. Hence, for every structure $\mathfrak{M}$ with $\mathfrak{M} \vDash \Gamma$, $\mathfrak{M} \vDash \varphi$, so $\Gamma \vDash \varphi$. □

Proposition 5.69. *If $\Gamma \subseteq \Gamma'$ and $\Gamma \vDash \varphi$, then $\Gamma' \vDash \varphi$.*

Proof. Suppose that $\Gamma \subseteq \Gamma'$ and $\Gamma \vDash \varphi$. Let $\mathfrak{M}$ be a structure such that $\mathfrak{M} \vDash \Gamma'$; then $\mathfrak{M} \vDash \Gamma$, and since $\Gamma \vDash \varphi$, we get that $\mathfrak{M} \vDash \varphi$. Hence, whenever $\mathfrak{M} \vDash \Gamma'$, $\mathfrak{M} \vDash \varphi$, so $\Gamma' \vDash \varphi$. □

Theorem 5.70 (Semantic Deduction Theorem). *$\Gamma \cup \{\varphi\} \vDash \psi$ iff $\Gamma \vDash \varphi \to \psi$.*

Proof. For the forward direction, let $\Gamma \cup \{\varphi\} \vDash \psi$ and let $\mathfrak{M}$ be a structure so that $\mathfrak{M} \vDash \Gamma$. If $\mathfrak{M} \vDash \varphi$, then $\mathfrak{M} \vDash \Gamma \cup \{\varphi\}$, so since $\Gamma \cup \{\varphi\}$ entails ψ, we get $\mathfrak{M} \vDash \psi$. Therefore, $\mathfrak{M} \vDash \varphi \to \psi$, so $\Gamma \vDash \varphi \to \psi$.

For the reverse direction, let $\Gamma \vDash \varphi \to \psi$ and $\mathfrak{M}$ be a structure so that $\mathfrak{M} \vDash \Gamma \cup \{\varphi\}$. Then $\mathfrak{M} \vDash \Gamma$, so $\mathfrak{M} \vDash \varphi \to \psi$, and since $\mathfrak{M} \vDash \varphi$, $\mathfrak{M} \vDash \psi$. Hence, whenever $\mathfrak{M} \vDash \Gamma \cup \{\varphi\}$, $\mathfrak{M} \vDash \psi$, so $\Gamma \cup \{\varphi\} \vDash \psi$. □

Proposition 5.71. *Let $\mathfrak{M}$ be a structure, and $\varphi(x)$ a formula with one free variable x, and t a closed term. Then:*

1. $\varphi(t) \vDash \exists x\, \varphi(x)$
2. $\forall x\, \varphi(x) \vDash \varphi(t)$

Proof. 1. Suppose $\mathfrak{M} \vDash \varphi(t)$. Let s be a variable assignment with $s(x) = \mathrm{Val}^{\mathfrak{M}}(t)$. Then $\mathfrak{M}, s \vDash \varphi(t)$ since $\varphi(t)$ is a sentence. By Proposition 5.63, $\mathfrak{M}, s \vDash \varphi(x)$. By Proposition 5.59, $\mathfrak{M} \vDash \exists x\, \varphi(x)$.

2. Exercise. □

Chapter 6

Theories and Their Models

6.1 Introduction

The development of the axiomatic method is a significant achievement in the history of science, and is of special importance in the history of mathematics. An axiomatic development of a field involves the clarification of many questions: What is the field about? What are the most fundamental concepts? How are they related? Can all the concepts of the field be defined in terms of these fundamental concepts? What laws do, and must, these concepts obey?

The axiomatic method and logic were made for each other. Formal logic provides the tools for formulating axiomatic theories, for proving theorems from the axioms of the theory in a precisely specified way, for studying the properties of all systems satisfying the axioms in a systematic way.

Definition 6.1. A set of sentences Γ is *closed* iff, whenever $\Gamma \vDash \varphi$ then $\varphi \in \Gamma$. The *closure* of a set of sentences Γ is $\{\varphi : \Gamma \vDash \varphi\}$.

We say that Γ is *axiomatized by* a set of sentences Δ if Γ is the closure of Δ.

We can think of an axiomatic theory as the set of sentences that is axiomatized by its set of axioms Δ. In other words, when we have a first-order language which contains non-logical symbols for the primitives of the axiomatically developed science we wish to study, together with a set of sentences that express the fundamental laws of the science, we can think of the theory as represented by all the sentences in this language that are entailed by the axioms. This ranges from simple examples with only a single primitive and simple axioms, such as the theory of partial orders, to complex theories such as Newtonian mechanics.

The important logical facts that make this formal approach to the axiomatic method so important are the following. Suppose Γ is an axiom system for a theory, i.e., a set of sentences.

1. We can state precisely when an axiom system captures an intended class of structures. That is, if we are interested in a certain class of structures, we will successfully capture that class by an axiom system Γ iff the structures are exactly those $\mathfrak{M}$ such that $\mathfrak{M} \models \Gamma$.

2. We may fail in this respect because there are $\mathfrak{M}$ such that $\mathfrak{M} \models \Gamma$, but $\mathfrak{M}$ is not one of the structures we intend. This may lead us to add axioms which are not true in $\mathfrak{M}$.

3. If we are successful at least in the respect that Γ is true in all the intended structures, then a sentence φ is true in all intended structures whenever $\Gamma \models \varphi$. Thus we can use logical tools (such as derivation methods) to show that sentences are true in all intended structures simply by showing that they are entailed by the axioms.

4. Sometimes we don't have intended structures in mind, but instead start from the axioms themselves: we begin with some primitives that we want to satisfy certain laws which we codify in an axiom system. One thing that we would like to verify right away is that the axioms do not contradict each other: if they do, there can be no concepts that obey these laws, and we have tried to set up an incoherent theory. We can verify that this doesn't happen by finding a model of Γ. And if there are models of our theory, we can use logical methods to investigate them, and we can also use logical methods to construct models.

5. The independence of the axioms is likewise an important question. It may happen that one of the axioms is actually a consequence of the others, and so is redundant. We can prove that an axiom φ in Γ is redundant by proving $\Gamma \setminus \{\varphi\} \models \varphi$. We can also prove that an axiom is not redundant by showing that $(\Gamma \setminus \{\varphi\}) \cup \{\neg\varphi\}$ is satisfiable. For instance, this is how it was shown that the parallel postulate is independent of the other axioms of geometry.

6. Another important question is that of definability of concepts in a theory: The choice of the language determines what the models of a theory consist of. But not every aspect of a theory must be represented separately in its models. For instance, every ordering $\leq$ determines a corresponding strict ordering $<$—given one, we can define the other. So it is not necessary that a model of a theory involving such an order must *also* contain the corresponding strict ordering. When is it the case, in general, that one relation can be defined in terms of others? When is it impossible to define a relation in terms of others (and hence must add it to the primitives of the language)?

6.2 Expressing Properties of Structures

It is often useful and important to express conditions on functions and relations, or more generally, that the functions and relations in a structure satisfy these conditions. For instance, we would like to have ways of distinguishing those structures for a language which "capture" what we want the predicate symbols to "mean" from those that do not. Of course we're completely free to specify which structures we "intend," e.g., we can specify that the interpretation of the predicate symbol $\leq$ must be an ordering, or that we are only interested in interpretations of $\mathcal{L}$ in which the domain consists of sets and $\in$ is interpreted by the "is an element of" relation. But can we do this with sentences of the language? In other words, which conditions on a structure $\mathfrak{M}$ can we express by a sentence (or perhaps a set of sentences) in the language of $\mathfrak{M}$? There are some conditions that we will not be able to express. For instance, there is no sentence of $\mathcal{L}_A$ which is only true in a structure $\mathfrak{M}$ if $|\mathfrak{M}| = \mathbb{N}$. We cannot express "the domain contains only natural numbers." But there are "structural properties" of structures that we perhaps can express. Which properties of structures can we express by sentences? Or, to put it another way, which collections of structures can we describe as those making a sentence (or set of sentences) true?

Definition 6.2 (Model of a set). Let Γ be a set of sentences in a language $\mathcal{L}$. We say that a structure $\mathfrak{M}$ *is a model of* Γ if $\mathfrak{M} \vDash \varphi$ for all $\varphi \in \Gamma$.

Example 6.3. The sentence $\forall x\, x \leq x$ is true in $\mathfrak{M}$ iff $\leq^{\mathfrak{M}}$ is a reflexive relation. The sentence $\forall x \forall y\, ((x \leq y \wedge y \leq x) \rightarrow x = y)$ is true in $\mathfrak{M}$ iff $\leq^{\mathfrak{M}}$ is anti-symmetric. The sentence $\forall x \forall y \forall z\, ((x \leq y \wedge y \leq z) \rightarrow x \leq z)$ is true in $\mathfrak{M}$ iff $\leq^{\mathfrak{M}}$ is transitive. Thus, the models of

$$\begin{aligned} \{ \quad & \forall x\, x \leq x, \\ & \forall x \forall y\, ((x \leq y \wedge y \leq x) \rightarrow x = y), \\ & \forall x \forall y \forall z\, ((x \leq y \wedge y \leq z) \rightarrow x \leq z) \quad \} \end{aligned}$$

are exactly those structures in which $\leq^{\mathfrak{M}}$ is reflexive, anti-symmetric, and transitive, i.e., a partial order. Hence, we can take them as axioms for the *first-order theory of partial orders*.

6.3 Examples of First-Order Theories

Example 6.4. The theory of strict linear orders in the language $\mathcal{L}_<$ is axiomatized by the set

$$\begin{aligned} \{ \quad & \forall x \neg x < x, \\ & \forall x \forall y\, ((x < y \vee y < x) \vee x = y), \\ & \forall x \forall y \forall z\, ((x < y \wedge y < z) \rightarrow x < z) \quad \} \end{aligned}$$

It completely captures the intended structures: every strict linear order is a model of this axiom system, and vice versa, if R is a linear order on a set X, then the structure $\mathfrak{M}$ with $|\mathfrak{M}| = X$ and $<^{\mathfrak{M}} = R$ is a model of this theory.

Example 6.5. The theory of groups in the language 1 (constant symbol), $\cdot$ (two-place function symbol) is axiomatized by

$$\begin{aligned}
&\forall x\,(x \cdot 1) = x\\
&\forall x\,\forall y\,\forall z\,(x \cdot (y \cdot z)) = ((x \cdot y) \cdot z)\\
&\forall x\,\exists y\,(x \cdot y) = 1
\end{aligned}$$

Example 6.6. The theory of Peano arithmetic is axiomatized by the following sentences in the language of arithmetic $\mathcal{L}_A$.

$$\begin{aligned}
&\forall x\,\forall y\,(x' = y' \to x = y)\\
&\forall x\,0 \neq x'\\
&\forall x\,(x + 0) = x\\
&\forall x\,\forall y\,(x + y') = (x + y)'\\
&\forall x\,(x \times 0) = 0\\
&\forall x\,\forall y\,(x \times y') = ((x \times y) + x)\\
&\forall x\,\forall y\,(x < y \leftrightarrow \exists z\,(z' + x) = y)
\end{aligned}$$

plus all sentences of the form

$$(\varphi(0) \wedge \forall x\,(\varphi(x) \to \varphi(x'))) \to \forall x\,\varphi(x)$$

Since there are infinitely many sentences of the latter form, this axiom system is infinite. The latter form is called the *induction schema*. (Actually, the induction schema is a bit more complicated than we let on here.)

The last axiom is an *explicit definition* of $<$.

Example 6.7. The theory of pure sets plays an important role in the foundations (and in the philosophy) of mathematics. A set is pure if all its elements are also pure sets. The empty set counts therefore as pure, but a set that has something as an element that is not a set would not be pure. So the pure sets are those that are formed just from the empty set and no "urelements," i.e., objects that are not themselves sets.

The following might be considered as an axiom system for a theory of pure sets:

$$\begin{aligned}
&\exists x\,\neg\exists y\,y \in x\\
&\forall x\,\forall y\,(\forall z(z \in x \leftrightarrow z \in y) \to x = y)\\
&\forall x\,\forall y\,\exists z\,\forall u\,(u \in z \leftrightarrow (u = x \vee u = y))\\
&\forall x\,\exists y\,\forall z\,(z \in y \leftrightarrow \exists u\,(z \in u \wedge u \in x))
\end{aligned}$$

plus all sentences of the form

$$\exists x \, \forall y \, (y \in x \leftrightarrow \varphi(y))$$

The first axiom says that there is a set with no elements (i.e., $\emptyset$ exists); the second says that sets are extensional; the third that for any sets X and Y, the set $\{X, Y\}$ exists; the fourth that for any set X, the set $\cup X$ exists, where $\cup X$ is the union of all the elements of X.

The sentences mentioned last are collectively called the *naive comprehension scheme*. It essentially says that for every $\varphi(x)$, the set $\{x : \varphi(x)\}$ exists—so at first glance a true, useful, and perhaps even necessary axiom. It is called "naive" because, as it turns out, it makes this theory unsatisfiable: if you take $\varphi(y)$ to be $\neg y \in y$, you get the sentence

$$\exists x \, \forall y \, (y \in x \leftrightarrow \neg y \in y)$$

and this sentence is not satisfied in any structure.

Example 6.8. In the area of *mereology*, the relation of *parthood* is a fundamental relation. Just like theories of sets, there are theories of parthood that axiomatize various conceptions (sometimes conflicting) of this relation.

The language of mereology contains a single two-place predicate symbol P, and $P(x, y)$ "means" that x is a part of y. When we have this interpretation in mind, a structure for this language is called a *parthood structure*. Of course, not every structure for a single two-place predicate will really deserve this name. To have a chance of capturing "parthood," $P^{\mathfrak{M}}$ must satisfy some conditions, which we can lay down as axioms for a theory of parthood. For instance, parthood is a partial order on objects: every object is a part (albeit an *improper* part) of itself; no two different objects can be parts of each other; a part of a part of an object is itself part of that object. Note that in this sense "is a part of" resembles "is a subset of," but does not resemble "is an element of" which is neither reflexive nor transitive.

$$\begin{aligned}
&\forall x \, P(x, x) \\
&\forall x \, \forall y \, ((P(x, y) \wedge P(y, x)) \rightarrow x = y) \\
&\forall x \, \forall y \, \forall z \, ((P(x, y) \wedge P(y, z)) \rightarrow P(x, z))
\end{aligned}$$

Moreover, any two objects have a mereological sum (an object that has these two objects as parts, and is minimal in this respect).

$$\forall x \, \forall y \, \exists z \, \forall u \, (P(z, u) \leftrightarrow (P(x, u) \wedge P(y, u)))$$

These are only some of the basic principles of parthood considered by metaphysicians. Further principles, however, quickly become hard to formulate or write down without first introducing some defined relations. For instance,

most metaphysicians interested in mereology also view the following as a valid principle: whenever an object x has a proper part y, it also has a part z that has no parts in common with y, and so that the fusion of y and z is x.

6.4 Expressing Relations in a Structure

One main use formulas can be put to is to express properties and relations in a structure $\mathfrak{M}$ in terms of the primitives of the language $\mathcal{L}$ of $\mathfrak{M}$. By this we mean the following: the domain of $\mathfrak{M}$ is a set of objects. The constant symbols, function symbols, and predicate symbols are interpreted in $\mathfrak{M}$ by some objects in$|\mathfrak{M}|$, functions on $|\mathfrak{M}|$, and relations on $|\mathfrak{M}|$. For instance, if A^2_0 is in $\mathcal{L}$, then $\mathfrak{M}$ assigns to it a relation $R = A^{2\mathfrak{M}}_0$. Then the formula $A^2_0(v_1, v_2)$ *expresses* that very relation, in the following sense: if a variable assignment s maps v_1 to $a \in |\mathfrak{M}|$ and v_2 to $b \in |\mathfrak{M}|$, then

$$Rab \quad \text{iff} \quad \mathfrak{M}, s \vDash A^2_0(v_1, v_2).$$

Note that we have to involve variable assignments here: we can't just say "Rab iff $\mathfrak{M} \vDash A^2_0(a, b)$" because a and b are not symbols of our language: they are elements of $|\mathfrak{M}|$.

Since we don't just have atomic formulas, but can combine them using the logical connectives and the quantifiers, more complex formulas can define other relations which aren't directly built into $\mathfrak{M}$. We're interested in how to do that, and specifically, which relations we can define in a structure.

Definition 6.9. Let $\varphi(v_1, \ldots, v_n)$ be a formula of $\mathcal{L}$ in which only $v_1, \ldots, v_n$ occur free, and let $\mathfrak{M}$ be a structure for $\mathcal{L}$. $\varphi(v_1, \ldots, v_n)$ *expresses the relation* $R \subseteq |\mathfrak{M}|^n$ iff

$$Ra_1 \ldots a_n \quad \text{iff} \quad \mathfrak{M}, s \vDash \varphi(v_1, \ldots, v_n)$$

for any variable assignment s with $s(v_i) = a_i$ $(i = 1, \ldots, n)$.

Example 6.10. In the standard model of arithmetic $\mathfrak{N}$, the formula $v_1 < v_2 \vee v_1 = v_2$ expresses the $\leq$ relation on $\mathbb{N}$. The formula $v_2 = v_1'$ expresses the successor relation, i.e., the relation $R \subseteq \mathbb{N}^2$ where Rnm holds if m is the successor of n. The formula $v_1 = v_2'$ expresses the predecessor relation. The formulas $\exists v_3\,(v_3 \neq 0 \wedge v_2 = (v_1 + v_3))$ and $\exists v_3\,(v_1 + v_3') = v_2$ both express the $<$ relation. This means that the predicate symbol $<$ is actually superfluous in the language of arithmetic; it can be defined.

This idea is not just interesting in specific structures, but generally whenever we use a language to describe an intended model or models, i.e., when we consider theories. These theories often only contain a few predicate symbols as basic symbols, but in the domain they are used to describe often many

other relations play an important role. If these other relations can be systematically expressed by the relations that interpret the basic predicate symbols of the language, we say we can *define* them in the language.

6.5 The Theory of Sets

Almost all of mathematics can be developed in the theory of sets. Developing mathematics in this theory involves a number of things. First, it requires a set of axioms for the relation $\in$. A number of different axiom systems have been developed, sometimes with conflicting properties of $\in$. The axiom system known as **ZFC**, Zermelo-Fraenkel set theory with the axiom of choice stands out: it is by far the most widely used and studied, because it turns out that its axioms suffice to prove almost all the things mathematicians expect to be able to prove. But before that can be established, it first is necessary to make clear how we can even *express* all the things mathematicians would like to express. For starters, the language contains no constant symbols or function symbols, so it seems at first glance unclear that we can talk about particular sets (such as $\emptyset$ or $\mathbb{N}$), can talk about operations on sets (such as $X \cup Y$ and $\wp(X)$), let alone other constructions which involve things other than sets, such as relations and functions.

To begin with, "is an element of" is not the only relation we are interested in: "is a subset of" seems almost as important. But we can *define* "is a subset of" in terms of "is an element of." To do this, we have to find a formula $\varphi(x, y)$ in the language of set theory which is satisfied by a pair of sets $\langle X, Y \rangle$ iff $X \subseteq Y$. But X is a subset of Y just in case all elements of X are also elements of Y. So we can define $\subseteq$ by the formula

$$\forall z\, (z \in x \to z \in y)$$

Now, whenever we want to use the relation $\subseteq$ in a formula, we could instead use that formula (with x and y suitably replaced, and the bound variable z renamed if necessary). For instance, extensionality of sets means that if any sets x and y are contained in each other, then x and y must be the same set. This can be expressed by $\forall x\, \forall y\, ((x \subseteq y \wedge y \subseteq x) \to x = y)$, or, if we replace $\subseteq$ by the above definition, by

$$\forall x\, \forall y\, ((\forall z\, (z \in x \to z \in y) \wedge \forall z\, (z \in y \to z \in x)) \to x = y).$$

This is in fact one of the axioms of **ZFC**, the "axiom of extensionality."

There is no constant symbol for $\emptyset$, but we can express "x is empty" by $\neg \exists y\, y \in x$. Then "$\emptyset$ exists" becomes the sentence $\exists x\, \neg \exists y\, y \in x$. This is another axiom of **ZFC**. (Note that the axiom of extensionality implies that there is only one empty set.) Whenever we want to talk about $\emptyset$ in the language of set theory, we would write this as "there is a set that's empty and ..." As an

example, to express the fact that $\emptyset$ is a subset of every set, we could write

$$\exists x\,(\neg\exists y\, y \in x \wedge \forall z\, x \subseteq z)$$

where, of course, $x \subseteq z$ would in turn have to be replaced by its definition.

To talk about operations on sets, such as $X \cup Y$ and $\wp(X)$, we have to use a similar trick. There are no function symbols in the language of set theory, but we can express the functional relations $X \cup Y = Z$ and $\wp(X) = Y$ by

$$\forall u\,((u \in x \vee u \in y) \leftrightarrow u \in z)$$
$$\forall u\,(u \subseteq x \leftrightarrow u \in y)$$

since the elements of $X \cup Y$ are exactly the sets that are either elements of X or elements of Y, and the elements of $\wp(X)$ are exactly the subsets of X. However, this doesn't allow us to use $x \cup y$ or $\wp(x)$ as if they were terms: we can only use the entire formulas that define the relations $X \cup Y = Z$ and $\wp(X) = Y$. In fact, we do not know that these relations are ever satisfied, i.e., we do not know that unions and power sets always exist. For instance, the sentence $\forall x\,\exists y\,\wp(x) = y$ is another axiom of **ZFC** (the power set axiom).

Now what about talk of ordered pairs or functions? Here we have to explain how we can think of ordered pairs and functions as special kinds of sets. One way to define the ordered pair $\langle x, y\rangle$ is as the set $\{\{x\}, \{x, y\}\}$. But like before, we cannot introduce a function symbol that names this set; we can only define the relation $\langle x, y\rangle = z$, i.e., $\{\{x\}, \{x, y\}\} = z$:

$$\forall u\,(u \in z \leftrightarrow (\forall v\,(v \in u \leftrightarrow v = x) \vee \forall v\,(v \in u \leftrightarrow (v = x \vee v = y))))$$

This says that the elements u of z are exactly those sets which either have x as its only element or have x and y as its only elements (in other words, those sets that are either identical to $\{x\}$ or identical to $\{x, y\}$). Once we have this, we can say further things, e.g., that $X \times Y = Z$:

$$\forall z\,(z \in Z \leftrightarrow \exists x\,\exists y\,(x \in X \wedge y \in Y \wedge \langle x, y\rangle = z))$$

A function $f\colon X \to Y$ can be thought of as the relation $f(x) = y$, i.e., as the set of pairs $\{\langle x, y\rangle : f(x) = y\}$. We can then say that a set f is a function from X to Y if (a) it is a relation $\subseteq X \times Y$, (b) it is total, i.e., for all $x \in X$ there is some $y \in Y$ such that $\langle x, y\rangle \in f$ and (c) it is functional, i.e., whenever $\langle x, y\rangle, \langle x, y'\rangle \in f$, $y = y'$ (because values of functions must be unique). So "f is a function from X to Y" can be written as:

$$\begin{aligned}&\forall u\,(u \in f \to \exists x\,\exists y\,(x \in X \wedge y \in Y \wedge \langle x, y\rangle = u)) \wedge\\ &\forall x\,(x \in X \to (\exists y\,(y \in Y \wedge \mathrm{maps}(f, x, y)) \wedge\\ &\qquad\qquad (\forall y\,\forall y'\,((\mathrm{maps}(f, x, y) \wedge \mathrm{maps}(f, x, y')) \to y = y')))\end{aligned}$$

where $\text{maps}(f, x, y)$ abbreviates $\exists v\,(v \in f \wedge \langle x, y\rangle = v)$ (this formula expresses "$f(x) = y$").

It is now also not hard to express that $f\colon X \to Y$ is injective, for instance:

$$f\colon X \to Y \wedge \forall x\,\forall x'\,((x \in X \wedge x' \in X \wedge \exists y\,(\text{maps}(f, x, y) \wedge \text{maps}(f, x', y))) \to x = x')$$

A function $f\colon X \to Y$ is injective iff, whenever f maps $x, x' \in X$ to a single y, $x = x'$. If we abbreviate this formula as $\text{inj}(f, X, Y)$, we're already in a position to state in the language of set theory something as non-trivial as Cantor's theorem: there is no injective function from $\wp(X)$ to X:

$$\forall X\,\forall Y\,(\wp(X) = Y \to \neg\exists f\,\text{inj}(f, Y, X))$$

One might think that set theory requires another axiom that guarantees the existence of a set for every defining property. If $\varphi(x)$ is a formula of set theory with the variable x free, we can consider the sentence

$$\exists y\,\forall x\,(x \in y \leftrightarrow \varphi(x)).$$

This sentence states that there is a set y whose elements are all and only those x that satisfy $\varphi(x)$. This schema is called the "comprehension principle." It looks very useful; unfortunately it is inconsistent. Take $\varphi(x) \equiv \neg x \in x$, then the comprehension principle states

$$\exists y\,\forall x\,(x \in y \leftrightarrow x \notin x),$$

i.e., it states the existence of a set of all sets that are not elements of themselves. No such set can exist—this is Russell's Paradox. **ZFC**, in fact, contains a restricted—and consistent—version of this principle, the separation principle:

$$\forall z\,\exists y\,\forall x\,(x \in y \leftrightarrow (x \in z \wedge \varphi(x)).$$

6.6 Expressing the Size of Structures

There are some properties of structures we can express even without using the non-logical symbols of a language. For instance, there are sentences which are true in a structure iff the domain of the structure has at least, at most, or exactly a certain number n of elements.

Proposition 6.11. *The sentence*

$$\begin{aligned}\varphi_{\geq n} \equiv \exists x_1\,\exists x_2\,\ldots\exists x_n\ & (x_1 \neq x_2 \wedge x_1 \neq x_3 \wedge x_1 \neq x_4 \wedge \cdots \wedge x_1 \neq x_n \wedge \\ & \quad x_2 \neq x_3 \wedge x_2 \neq x_4 \wedge \cdots \wedge x_2 \neq x_n \wedge \\ & \quad \vdots \\ & \quad x_{n-1} \neq x_n)\end{aligned}$$

is true in a structure $\mathfrak{M}$ iff $|\mathfrak{M}|$ contains at least n elements. Consequently, $\mathfrak{M} \models \neg\varphi_{\geq n+1}$ iff $|\mathfrak{M}|$ contains at most n elements.

Proposition 6.12. *The sentence*

$$\begin{aligned}\varphi_{=n} \equiv \exists x_1\, \exists x_2\, \ldots \exists x_n \\ (x_1 \neq x_2 \wedge x_1 \neq x_3 \wedge x_1 \neq x_4 \wedge \cdots \wedge x_1 \neq x_n \wedge \\ x_2 \neq x_3 \wedge x_2 \neq x_4 \wedge \cdots \wedge x_2 \neq x_n \wedge \\ \vdots \\ x_{n-1} \neq x_n \wedge \\ \forall y\, (y = x_1 \vee \cdots \vee y = x_n))\end{aligned}$$

is true in a structure $\mathfrak{M}$ iff $|\mathfrak{M}|$ contains exactly n elements.

Proposition 6.13. *A structure is infinite iff it is a model of*

$$\{\varphi_{\geq 1}, \varphi_{\geq 2}, \varphi_{\geq 3}, \ldots\}.$$

There is no single purely logical sentence which is true in $\mathfrak{M}$ iff $|\mathfrak{M}|$ is infinite. However, one can give sentences with non-logical predicate symbols which only have infinite models (although not every infinite structure is a model of them). The property of being a finite structure, and the property of being a non-enumerable structure cannot even be expressed with an infinite set of sentences. These facts follow from the compactness and Löwenheim-Skolem theorems.

Part III

Proofs and Completeness

Chapter 7

The Sequent Calculus

7.1 Rules and Derivations

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

Definition 7.1 (Sequent). A *sequent* is an expression of the form

$$\Gamma \Rightarrow \Delta$$

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

The intuitive idea behind a sequent is: if all of the sentences in the antecedent hold, then at least one of the sentences in the succedent holds. That is, if $\Gamma = \langle \varphi_1, \ldots, \varphi_m \rangle$ and $\Delta = \langle \psi_1, \ldots, \psi_n \rangle$, then $\Gamma \Rightarrow \Delta$ holds iff

$$(\varphi_1 \wedge \cdots \wedge \varphi_m) \rightarrow (\psi_1 \vee \cdots \vee \psi_n)$$

holds. There are two special cases: where Γ is empty and when Δ is empty. When Γ is empty, i.e., $m = 0$, $\Rightarrow \Delta$ holds iff $\psi_1 \vee \cdots \vee \psi_n$ holds. When Δ is empty, i.e., $n = 0$, $\Gamma \Rightarrow$ holds iff $\neg(\varphi_1 \wedge \cdots \wedge \varphi_m)$ does. We say a sequent is valid iff the corresponding sentence is valid.

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

Definition 7.2 (Initial Sequent). An *initial sequent* is a sequent of one of the following forms:

1. $\varphi \Rightarrow \varphi$
2. $\bot \Rightarrow$

for any sentence φ in the language.

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

7.2 Propositional Rules

Rules for $\neg$

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

Rules for $\wedge$

$$\frac{\varphi, \Gamma \Rightarrow \Delta}{\varphi \wedge \psi, \Gamma \Rightarrow \Delta}\,\wedge\mathrm{L} \qquad \frac{\psi, \Gamma \Rightarrow \Delta}{\varphi \wedge \psi, \Gamma \Rightarrow \Delta}\,\wedge\mathrm{L} \qquad \frac{\Gamma \Rightarrow \Delta, \varphi \qquad \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \varphi \wedge \psi}\,\wedge\mathrm{R}$$

Rules for $\vee$

$$\frac{\varphi, \Gamma \Rightarrow \Delta \qquad \psi, \Gamma \Rightarrow \Delta}{\varphi \vee \psi, \Gamma \Rightarrow \Delta}\,\vee\mathrm{L} \qquad \frac{\Gamma \Rightarrow \Delta, \varphi}{\Gamma \Rightarrow \Delta, \varphi \vee \psi}\,\vee\mathrm{R} \qquad \frac{\Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \varphi \vee \psi}\,\vee\mathrm{R}$$

Rules for $\rightarrow$

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

7.3 Quantifier Rules

Rules for $\forall$

$$\frac{\varphi(t), \Gamma \Rightarrow \Delta}{\forall x\, \varphi(x), \Gamma \Rightarrow \Delta}\ \forall\text{L} \qquad \frac{\Gamma \Rightarrow \Delta, \varphi(a)}{\Gamma \Rightarrow \Delta, \forall x\, \varphi(x)}\ \forall\text{R}$$

In $\forall$L, t is a closed term (i.e., one without variables). In $\forall$R, a is a constant symbol which must not occur anywhere in the lower sequent of the $\forall$R rule. We call a the *eigenvariable* of the $\forall$R inference.[1]

Rules for $\exists$

$$\frac{\varphi(a), \Gamma \Rightarrow \Delta}{\exists x\, \varphi(x), \Gamma \Rightarrow \Delta}\ \exists\text{L} \qquad \frac{\Gamma \Rightarrow \Delta, \varphi(t)}{\Gamma \Rightarrow \Delta, \exists x\, \varphi(x)}\ \exists\text{R}$$

Again, t is a closed term, and a is a constant symbol which does not occur in the lower sequent of the $\exists$L rule. We call a the *eigenvariable* of the $\exists$L inference.

The condition that an eigenvariable not occur in the lower sequent of the $\forall$R or $\exists$L inference is called the *eigenvariable condition*.

Recall the convention that when φ is a formula with the variable x free, we indicate this by writing $\varphi(x)$. In the same context, $\varphi(t)$ then is short for $\varphi[t/x]$. So we could also write the $\exists$R rule as:

$$\frac{\Gamma \Rightarrow \Delta, \varphi[t/x]}{\Gamma \Rightarrow \Delta, \exists x\, \varphi}\ \exists\text{R}$$

Note that t may already occur in φ, e.g., φ might be $P(t, x)$. Thus, inferring $\Gamma \Rightarrow \Delta, \exists x\, P(t, x)$ from $\Gamma \Rightarrow \Delta, P(t, t)$ is a correct application of $\exists$R—you may "replace" one or more, and not necessarily all, occurrences of t in the premise by the bound variable x. However, the eigenvariable conditions in $\forall$R and $\exists$L require that the constant symbol a does not occur in φ. So, you cannot correctly infer $\Gamma \Rightarrow \Delta, \forall x\, P(a, x)$ from $\Gamma \Rightarrow \Delta, P(a, a)$ using $\forall$R.

In $\exists$R and $\forall$L there are no restrictions on the term t. On the other hand, in the $\exists$L and $\forall$R rules, the eigenvariable condition requires that the constant symbol a does not occur anywhere outside of $\varphi(a)$ in the upper sequent. It is necessary to ensure that the system is sound, i.e., only derives sequents that are valid. Without this condition, the following would be allowed:

[1] We use the term "eigenvariable" even though a in the above rule is a constant symbol. This has historical reasons.

$$\dfrac{\dfrac{\varphi(a) \Rightarrow \varphi(a)}{\exists x\, \varphi(x) \Rightarrow \varphi(a)}\,{*\exists}\mathrm{L}}{\exists x\, \varphi(x) \Rightarrow \forall x\, \varphi(x)}\,\forall\mathrm{R} \qquad \dfrac{\dfrac{\varphi(a) \Rightarrow \varphi(a)}{\varphi(a) \Rightarrow \forall x\, \varphi(x)}\,{*\forall}\mathrm{R}}{\exists x\, \varphi(x) \Rightarrow \forall x\, \varphi(x)}\,\exists\mathrm{L}$$

However, $\exists x\, \varphi(x) \Rightarrow \forall x\, \varphi(x)$ is not valid.

7.4 Structural Rules

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

Weakening

$$\dfrac{\Gamma \Rightarrow \Delta}{\varphi, \Gamma \Rightarrow \Delta}\,\mathrm{WL} \qquad \dfrac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \varphi}\,\mathrm{WR}$$

Contraction

$$\dfrac{\varphi, \varphi, \Gamma \Rightarrow \Delta}{\varphi, \Gamma \Rightarrow \Delta}\,\mathrm{CL} \qquad \dfrac{\Gamma \Rightarrow \Delta, \varphi, \varphi}{\Gamma \Rightarrow \Delta, \varphi}\,\mathrm{CR}$$

Exchange

$$\dfrac{\Gamma, \varphi, \psi, \Pi \Rightarrow \Delta}{\Gamma, \psi, \varphi, \Pi \Rightarrow \Delta}\,\mathrm{XL} \qquad \dfrac{\Gamma \Rightarrow \Delta, \varphi, \psi, \Lambda}{\Gamma \Rightarrow \Delta, \psi, \varphi, \Lambda}\,\mathrm{XR}$$

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

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

$$\dfrac{\Gamma \Rightarrow \Delta, \varphi \qquad \varphi, \Pi \Rightarrow \Lambda}{\Gamma, \Pi \Rightarrow \Delta, \Lambda}\,\mathrm{Cut}$$

7.5 Derivations

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

Definition 7.3 (LK derivation). An **LK**-*derivation* of a sequent S is a finite tree of sequents satisfying the following conditions:

1. The topmost sequents of the tree are initial sequents.
2. The bottommost sequent of the tree is S.
3. Every sequent in the tree except S is a premise of a correct application of an inference rule whose conclusion stands directly below that sequent in the tree.

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

Example 7.4. Every initial sequent, e.g., $\chi \Rightarrow \chi$ is a derivation. We can obtain a new derivation from this by applying, say, the WL rule,

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

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

$$\frac{\chi \Rightarrow \chi}{\theta, \chi \Rightarrow \chi}\,\text{WL}$$

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

$$\dfrac{\dfrac{\chi \Rightarrow \chi}{\theta, \chi \Rightarrow \chi}\,\text{WL}}{\chi, \theta \Rightarrow \chi}\,\text{XL}$$

In this application of the rule, which was given as

$$\frac{\Gamma, \varphi, \psi, \Pi \Rightarrow \Delta}{\Gamma, \psi, \varphi, \Pi \Rightarrow \Delta,}\,\text{XL}$$

both Γ and Π were empty, Δ is χ, and the roles of φ and ψ are played by θ and χ, respectively. In much the same way, we also see that

$$\dfrac{\theta \Rightarrow \theta}{\chi, \theta \Rightarrow \theta}\,\text{WL}$$

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

$$\dfrac{\Gamma \Rightarrow \Delta, \varphi \qquad \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \varphi \wedge \psi}\,\wedge\text{R}$$

In our case, the premises must match the last sequents of the derivations ending in the premises. That means that Γ is χ, θ, Δ is empty, φ is χ and ψ is θ. So the conclusion, if the inference should be correct, is $\chi, \theta \Rightarrow \chi \wedge \theta$.

$$\dfrac{\dfrac{\dfrac{\chi \Rightarrow \chi}{\theta, \chi \Rightarrow \chi}\,\text{WL}}{\chi, \theta \Rightarrow \chi}\,\text{XL} \qquad \dfrac{\theta \Rightarrow \theta}{\chi, \theta \Rightarrow \theta}\,\text{WL}}{\chi, \theta \Rightarrow \chi \wedge \theta}\,\wedge\text{R}$$

Of course, we can also reverse the premises, then φ would be θ and ψ would be χ.

$$\dfrac{\dfrac{\theta \Rightarrow \theta}{\chi, \theta \Rightarrow \theta}\,\text{WL} \qquad \dfrac{\dfrac{\chi \Rightarrow \chi}{\theta, \chi \Rightarrow \chi}\,\text{WL}}{\chi, \theta \Rightarrow \chi}\,\text{XL}}{\chi, \theta \Rightarrow \theta \wedge \chi}\,\wedge\text{R}$$

7.6 Examples of Derivations

Example 7.5. Give an **LK**-derivation for the sequent $\varphi \wedge \psi \Rightarrow \varphi$.

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

$$\dfrac{}{\varphi \wedge \psi \Rightarrow \varphi}$$

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

$$\dfrac{}{\varphi \wedge \psi \Rightarrow \varphi}\,\wedge\text{L}$$

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

$$\dfrac{\varphi \Rightarrow \varphi}{\varphi \wedge \psi \Rightarrow \varphi}\,\wedge\text{L}$$

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

Example 7.6. Give an **LK**-derivation for the sequent $\neg\varphi \vee \psi \Rightarrow \varphi \to \psi$.

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

$$\dfrac{}{\neg\varphi \vee \psi \Rightarrow \varphi \to \psi}$$

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

$$\dfrac{\dfrac{}{\varphi, \neg\varphi \vee \psi \Rightarrow \psi}}{\neg\varphi \vee \psi \Rightarrow \varphi \to \psi}\,\to\text{R}$$

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

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\neg\varphi, \varphi \Rightarrow \psi} \qquad \dfrac{}{\psi, \varphi \Rightarrow \psi}}{\neg\varphi \vee \psi, \varphi \Rightarrow \psi}\,\vee\text{L}}{\varphi, \neg\varphi \vee \psi \Rightarrow \psi}\,\text{XR}}{\neg\varphi \vee \psi \Rightarrow \varphi \to \psi}\,\to\text{R}$$

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

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\neg\varphi, \varphi \Rightarrow \psi} \qquad \dfrac{\dfrac{\dfrac{}{\psi \Rightarrow \psi}}{\varphi, \psi \Rightarrow \psi}\,\text{WL}}{\psi, \varphi \Rightarrow \psi}\,\text{XL}}{\neg\varphi \vee \psi, \varphi \Rightarrow \psi}\,\vee\text{L}}{\varphi, \neg\varphi \vee \psi \Rightarrow \psi}\,\text{XR}}{\neg\varphi \vee \psi \Rightarrow \varphi \to \psi}\,\to\text{R}$$

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

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\varphi \Rightarrow \psi, \varphi}}{\neg\varphi, \varphi \Rightarrow \psi}\neg\text{L} \quad \dfrac{\dfrac{\psi \Rightarrow \psi}{\varphi, \psi \Rightarrow \psi}\text{WL}}{\psi, \varphi \Rightarrow \psi}\text{XL}}{\neg\varphi \vee \psi, \varphi \Rightarrow \psi}\vee\text{L}}{\dfrac{\varphi, \neg\varphi \vee \psi \Rightarrow \psi}{\neg\varphi \vee \psi \Rightarrow \varphi \to \psi}\to\text{R}}\text{XR}$$

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

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\varphi \Rightarrow \varphi}}{\varphi \Rightarrow \varphi, \psi}\text{WR}}{\varphi \Rightarrow \varphi, \psi}\text{XR}}{\varphi \Rightarrow \psi, \varphi}\neg\text{L} \quad \dfrac{\dfrac{\psi \Rightarrow \psi}{\varphi, \psi \Rightarrow \psi}\text{WL}}{\psi, \varphi \Rightarrow \psi}\text{XL}}{\neg\varphi, \varphi \Rightarrow \psi}\vee\text{L}}{\dfrac{\varphi, \neg\varphi \vee \psi \Rightarrow \psi}{\neg\varphi \vee \psi \Rightarrow \varphi \to \psi}\to\text{R}}\text{XR}$$

Example 7.7. Give an **LK**-derivation of the sequent $\neg\varphi \vee \neg\psi \Rightarrow \neg(\varphi \wedge \psi)$

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

$$\dfrac{}{\neg\varphi \vee \neg\psi \Rightarrow \neg(\varphi \wedge \psi)}$$

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

$$\dfrac{\dfrac{}{\varphi \wedge \psi, \neg\varphi \vee \neg\psi \Rightarrow}}{\neg\varphi \vee \neg\psi \Rightarrow \neg(\varphi \wedge \psi)}\neg\text{R}$$

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

$$\dfrac{\dfrac{\dfrac{}{\varphi, \neg\varphi \vee \neg\psi \Rightarrow}}{\varphi \wedge \psi, \neg\varphi \vee \neg\psi \Rightarrow}\wedge\text{L}}{\neg\varphi \vee \neg\psi \Rightarrow \neg(\varphi \wedge \psi)}\neg\text{R}$$

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

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\varphi \Rightarrow \varphi}{\neg\varphi, \varphi \Rightarrow}\,\neg\mathrm{L} \qquad \dfrac{\dfrac{}{\varphi \Rightarrow \psi}\,?}{\neg\psi, \varphi \Rightarrow}\,\neg\mathrm{L}}{\neg\varphi \vee \neg\psi, \varphi \Rightarrow}\,\vee\mathrm{L}}{\varphi, \neg\varphi \vee \neg\psi \Rightarrow}\,\mathrm{XL}}{\varphi \wedge \psi, \neg\varphi \vee \neg\psi \Rightarrow}\,\wedge\mathrm{L}}{\neg\varphi \vee \neg\psi \Rightarrow \neg(\varphi \wedge \psi)}\,\neg\mathrm{R}$$

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

$$\dfrac{\dfrac{\dfrac{\dfrac{\neg\varphi, \varphi \wedge \psi \Rightarrow}{} \qquad \dfrac{\neg\psi, \varphi \wedge \psi \Rightarrow}{}}{\neg\varphi \vee \neg\psi, \varphi \wedge \psi \Rightarrow}\,\vee\mathrm{L}}{\varphi \wedge \psi, \neg\varphi \vee \neg\psi \Rightarrow}\,\mathrm{XL}}{\neg\varphi \vee \neg\psi \Rightarrow \neg(\varphi \wedge \psi)}\,\neg\mathrm{R}$$

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

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\varphi \Rightarrow \varphi}{\varphi \wedge \psi \Rightarrow \varphi}\,\wedge\mathrm{L}}{\neg\varphi, \varphi \wedge \psi \Rightarrow}\,\neg\mathrm{L} \qquad \dfrac{\dfrac{\psi \Rightarrow \psi}{\varphi \wedge \psi \Rightarrow \psi}\,\wedge\mathrm{L}}{\neg\psi, \varphi \wedge \psi \Rightarrow}\,\neg\mathrm{L}}{\neg\varphi \vee \neg\psi, \varphi \wedge \psi \Rightarrow}\,\vee\mathrm{L}}{\varphi \wedge \psi, \neg\varphi \vee \neg\psi \Rightarrow}\,\mathrm{XL}}{\neg\varphi \vee \neg\psi \Rightarrow \neg(\varphi \wedge \psi)}\,\neg\mathrm{R}$$

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

Example 7.8. So far we haven't used the contraction rule, but it is sometimes required. Here's an example where that happens. Suppose we want to prove $\Rightarrow \varphi \vee \neg\varphi$. Applying $\vee$R backwards would give us one of these two derivations:

$$\dfrac{\dfrac{}{\Rightarrow \varphi}}{\Rightarrow \varphi \vee \neg\varphi}\,\vee\mathrm{R} \qquad \dfrac{\dfrac{\dfrac{}{\varphi \Rightarrow}}{\Rightarrow \neg\varphi}\,\neg\mathrm{R}}{\Rightarrow \varphi \vee \neg\varphi}\,\vee\mathrm{R}$$

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

$$\dfrac{\dfrac{\dfrac{}{\Rightarrow \varphi \vee \neg\varphi, \varphi}}{\Rightarrow \varphi \vee \neg\varphi, \varphi \vee \neg\varphi}\,\vee\mathrm{R}}{\Rightarrow \varphi \vee \neg\varphi}\,\mathrm{C}\mathrm{R}$$

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

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\varphi \Rightarrow \varphi}}{\Rightarrow \varphi, \neg\varphi}\,\neg\mathrm{R}}{\Rightarrow \varphi, \varphi \vee \neg\varphi}\,\vee\mathrm{R}}{\Rightarrow \varphi \vee \neg\varphi, \varphi}\,\mathrm{X}\mathrm{R}}{\Rightarrow \varphi \vee \neg\varphi, \varphi \vee \neg\varphi}\,\vee\mathrm{R}}{\Rightarrow \varphi \vee \neg\varphi}\,\mathrm{C}\mathrm{R}$$

7.7 Derivations with Quantifiers

Example 7.9. Give an **LK**-derivation of the sequent $\exists x\, \neg\varphi(x) \Rightarrow \neg\forall x\, \varphi(x)$.

When dealing with quantifiers, we have to make sure not to violate the eigenvariable condition, and sometimes this requires us to play around with the order of carrying out certain inferences. In general, it helps to try and take care of rules subject to the eigenvariable condition first (they will be lower down in the finished proof). Also, it is a good idea to try and look ahead and try to guess what the initial sequent might look like. In our case, it will have to be something like $\varphi(a) \Rightarrow \varphi(a)$. That means that when we are "reversing" the quantifier rules, we will have to pick the same term—what we will call a—for both the $\forall$ and the $\exists$ rule. If we picked different terms for each rule, we would end up with something like $\varphi(a) \Rightarrow \varphi(b)$, which, of course, is not derivable.

Starting as usual, we write

$$\dfrac{}{\exists x\, \neg\varphi(x) \Rightarrow \neg\forall x\, \varphi(x)}$$

We could either carry out the $\exists$L rule or the $\neg$R rule. Since the $\exists$L rule is subject to the eigenvariable condition, it's a good idea to take care of it sooner rather than later, so we'll do that one first.

$$\dfrac{\dfrac{}{\neg\varphi(a) \Rightarrow \neg\forall x\, \varphi(x)}}{\exists x\, \neg\varphi(x) \Rightarrow \neg\forall x\, \varphi(x)}\,\exists\mathrm{L}$$

Applying the $\neg$L and $\neg$R rules backwards, we get

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\forall x\, \varphi(x) \Rightarrow \varphi(a)}}{\neg\varphi(a), \forall x\, \varphi(x) \Rightarrow}\,\neg\mathrm{L}}{\forall x\, \varphi(x), \neg\varphi(a) \Rightarrow}\,\mathrm{X}\mathrm{L}}{\neg\varphi(a) \Rightarrow \neg\forall x\varphi(x)}\,\neg\mathrm{R}}{\exists x\neg\varphi(x) \Rightarrow \neg\forall x\varphi(x)}\,\exists\mathrm{L}$$

At this point, our only option is to carry out the $\forall$L rule. Since this rule is not subject to the eigenvariable restriction, we're in the clear. Remember, we want to try and obtain an initial sequent (of the form $\varphi(a) \Rightarrow \varphi(a)$), so we should choose a as our argument for φ when we apply the rule.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\varphi(a) \Rightarrow \varphi(a)}{\forall x\, \varphi(x) \Rightarrow \varphi(a)}\,\forall\text{L}}{\neg\varphi(a), \forall x\, \varphi(x) \Rightarrow}\,\neg\text{L}}{\forall x\, \varphi(x), \neg\varphi(a) \Rightarrow}\,\text{XL}}{\neg\varphi(a) \Rightarrow \neg\forall x\, \varphi(x)}\,\neg\text{R}}{\exists x\, \neg\varphi(x) \Rightarrow \neg\forall x\, \varphi(x)}\,\exists\text{L}$$

It is important, especially when dealing with quantifiers, to double check at this point that the eigenvariable condition has not been violated. Since the only rule we applied that is subject to the eigenvariable condition was $\exists$L, and the eigenvariable a does not occur in its lower sequent (the end-sequent), this is a correct derivation.

7.8 Proof-Theoretic Notions

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

Definition 7.10 (Theorems). A sentence φ is a *theorem* if there is a derivation in **LK** of the sequent $\Rightarrow \varphi$. We write $\vdash \varphi$ if φ is a theorem and $\nvdash \varphi$ if it is not.

Definition 7.11 (Derivability). A sentence φ is *derivable from* a set of sentences Γ, $\Gamma \vdash \varphi$, iff there is a finite subset $\Gamma_0 \subseteq \Gamma$ and a sequence Γ_0' of the sentences in Γ_0 such that **LK** derives $\Gamma_0' \Rightarrow \varphi$. If φ is not derivable from Γ we write $\Gamma \nvdash \varphi$.

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

$$\dfrac{\dfrac{\dfrac{\vdots}{\psi, \psi, \chi \Rightarrow \varphi}}{\psi, \chi \Rightarrow \varphi}\,\mathrm{CL}}{\chi, \psi \Rightarrow \varphi}\,\mathrm{XL}}{\chi, \chi, \psi \Rightarrow \varphi}\,\mathrm{WL}$$

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

Definition 7.12 (Consistency). A set of sentences Γ is *inconsistent* iff there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that **LK** derives $\Gamma_0 \Rightarrow$. If Γ is not inconsistent, i.e., if for every finite $\Gamma_0 \subseteq \Gamma$, **LK** does not derive $\Gamma_0 \Rightarrow$, we say it is *consistent*.

Proposition 7.13 (Reflexivity). *If $\varphi \in \Gamma$, then $\Gamma \vdash \varphi$.*

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

Proposition 7.14 (Monotonicity). *If $\Gamma \subseteq \Delta$ and $\Gamma \vdash \varphi$, then $\Delta \vdash \varphi$.*

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

Proposition 7.15 (Transitivity). *If $\Gamma \vdash \varphi$ and $\{\varphi\} \cup \Delta \vdash \psi$, then $\Gamma \cup \Delta \vdash \psi$.*

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

$$\dfrac{\begin{array}{cc}\vdots\,\pi_0 & \vdots\,\pi_1\\ \Gamma_0 \Rightarrow \varphi & \varphi, \Delta_0 \Rightarrow \psi\end{array}}{\Gamma_0, \Delta_0 \Rightarrow \psi}\,\mathrm{Cut}$$

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

Note that this means that in particular if $\Gamma \vdash \varphi$ and $\varphi \vdash \psi$, then $\Gamma \vdash \psi$. It follows also that if $\varphi_1, \ldots, \varphi_n \vdash \psi$ and $\Gamma \vdash \varphi_i$ for each i, then $\Gamma \vdash \psi$.

Proposition 7.16. *Γ is inconsistent iff $\Gamma \vdash \varphi$ for every sentence φ.*

Proof. Exercise. □

Proposition 7.17 (Compactness). *1. If $\Gamma \vdash \varphi$ then there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \vdash \varphi$.*

2. *If every finite subset of* Γ *is consistent, then* Γ *is consistent.*

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

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

7.9 Derivability and Consistency

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

Proposition 7.18. *If* $\Gamma \vdash \varphi$ *and* $\Gamma \cup \{\varphi\}$ *is inconsistent, then* Γ *is inconsistent.*

Proof. There are finite Γ_0 and $\Gamma_1 \subseteq \Gamma$ such that **LK** derives $\Gamma_0 \Rightarrow \varphi$ and $\varphi, \Gamma_1 \Rightarrow$. Let the **LK**-derivation of $\Gamma_0 \Rightarrow \varphi$ be π_0 and the **LK**-derivation of $\Gamma_1, \varphi \Rightarrow$ be π_1. We can then derive

$$\dfrac{\begin{array}{cc}\vdots\,\pi_0 & \vdots\,\pi_1\\ \Gamma_0 \Rightarrow \varphi & \varphi, \Gamma_1 \Rightarrow\end{array}}{\Gamma_0, \Gamma_1 \Rightarrow}\,\text{Cut}$$

Since $\Gamma_0 \subseteq \Gamma$ and $\Gamma_1 \subseteq \Gamma$, $\Gamma_0 \cup \Gamma_1 \subseteq \Gamma$, hence Γ is inconsistent. □

Proposition 7.19. $\Gamma \vdash \varphi$ *iff* $\Gamma \cup \{\neg\varphi\}$ *is inconsistent.*

Proof. First suppose $\Gamma \vdash \varphi$, i.e., there is a derivation π_0 of $\Gamma \Rightarrow \varphi$. By adding a $\neg$L rule, we obtain a derivation of $\neg\varphi, \Gamma \Rightarrow$, i.e., $\Gamma \cup \{\neg\varphi\}$ is inconsistent.

If $\Gamma \cup \{\neg\varphi\}$ is inconsistent, there is a derivation π_1 of $\neg\varphi, \Gamma \Rightarrow$. The following is a derivation of $\Gamma \Rightarrow \varphi$:

$$\dfrac{\dfrac{\varphi \Rightarrow \varphi}{\Rightarrow \varphi, \neg\varphi}\,\neg\text{R} \qquad \begin{array}{c}\vdots\,\pi_1\\ \neg\varphi, \Gamma \Rightarrow\end{array}}{\Gamma \Rightarrow \varphi}\,\text{Cut}$$

□

Proposition 7.20. *If* $\Gamma \vdash \varphi$ *and* $\neg\varphi \in \Gamma$, *then* Γ *is inconsistent.*

Proof. Suppose $\Gamma \vdash \varphi$ and $\neg\varphi \in \Gamma$. Then there is a derivation π of a sequent $\Gamma_0 \Rightarrow \varphi$. The sequent $\neg\varphi, \Gamma_0 \Rightarrow$ is also derivable:

$$\dfrac{\dfrac{\vdots\,\pi}{\Gamma_0 \Rightarrow \varphi} \qquad \dfrac{\dfrac{\varphi \Rightarrow \varphi}{\neg\varphi, \varphi \Rightarrow}\,\neg\mathrm{L}}{\varphi, \neg\varphi \Rightarrow}\,\mathrm{XL}}{\Gamma, \neg\varphi \Rightarrow}\,\mathrm{Cut}$$

Since $\neg\varphi \in \Gamma$ and $\Gamma_0 \subseteq \Gamma$, this shows that Γ is inconsistent. □

Proposition 7.21. *If $\Gamma \cup \{\varphi\}$ and $\Gamma \cup \{\neg\varphi\}$ are both inconsistent, then Γ is inconsistent.*

Proof. There are finite sets $\Gamma_0 \subseteq \Gamma$ and $\Gamma_1 \subseteq \Gamma$ and **LK**-derivations π_0 and π_1 of $\varphi, \Gamma_0 \Rightarrow$ and $\neg\varphi, \Gamma_1 \Rightarrow$, respectively. We can then derive

$$\dfrac{\dfrac{\vdots\,\pi_0}{\varphi, \Gamma_0 \Rightarrow}}{\Gamma_0 \Rightarrow \neg\varphi}\,\neg\mathrm{R} \qquad \dfrac{\vdots\,\pi_1}{\neg\varphi, \Gamma_1 \Rightarrow}}{\Gamma_0, \Gamma_1 \Rightarrow}\,\mathrm{Cut}$$

Since $\Gamma_0 \subseteq \Gamma$ and $\Gamma_1 \subseteq \Gamma$, $\Gamma_0 \cup \Gamma_1 \subseteq \Gamma$. Hence Γ is inconsistent. □

7.10 Derivability and the Propositional Connectives

We establish that the derivability relation $\vdash$ of the sequent calculus is strong enough to establish some basic facts involving the propositional connectives, such as that $\varphi \wedge \psi \vdash \varphi$ and $\varphi, \varphi \to \psi \vdash \psi$ (modus ponens). These facts are needed for the proof of the completeness theorem.

Proposition 7.22. *1. Both $\varphi \wedge \psi \vdash \varphi$ and $\varphi \wedge \psi \vdash \psi$.*

2. $\varphi, \psi \vdash \varphi \wedge \psi$.

Proof. 1. Both sequents $\varphi \wedge \psi \Rightarrow \varphi$ and $\varphi \wedge \psi \Rightarrow \psi$ are derivable:

$$\dfrac{\varphi \Rightarrow \varphi}{\varphi \wedge \psi \Rightarrow \varphi}\,\wedge\mathrm{L} \qquad \dfrac{\psi \Rightarrow \psi}{\varphi \wedge \psi \Rightarrow \psi}\,\wedge\mathrm{L}$$

2. Here is a derivation of the sequent $\varphi, \psi \Rightarrow \varphi \wedge \psi$:

$$\dfrac{\varphi \Rightarrow \varphi \qquad \psi \Rightarrow \psi}{\varphi, \psi \Rightarrow \varphi \wedge \psi}\,\wedge\mathrm{R}$$

□

Proposition 7.23. *1. $\varphi \vee \psi, \neg\varphi, \neg\psi$ is inconsistent.*

2. Both $\varphi \vdash \varphi \vee \psi$ and $\psi \vdash \varphi \vee \psi$.

Proof. 1. We give a derivation of the sequent $\varphi \vee \psi, \neg\varphi, \neg\psi \Rightarrow$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\varphi \Rightarrow \varphi}{\neg\varphi, \varphi \Rightarrow}\neg\mathrm{L}}{\varphi, \neg\varphi, \neg\psi \Rightarrow} \quad \dfrac{\dfrac{\dfrac{\psi \Rightarrow \psi}{\neg\psi, \psi \Rightarrow}\neg\mathrm{L}}{\psi, \neg\varphi, \neg\psi \Rightarrow}}{\varphi \vee \psi, \neg\varphi, \neg\psi \Rightarrow}\vee\mathrm{L}$$

(Recall that double inference lines indicate several weakening, contraction, and exchange inferences.)

2. Both sequents $\varphi \Rightarrow \varphi \vee \psi$ and $\psi \Rightarrow \varphi \vee \psi$ have derivations:

$$\dfrac{\varphi \Rightarrow \varphi}{\varphi \Rightarrow \varphi \vee \psi}\vee\mathrm{R} \qquad \dfrac{\psi \Rightarrow \psi}{\psi \Rightarrow \varphi \vee \psi}\vee\mathrm{R}$$ □

Proposition 7.24. *1. $\varphi, \varphi \to \psi \vdash \psi$.*

2. Both $\neg\varphi \vdash \varphi \to \psi$ and $\psi \vdash \varphi \to \psi$.

Proof. 1. The sequent $\varphi \to \psi, \varphi \Rightarrow \psi$ is derivable:

$$\dfrac{\varphi \Rightarrow \varphi \qquad \psi \Rightarrow \psi}{\varphi \to \psi, \varphi \Rightarrow \psi}\to\mathrm{L}$$

2. Both sequents $\neg\varphi \Rightarrow \varphi \to \psi$ and $\psi \Rightarrow \varphi \to \psi$ are derivable:

$$\dfrac{\dfrac{\dfrac{\dfrac{\varphi \Rightarrow \varphi}{\neg\varphi, \varphi \Rightarrow}\neg\mathrm{L}}{\varphi, \neg\varphi \Rightarrow}\mathrm{XL}}{\varphi, \neg\varphi \Rightarrow \psi}\mathrm{WR}}{\neg\varphi \Rightarrow \varphi \to \psi}\to\mathrm{R} \qquad \dfrac{\dfrac{\psi \Rightarrow \psi}{\varphi, \psi \Rightarrow \psi}\mathrm{WL}}{\psi \Rightarrow \varphi \to \psi}\to\mathrm{R}$$ □

7.11 Derivability and the Quantifiers

The completeness theorem also requires that the sequent calculus rules rules yield the facts about $\vdash$ established in this section.

Theorem 7.25. *If c is a constant not occurring in Γ or $\varphi(x)$ and $\Gamma \vdash \varphi(c)$, then $\Gamma \vdash \forall x\, \varphi(x)$.*

Proof. Let π_0 be an **LK**-derivation of $\Gamma_0 \Rightarrow \varphi(c)$ for some finite $\Gamma_0 \subseteq \Gamma$. By adding a $\forall$R inference, we obtain a derivation of $\Gamma_0 \Rightarrow \forall x\, \varphi(x)$, since c does not occur in Γ or $\varphi(x)$ and thus the eigenvariable condition is satisfied. □

Proposition 7.26. *1. $\varphi(t) \vdash \exists x\, \varphi(x)$.*

2. $\forall x\, \varphi(x) \vdash \varphi(t)$.

Proof. 1. The sequent $\varphi(t) \Rightarrow \exists x\, \varphi(x)$ is derivable:

$$\dfrac{\varphi(t) \;\Rightarrow\; \varphi(t)}{\varphi(t) \;\Rightarrow\; \exists x\, \varphi(x)}\,\exists\mathrm{R}$$

2. The sequent $\forall x\, \varphi(x) \Rightarrow \varphi(t)$ is derivable:

$$\dfrac{\varphi(t) \;\Rightarrow\; \varphi(t)}{\forall x\, \varphi(x) \;\Rightarrow\; \varphi(t)}\,\forall\mathrm{L}$$

□

7.12 Soundness

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

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

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

Because all these proof-theoretic properties are defined via derivability in the sequent calculus of certain sequents, proving (1)–(3) above requires proving something about the semantic properties of derivable sequents. We will first define what it means for a sequent to be *valid*, and then show that every derivable sequent is valid. (1)–(3) then follow as corollaries from this result.

Definition 7.27. A structure $\mathfrak{M}$ *satisfies* a sequent $\Gamma \Rightarrow \Delta$ iff either $\mathfrak{M} \nvDash \varphi$ for some $\varphi \in \Gamma$ or $\mathfrak{M} \vDash \varphi$ for some $\varphi \in \Delta$.

A sequent is *valid* iff every structure $\mathfrak{M}$ satisfies it.

Theorem 7.28 (Soundness). *If* **LK** *derives* $\Theta \Rightarrow \Xi$, *then* $\Theta \Rightarrow \Xi$ *is valid.*

Proof. Let π be a derivation of $\Theta \Rightarrow \Xi$. We proceed by induction on the number of inferences n in π.

If the number of inferences is 0, then π consists only of an initial sequent. Every initial sequent $\varphi \Rightarrow \varphi$ is obviously valid, since for every $\mathfrak{M}$, either $\mathfrak{M} \nvDash \varphi$ or $\mathfrak{M} \vDash \varphi$.

If the number of inferences is greater than 0, we distinguish cases according to the type of the lowermost inference. By induction hypothesis, we can assume that the premises of that inference are valid, since the number of inferences in the derivation of any premise is smaller than n.

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

1. The last inference is a weakening. Then $\Theta \Rightarrow \Xi$ is either $\varphi, \Gamma \Rightarrow \Delta$ (if the last inference is WL) or $\Gamma \Rightarrow \Delta, \varphi$ (if it's WR), and the derivation ends in one of

$$\dfrac{\begin{array}{c}\vdots\\ \Gamma \Rightarrow \Delta\end{array}}{\varphi, \Gamma \Rightarrow \Delta}\,\mathrm{WL} \qquad \dfrac{\begin{array}{c}\vdots\\ \Gamma \Rightarrow \Delta\end{array}}{\Gamma \Rightarrow \Delta, \varphi}\,\mathrm{WR}$$

   By induction hypothesis, $\Gamma \Rightarrow \Delta$ is valid, i.e., for every structure $\mathfrak{M}$, either there is some $\chi \in \Gamma$ such that $\mathfrak{M} \nvDash \chi$ or there is some $\chi \in \Delta$ such that $\mathfrak{M} \vDash \chi$.

   If $\mathfrak{M} \nvDash \chi$ for some $\chi \in \Gamma$, then $\chi \in \Theta$ as well since $\Theta = \varphi, \Gamma$, and so $\mathfrak{M} \nvDash \chi$ for some $\chi \in \Theta$. Similarly, if $\mathfrak{M} \vDash \chi$ for some $\chi \in \Delta$, as $\chi \in \Xi$, $\mathfrak{M} \vDash \chi$ for some $\chi \in \Xi$. Consequently, $\Theta \Rightarrow \Xi$ is valid.

2. The last inference is ¬L: Then the premise of the last inference is $\Gamma \Rightarrow \Delta, \varphi$ and the conclusion is $\neg\varphi, \Gamma \Rightarrow \Delta$, i.e., the derivation ends in

$$\dfrac{\begin{array}{c}\vdots\\ \Gamma \Rightarrow \Delta, \varphi\end{array}}{\neg\varphi, \Gamma \Rightarrow \Delta}\,\neg\mathrm{L}$$

   and $\Theta = \neg\varphi, \Gamma$ while $\Xi = \Delta$.

   The induction hypothesis tells us that $\Gamma \Rightarrow \Delta, \varphi$ is valid, i.e., for every $\mathfrak{M}$, either (a) for some $\chi \in \Gamma$, $\mathfrak{M} \nvDash \chi$, or (b) for some $\chi \in \Delta$, $\mathfrak{M} \vDash \chi$, or (c) $\mathfrak{M} \vDash \varphi$. We want to show that $\Theta \Rightarrow \Xi$ is also valid. Let $\mathfrak{M}$ be a structure. If (a) holds, then there is $\chi \in \Gamma$ so that $\mathfrak{M} \nvDash \chi$, but $\chi \in \Theta$ as well. If (b) holds, there is $\chi \in \Delta$ such that $\mathfrak{M} \vDash \chi$, but $\chi \in \Xi$ as well. Finally, if $\mathfrak{M} \vDash \varphi$, then $\mathfrak{M} \nvDash \neg\varphi$. Since $\neg\varphi \in \Theta$, there is $\chi \in \Theta$ such that $\mathfrak{M} \nvDash \chi$. Consequently, $\Theta \Rightarrow \Xi$ is valid.

3. The last inference is ¬R: Exercise.

4. The last inference is ∧L: There are two variants: $\varphi \wedge \psi$ may be inferred on the left from φ or from ψ on the left side of the premise. In the first case, the π ends in

$$\dfrac{\begin{array}{c}\vdots\\ \varphi, \Gamma \Rightarrow \Delta\end{array}}{\varphi \wedge \psi, \Gamma \Rightarrow \Delta} \wedge\mathrm{L}$$

and $\Theta = \varphi \wedge \psi, \Gamma$ while $\Xi = \Delta$. Consider a structure $\mathfrak{M}$. Since by induction hypothesis, $\varphi, \Gamma \Rightarrow \Delta$ is valid, (a) $\mathfrak{M} \nvDash \varphi$, (b) $\mathfrak{M} \nvDash \chi$ for some $\chi \in \Gamma$, or (c) $\mathfrak{M} \vDash \chi$ for some $\chi \in \Delta$. In case (a), $\mathfrak{M} \nvDash \varphi \wedge \psi$, so there is $\chi \in \Theta$ (namely, $\varphi \wedge \psi$) such that $\mathfrak{M} \nvDash \chi$. In case (b), there is $\chi \in \Gamma$ such that $\mathfrak{M} \nvDash \chi$, and $\chi \in \Theta$ as well. In case (c), there is $\chi \in \Delta$ such that $\mathfrak{M} \vDash \chi$, and $\chi \in \Xi$ as well since $\Xi = \Delta$. So in each case, $\mathfrak{M}$ satisfies $\varphi \wedge \psi, \Gamma \Rightarrow \Delta$. Since $\mathfrak{M}$ was arbitrary, $\Gamma \Rightarrow \Delta$ is valid. The case where $\varphi \wedge \psi$ is inferred from ψ is handled the same, changing φ to ψ.

5. The last inference is ∨R: There are two variants: $\varphi \vee \psi$ may be inferred on the right from φ or from ψ on the right side of the premise. In the first case, π ends in

$$\dfrac{\begin{array}{c}\vdots\\ \Gamma \Rightarrow \Delta, \varphi\end{array}}{\Gamma \Rightarrow \Delta, \varphi \vee \psi} \vee\mathrm{R}$$

Now $\Theta = \Gamma$ and $\Xi = \Delta, \varphi \vee \psi$. Consider a structure $\mathfrak{M}$. Since $\Gamma \Rightarrow \Delta, \varphi$ is valid, (a) $\mathfrak{M} \vDash \varphi$, (b) $\mathfrak{M} \nvDash \chi$ for some $\chi \in \Gamma$, or (c) $\mathfrak{M} \vDash \chi$ for some $\chi \in \Delta$. In case (a), $\mathfrak{M} \vDash \varphi \vee \psi$. In case (b), there is $\chi \in \Gamma$ such that $\mathfrak{M} \nvDash \chi$. In case (c), there is $\chi \in \Delta$ such that $\mathfrak{M} \vDash \chi$. So in each case, $\mathfrak{M}$ satisfies $\Gamma \Rightarrow \Delta, \varphi \vee \psi$, i.e., $\Theta \Rightarrow \Xi$. Since $\mathfrak{M}$ was arbitrary, $\Theta \Rightarrow \Xi$ is valid. The case where $\varphi \vee \psi$ is inferred from ψ is handled the same, changing φ to ψ.

6. The last inference is →R: Then π ends in

$$\dfrac{\begin{array}{c}\vdots\\ \varphi, \Gamma \Rightarrow \Delta, \psi\end{array}}{\Gamma \Rightarrow \Delta, \varphi \to \psi} \to\mathrm{R}$$

Again, the induction hypothesis says that the premise is valid; we want to show that the conclusion is valid as well. Let $\mathfrak{M}$ be arbitrary. Since $\varphi, \Gamma \Rightarrow \Delta, \psi$ is valid, at least one of the following cases obtains: (a) $\mathfrak{M} \nvDash \varphi$, (b) $\mathfrak{M} \vDash \psi$, (c) $\mathfrak{M} \nvDash \chi$ for some $\chi \in \Gamma$, or (d) $\mathfrak{M} \vDash \chi$ for some $\chi \in \Delta$. In cases (a) and (b), $\mathfrak{M} \vDash \varphi \to \psi$ and so there is a $\chi \in \Delta, \varphi \to \psi$ such that $\mathfrak{M} \vDash \chi$. In case (c), for some $\chi \in \Gamma$, $\mathfrak{M} \nvDash \chi$. In case (d), for some $\chi \in \Delta$, $\mathfrak{M} \vDash \chi$. In each case, $\mathfrak{M}$ satisfies $\Gamma \Rightarrow \Delta, \varphi \to \psi$. Since $\mathfrak{M}$ was arbitrary, $\Gamma \Rightarrow \Delta, \varphi \to \psi$ is valid.

7. The last inference is $\forall$L: Then there is a formula $\varphi(x)$ and a closed term t such that π ends in

$$\dfrac{\begin{array}{c}\vdots\\ \varphi(t), \Gamma \Rightarrow \Delta\end{array}}{\forall x\, \varphi(x), \Gamma \Rightarrow \Delta}\ \forall\text{L}$$

We want to show that the conclusion $\forall x\, \varphi(x), \Gamma \Rightarrow \Delta$ is valid. Consider a structure $\mathfrak{M}$. Since the premise $\varphi(t), \Gamma \Rightarrow \Delta$ is valid, (a) $\mathfrak{M} \nvDash \varphi(t)$, (b) $\mathfrak{M} \nvDash \chi$ for some $\chi \in \Gamma$, or (c) $\mathfrak{M} \vDash \chi$ for some $\chi \in \Delta$. In case (a), by Proposition 5.71, if $\mathfrak{M} \vDash \forall x\, \varphi(x)$, then $\mathfrak{M} \vDash \varphi(t)$. Since $\mathfrak{M} \nvDash \varphi(t)$, $\mathfrak{M} \nvDash \forall x\, \varphi(x)$. In case (b) and (c), $\mathfrak{M}$ also satisfies $\forall x\, \varphi(x), \Gamma \Rightarrow \Delta$. Since $\mathfrak{M}$ was arbitrary, $\forall x\, \varphi(x), \Gamma \Rightarrow \Delta$ is valid.

8. The last inference is $\exists$R: Exercise.

9. The last inference is $\forall$R: Then there is a formula $\varphi(x)$ and a constant symbol a such that π ends in

$$\dfrac{\begin{array}{c}\vdots\\ \Gamma \Rightarrow \Delta, \varphi(a)\end{array}}{\Gamma \Rightarrow \Delta, \forall x\, \varphi(x)}\ \forall\text{R}$$

where the eigenvariable condition is satisfied, i.e., a does not occur in $\varphi(x)$, Γ, or Δ. By induction hypothesis, the premise of the last inference is valid. We have to show that the conclusion is valid as well, i.e., that for any structure $\mathfrak{M}$, (a) $\mathfrak{M} \vDash \forall x\, \varphi(x)$, (b) $\mathfrak{M} \nvDash \chi$ for some $\chi \in \Gamma$, or (c) $\mathfrak{M} \vDash \chi$ for some $\chi \in \Delta$.

Suppose $\mathfrak{M}$ is an arbitrary structure. If (b) or (c) holds, we are done, so suppose neither holds: for all $\chi \in \Gamma$, $\mathfrak{M} \vDash \chi$, and for all $\chi \in \Delta$, $\mathfrak{M} \nvDash \chi$. We have to show that (a) holds, i.e., $\mathfrak{M} \vDash \forall x\, \varphi(x)$. By Proposition 5.59, if suffices to show that $\mathfrak{M}, s \vDash \varphi(x)$ for all variable assignments s. So let s

be an arbitrary variable assignment. Consider the structure $\mathfrak{M}'$ which is just like $\mathfrak{M}$ except $a^{\mathfrak{M}'} = s(x)$. By Corollary 5.61, for any $\chi \in \Gamma$, $\mathfrak{M}' \vDash \chi$ since a does not occur in Γ, and for any $\chi \in \Delta$, $\mathfrak{M}' \nvDash \chi$. But the premise is valid, so $\mathfrak{M}' \vDash \varphi(a)$. By Proposition 5.58, $\mathfrak{M}', s \vDash \varphi(a)$, since $\varphi(a)$ is a sentence. Now $s \sim_x s$ with $s(x) = \mathrm{Val}_s^{\mathfrak{M}'}(a)$, since we've defined $\mathfrak{M}'$ in just this way. So Proposition 5.63 applies, and we get $\mathfrak{M}', s \vDash \varphi(x)$. Since a does not occur in $\varphi(x)$, by Proposition 5.60, $\mathfrak{M}, s \vDash \varphi(x)$. Since s was arbitrary, we've completed the proof that $\mathfrak{M}, s \vDash \varphi(x)$ for all variable assignments.

10. The last inference is $\exists$L: Exercise.

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

1. The last inference is a cut: then π ends in

$$\dfrac{\begin{array}{cc}\vdots & \vdots \\ \Gamma \Rightarrow \Delta, \varphi & \varphi, \Pi \Rightarrow \Lambda\end{array}}{\Gamma, \Pi \Rightarrow \Delta, \Lambda}\ \mathrm{Cut}$$

   Let $\mathfrak{M}$ be a structure. By induction hypothesis, the premises are valid, so $\mathfrak{M}$ satisfies both premises. We distinguish two cases: (a) $\mathfrak{M} \nvDash \varphi$ and (b) $\mathfrak{M} \vDash \varphi$. In case (a), in order for $\mathfrak{M}$ to satisfy the left premise, it must satisfy $\Gamma \Rightarrow \Delta$. But then it also satisfies the conclusion. In case (b), in order for $\mathfrak{M}$ to satisfy the right premise, it must satisfy $\Pi \setminus \Lambda$. Again, $\mathfrak{M}$ satisfies the conclusion.

2. The last inference is $\wedge$R. Then π ends in

$$\dfrac{\begin{array}{cc}\vdots & \vdots \\ \Gamma \Rightarrow \Delta, \varphi & \Gamma \Rightarrow \Delta, \psi\end{array}}{\Gamma \Rightarrow \Delta, \varphi \wedge \psi}\ \wedge\mathrm{R}$$

   Consider a structure $\mathfrak{M}$. If $\mathfrak{M}$ satisfies $\Gamma \Rightarrow \Delta$, we are done. So suppose it doesn't. Since $\Gamma \Rightarrow \Delta, \varphi$ is valid by induction hypothesis, $\mathfrak{M} \vDash \varphi$. Similarly, since $\Gamma \Rightarrow \Delta, \psi$ is valid, $\mathfrak{M} \vDash \psi$. But then $\mathfrak{M} \vDash \varphi \wedge \psi$.

3. The last inference is $\vee$L: Exercise.

4. The last inference is $\rightarrow$L. Then π ends in

$$\frac{\begin{array}{c}\vdots\\ \Gamma \Rightarrow \Delta, \varphi\end{array} \qquad \begin{array}{c}\vdots\\ \psi, \Pi \Rightarrow \Lambda\end{array}}{\varphi \to \psi, \Gamma, \Pi \Rightarrow \Delta, \Lambda} \to\text{L}$$

Again, consider a structure $\mathfrak{M}$ and suppose $\mathfrak{M}$ doesn't satisfy $\Gamma, \Pi \Rightarrow \Delta, \Lambda$. We have to show that $\mathfrak{M} \nvDash \varphi \to \psi$. If $\mathfrak{M}$ doesn't satisfy $\Gamma, \Pi \Rightarrow \Delta, \Lambda$, it satisfies neither $\Gamma \Rightarrow \Delta$ nor $\Pi \Rightarrow \Lambda$. Since, $\Gamma \Rightarrow \Delta, \varphi$ is valid, we have $\mathfrak{M} \vDash \varphi$. Since $\psi, \Pi \Rightarrow \Lambda$ is valid, we have $\mathfrak{M} \nvDash \psi$. But then $\mathfrak{M} \nvDash \varphi \to \psi$, which is what we wanted to show. □

Corollary 7.29. *If $\vdash \varphi$ then φ is valid.*

Corollary 7.30. *If $\Gamma \vdash \varphi$ then $\Gamma \vDash \varphi$.*

Proof. If $\Gamma \vdash \varphi$ then for some finite subset $\Gamma_0 \subseteq \Gamma$, there is a derivation of $\Gamma_0 \Rightarrow \varphi$. By Theorem 7.28, every structure $\mathfrak{M}$ either makes some $\psi \in \Gamma_0$ false or makes φ true. Hence, if $\mathfrak{M} \vDash \Gamma$ then also $\mathfrak{M} \vDash \varphi$. □

Corollary 7.31. *If Γ is satisfiable, then it is consistent.*

Proof. We prove the contrapositive. Suppose that Γ is not consistent. Then there is a finite $\Gamma_0 \subseteq \Gamma$ and a derivation of $\Gamma_0 \Rightarrow \quad$. By Theorem 7.28, $\Gamma_0 \Rightarrow$ is valid. In other words, for every structure $\mathfrak{M}$, there is $\chi \in \Gamma_0$ so that $\mathfrak{M} \nvDash \chi$, and since $\Gamma_0 \subseteq \Gamma$, that χ is also in Γ. Thus, no $\mathfrak{M}$ satisfies Γ, and Γ is not satisfiable. □

7.13 Derivations with Identity predicate

Derivations with identity predicate require additional initial sequents and inference rules.

Definition 7.32 (Initial sequents for $=$). If t is a closed term, then $\Rightarrow t = t$ is an initial sequent.

The rules for $=$ are (t_1 and t_2 are closed terms):

$$\frac{t_1 = t_2, \Gamma \Rightarrow \Delta, \varphi(t_1)}{t_1 = t_2, \Gamma \Rightarrow \Delta, \varphi(t_2)} = \qquad \frac{t_1 = t_2, \Gamma \Rightarrow \Delta, \varphi(t_2)}{t_1 = t_2, \Gamma \Rightarrow \Delta, \varphi(t_1)} =$$

Example 7.33. If s and t are closed terms, then $s = t, \varphi(s) \vdash \varphi(t)$:

$$\dfrac{\dfrac{\varphi(s) \Rightarrow \varphi(s)}{s = t, \varphi(s) \Rightarrow \varphi(s)}\,\mathrm{WL}}{s = t, \varphi(s) \Rightarrow \varphi(t)}\,=$$

This may be familiar as the principle of substitutability of identicals, or Leibniz' Law.

LK proves that $=$ is symmetric and transitive:

$$\dfrac{\dfrac{\Rightarrow t_1 = t_1}{t_1 = t_2 \Rightarrow t_1 = t_1}\,\mathrm{WL}}{t_1 = t_2 \Rightarrow t_2 = t_1}\,= \qquad \dfrac{\dfrac{\dfrac{t_1 = t_2 \Rightarrow t_1 = t_2}{t_2 = t_3, t_1 = t_2 \Rightarrow t_1 = t_2}\,\mathrm{WL}}{t_2 = t_3, t_1 = t_2 \Rightarrow t_1 = t_3}\,=}{t_1 = t_2, t_2 = t_3 \Rightarrow t_1 = t_3}\,\mathrm{XL}$$

In the derivation on the left, the formula $x = t_1$ is our $\varphi(x)$. On the right, we take $\varphi(x)$ to be $t_1 = x$.

7.14 Soundness with Identity predicate

Proposition 7.34. **LK** *with initial sequents and rules for identity is sound.*

Proof. Initial sequents of the form $\Rightarrow t = t$ are valid, since for every structure $\mathfrak{M}$, $\mathfrak{M} \vDash t = t$. (Note that we assume the term t to be closed, i.e., it contains no variables, so variable assignments are irrelevant).

Suppose the last inference in a derivation is $=$. Then the premise is $t_1 = t_2, \Gamma \Rightarrow \Delta, \varphi(t_1)$ and the conclusion is $t_1 = t_2, \Gamma \Rightarrow \Delta, \varphi(t_2)$. Consider a structure $\mathfrak{M}$. We need to show that the conclusion is valid, i.e., if $\mathfrak{M} \vDash t_1 = t_2$ and $\mathfrak{M} \vDash \Gamma$, then either $\mathfrak{M} \vDash \chi$ for some $\chi \in \Delta$ or $\mathfrak{M} \vDash \varphi(t_2)$.

By induction hypothesis, the premise is valid. This means that if $\mathfrak{M} \vDash t_1 = t_2$ and $\mathfrak{M} \vDash \Gamma$ either (a) for some $\chi \in \Delta$, $\mathfrak{M} \vDash \chi$ or (b) $\mathfrak{M} \vDash \varphi(t_1)$. In case (a) we are done. Consider case (b). Let s be a variable assignment with $s(x) = \mathrm{Val}^{\mathfrak{M}}(t_1)$. By Proposition 5.58, $\mathfrak{M}, s \vDash \varphi(t_1)$. Since $s \sim_x s$, by Proposition 5.63, $\mathfrak{M}, s \vDash \varphi(x)$. since $\mathfrak{M} \vDash t_1 = t_2$, we have $\mathrm{Val}^{\mathfrak{M}}(t_1) = \mathrm{Val}^{\mathfrak{M}}(t_2)$, and hence $s(x) = \mathrm{Val}^{\mathfrak{M}}(t_2)$. By applying Proposition 5.63 again, we also have $\mathfrak{M}, s \vDash \varphi(t_2)$. By Proposition 5.58, $\mathfrak{M} \vDash \varphi(t_2)$. □

Chapter 8

The Completeness Theorem

8.1 Introduction

The completeness theorem is one of the most fundamental results about logic. It comes in two formulations, the equivalence of which we'll prove. In its first formulation it says something fundamental about the relationship between semantic consequence and our derivation system: if a sentence φ follows from some sentences Γ, then there is also a derivation that establishes $\Gamma \vdash \varphi$. Thus, the derivation system is as strong as it can possibly be without proving things that don't actually follow.

In its second formulation, it can be stated as a model existence result: every consistent set of sentences is satisfiable. Consistency is a proof-theoretic notion: it says that our derivation system is unable to produce certain derivations. But who's to say that just because there are no derivations of a certain sort from Γ, it's guaranteed that there is a structure $\mathfrak{M}$? Before the completeness theorem was first proved—in fact before we had the derivation systems we now do—the great German mathematician David Hilbert held the view that consistency of mathematical theories guarantees the existence of the objects they are about. He put it as follows in a letter to Gottlob Frege:

> If the arbitrarily given axioms do not contradict one another with all their consequences, then they are true and the things defined by the axioms exist. This is for me the criterion of truth and existence.

Frege vehemently disagreed. The second formulation of the completeness theorem shows that Hilbert was right in at least the sense that if the axioms are consistent, then *some* structure exists that makes them all true.

These aren't the only reasons the completeness theorem—or rather, its proof—is important. It has a number of important consequences, some of which we'll discuss separately. For instance, since any derivation that shows $\Gamma \vdash \varphi$ is finite and so can only use finitely many of the sentences in Γ, it follows by the completeness theorem that if φ is a consequence of Γ, it is already

a consequence of a finite subset of Γ. This is called *compactness*. Equivalently, if every finite subset of Γ is consistent, then Γ itself must be consistent.

Although the compactness theorem follows from the completeness theorem via the detour through derivations, it is also possible to use the *the proof of* the completeness theorem to establish it directly. For what the proof does is take a set of sentences with a certain property—consistency—and constructs a structure out of this set that has certain properties (in this case, that it satisfies the set). Almost the very same construction can be used to directly establish compactness, by starting from "finitely satisfiable" sets of sentences instead of consistent ones. The construction also yields other consequences, e.g., that any satisfiable set of sentences has a finite or denumerable model. (This result is called the Löwenheim-Skolem theorem.) In general, the construction of structures from sets of sentences is used often in logic, and sometimes even in philosophy.

8.2 Outline of the Proof

The proof of the completeness theorem is a bit complex, and upon first reading it, it is easy to get lost. So let us outline the proof. The first step is a shift of perspective, that allows us to see a route to a proof. When completeness is thought of as "whenever $\Gamma \vDash \varphi$ then $\Gamma \vdash \varphi$," it may be hard to even come up with an idea: for to show that $\Gamma \vdash \varphi$ we have to find a derivation, and it does not look like the hypothesis that $\Gamma \vDash \varphi$ helps us for this in any way. For some proof systems it is possible to directly construct a derivation, but we will take a slightly different approach. The shift in perspective required is this: completeness can also be formulated as: "if Γ is consistent, it is satisfiable." Perhaps we can use the information in Γ together with the hypothesis that it is consistent to construct a structure that satisfies every sentence in Γ. After all, we know what kind of structure we are looking for: one that is as Γ describes it!

If Γ contains only atomic sentences, it is easy to construct a model for it. Suppose the atomic sentences are all of the form $P(a_1, \ldots, a_n)$ where the a_i are constant symbols. All we have to do is come up with a domain $|\mathfrak{M}|$ and an assignment for P so that $\mathfrak{M} \vDash P(a_1, \ldots, a_n)$. But that's not very hard: put $|\mathfrak{M}| = \mathbb{N}$, $c_i^{\mathfrak{M}} = i$, and for every $P(a_1, \ldots, a_n) \in \Gamma$, put the tuple $\langle k_1, \ldots, k_n \rangle$ into $P^{\mathfrak{M}}$, where k_i is the index of the constant symbol a_i (i.e., $a_i \equiv c_{k_i}$).

Now suppose Γ contains some formula $\neg\psi$, with ψ atomic. We might worry that the construction of $\mathfrak{M}$ interferes with the possibility of making $\neg\psi$ true. But here's where the consistency of Γ comes in: if $\neg\psi \in \Gamma$, then $\psi \notin \Gamma$, or else Γ would be inconsistent. And if $\psi \notin \Gamma$, then according to our construction of $\mathfrak{M}$, $\mathfrak{M} \nvDash \psi$, so $\mathfrak{M} \vDash \neg\psi$. So far so good.

What if Γ contains complex, non-atomic formulas? Say it contains $\varphi \wedge \psi$. To make that true, we should proceed as if both φ and ψ were in Γ. And if

$\varphi \vee \psi \in \Gamma$, then we will have to make at least one of them true, i.e., proceed as if one of them was in Γ.

This suggests the following idea: we add additional formulas to Γ so as to (a) keep the resulting set consistent and (b) make sure that for every possible atomic sentence φ, either φ is in the resulting set, or $\neg\varphi$ is, and (c) such that, whenever $\varphi \wedge \psi$ is in the set, so are both φ and ψ, if $\varphi \vee \psi$ is in the set, at least one of φ or ψ is also, etc. We keep doing this (potentially forever). Call the set of all formulas so added Γ^*. Then our construction above would provide us with a structure $\mathfrak{M}$ for which we could prove, by induction, that it satisfies all sentences in Γ^*, and hence also all sentence in Γ since $\Gamma \subseteq \Gamma^*$. It turns out that guaranteeing (a) and (b) is enough. A set of sentences for which (b) holds is called *complete*. So our task will be to extend the consistent set Γ to a consistent and complete set Γ^*.

There is one wrinkle in this plan: if $\exists x\, \varphi(x) \in \Gamma$ we would hope to be able to pick some constant symbol c and add $\varphi(c)$ in this process. But how do we know we can always do that? Perhaps we only have a few constant symbols in our language, and for each one of them we have $\neg\varphi(c) \in \Gamma$. We can't also add $\varphi(c)$, since this would make the set inconsistent, and we wouldn't know whether $\mathfrak{M}$ has to make $\varphi(c)$ or $\neg\varphi(c)$ true. Moreover, it might happen that Γ contains only sentences in a language that has no constant symbols at all (e.g., the language of set theory).

The solution to this problem is to simply add infinitely many constants at the beginning, plus sentences that connect them with the quantifiers in the right way. (Of course, we have to verify that this cannot introduce an inconsistency.)

Our original construction works well if we only have constant symbols in the atomic sentences. But the language might also contain function symbols. In that case, it might be tricky to find the right functions on $\mathbb{N}$ to assign to these function symbols to make everything work. So here's another trick: instead of using i to interpret c_i, just take the set of constant symbols itself as the domain. Then $\mathfrak{M}$ can assign every constant symbol to itself: $c_i^{\mathfrak{M}} = c_i$. But why not go all the way: let $|\mathfrak{M}|$ be all *terms* of the language! If we do this, there is an obvious assignment of functions (that take terms as arguments and have terms as values) to function symbols: we assign to the function symbol f_i^n the function which, given n terms $t_1, \ldots, t_n$ as input, produces the term $f_i^n(t_1, \ldots, t_n)$ as value.

The last piece of the puzzle is what to do with $=$. The predicate symbol $=$ has a fixed interpretation: $\mathfrak{M} \vDash t = t'$ iff $\mathrm{Val}^{\mathfrak{M}}(t) = \mathrm{Val}^{\mathfrak{M}}(t')$. Now if we set things up so that the value of a term t is t itself, then this structure will make *no* sentence of the form $t = t'$ true unless t and t' are one and the same term. And of course this is a problem, since basically every interesting theory in a language with function symbols will have as theorems sentences $t = t'$ where t and t' are not the same term (e.g., in theories of arithmetic: $(\mathsf{0} + \mathsf{0}) = \mathsf{0}$). To

solve this problem, we change the domain of $\mathfrak{M}$: instead of using terms as the objects in $|\mathfrak{M}|$, we use sets of terms, and each set is so that it contains all those terms which the sentences in Γ require to be equal. So, e.g., if Γ is a theory of arithmetic, one of these sets will contain: o, $(\mathsf{o}+\mathsf{o})$, $(\mathsf{o}\times\mathsf{o})$, etc. This will be the set we assign to o, and it will turn out that this set is also the value of all the terms in it, e.g., also of $(\mathsf{o}+\mathsf{o})$. Therefore, the sentence $(\mathsf{o}+\mathsf{o})=\mathsf{o}$ will be true in this revised structure.

So here's what we'll do. First we investigate the properties of complete consistent sets, in particular we prove that a complete consistent set contains $\varphi\wedge\psi$ iff it contains both φ and ψ, $\varphi\vee\psi$ iff it contains at least one of them, etc. (Proposition 8.2). Then we define and investigate "saturated" sets of sentences. A saturated set is one which contains conditionals that link each quantified sentence to instances of it (Definition 8.5). We show that any consistent set Γ can always be extended to a saturated set Γ' (Lemma 8.6). If a set is consistent, saturated, and complete it also has the property that it contains $\exists x\,\varphi(x)$ iff it contains $\varphi(t)$ for some closed term t and $\forall x\,\varphi(x)$ iff it contains $\varphi(t)$ for all closed terms t (Proposition 8.7). We'll then take the saturated consistent set Γ' and show that it can be extended to a saturated, consistent, and complete set Γ^* (Lemma 8.8). This set Γ^* is what we'll use to define our term model $\mathfrak{M}(\Gamma^*)$. The term model has the set of closed terms as its domain, and the interpretation of its predicate symbols is given by the atomic sentences in Γ^* (Definition 8.9). We'll use the properties of saturated, complete consistent sets to show that indeed $\mathfrak{M}(\Gamma^*)\vDash\varphi$ iff $\varphi\in\Gamma^*$ (Lemma 8.12), and thus in particular, $\mathfrak{M}(\Gamma^*)\vDash\Gamma$. Finally, we'll consider how to define a term model if Γ contains $=$ as well (Definition 8.16) and show that it satisfies Γ^* (Lemma 8.19).

8.3 Complete Consistent Sets of Sentences

Definition 8.1 (Complete set). A set Γ of sentences is *complete* iff for any sentence φ, either $\varphi\in\Gamma$ or $\neg\varphi\in\Gamma$.

Complete sets of sentences leave no questions unanswered. For any sentence φ, Γ "says" if φ is true or false. The importance of complete sets extends beyond the proof of the completeness theorem. A theory which is complete and axiomatizable, for instance, is always decidable.

Complete consistent sets are important in the completeness proof since we can guarantee that every consistent set of sentences Γ is contained in a complete consistent set Γ^*. A complete consistent set contains, for each sentence φ, either φ or its negation $\neg\varphi$, but not both. This is true in particular for atomic sentences, so from a complete consistent set in a language suitably expanded by constant symbols, we can construct a structure where the interpretation of predicate symbols is defined according to which atomic sentences are in Γ^*. This structure can then be shown to make all sentences in Γ^* (and hence also

all those in Γ) true. The proof of this latter fact requires that $\neg\varphi \in \Gamma^*$ iff $\varphi \notin \Gamma^*$, $(\varphi \vee \psi) \in \Gamma^*$ iff $\varphi \in \Gamma^*$ or $\psi \in \Gamma^*$, etc.

In what follows, we will often tacitly use the properties of reflexivity, monotonicity, and transitivity of $\vdash$ (see section 7.8).

Proposition 8.2. *Suppose Γ is complete and consistent. Then:*

1. *If $\Gamma \vdash \varphi$, then $\varphi \in \Gamma$.*
2. *$\varphi \wedge \psi \in \Gamma$ iff both $\varphi \in \Gamma$ and $\psi \in \Gamma$.*
3. *$\varphi \vee \psi \in \Gamma$ iff either $\varphi \in \Gamma$ or $\psi \in \Gamma$.*
4. *$\varphi \to \psi \in \Gamma$ iff either $\varphi \notin \Gamma$ or $\psi \in \Gamma$.*

Proof. Let us suppose for all of the following that Γ is complete and consistent.

1. If $\Gamma \vdash \varphi$, then $\varphi \in \Gamma$.

   Suppose that $\Gamma \vdash \varphi$. Suppose to the contrary that $\varphi \notin \Gamma$. Since Γ is complete, $\neg\varphi \in \Gamma$. By Proposition 7.20, Γ is inconsistent. This contradicts the assumption that Γ is consistent. Hence, it cannot be the case that $\varphi \notin \Gamma$, so $\varphi \in \Gamma$.

2. $\varphi \wedge \psi \in \Gamma$ iff both $\varphi \in \Gamma$ and $\psi \in \Gamma$:

   For the forward direction, suppose $\varphi \wedge \psi \in \Gamma$. Then by Proposition 7.22, item (1), $\Gamma \vdash \varphi$ and $\Gamma \vdash \psi$. By (1), $\varphi \in \Gamma$ and $\psi \in \Gamma$, as required.

   For the reverse direction, let $\varphi \in \Gamma$ and $\psi \in \Gamma$. By Proposition 7.22, item (2), $\Gamma \vdash \varphi \wedge \psi$. By (1), $\varphi \wedge \psi \in \Gamma$.

3. First we show that if $\varphi \vee \psi \in \Gamma$, then either $\varphi \in \Gamma$ or $\psi \in \Gamma$. Suppose $\varphi \vee \psi \in \Gamma$ but $\varphi \notin \Gamma$ and $\psi \notin \Gamma$. Since Γ is complete, $\neg\varphi \in \Gamma$ and $\neg\psi \in \Gamma$. By Proposition 7.23, item (1), Γ is inconsistent, a contradiction. Hence, either $\varphi \in \Gamma$ or $\psi \in \Gamma$.

   For the reverse direction, suppose that $\varphi \in \Gamma$ or $\psi \in \Gamma$. By Proposition 7.23, item (2), $\Gamma \vdash \varphi \vee \psi$. By (1), $\varphi \vee \psi \in \Gamma$, as required.

4. Exercise. □

8.4 Henkin Expansion

Part of the challenge in proving the completeness theorem is that the model we construct from a complete consistent set Γ must make all the quantified formulas in Γ true. In order to guarantee this, we use a trick due to Leon Henkin. In essence, the trick consists in expanding the language by infinitely many constant symbols and adding, for each formula with one free variable

$\varphi(x)$ a formula of the form $\exists x\, \varphi(x) \to \varphi(c)$, where c is one of the new constant symbols. When we construct the structure satisfying Γ, this will guarantee that each true existential sentence has a witness among the new constants.

Proposition 8.3. *If Γ is consistent in $\mathcal{L}$ and $\mathcal{L}'$ is obtained from $\mathcal{L}$ by adding a denumerable set of new constant symbols d_0, d_1, ..., then Γ is consistent in $\mathcal{L}'$.*

Definition 8.4 (Saturated set). A set Γ of formulas of a language $\mathcal{L}$ is *saturated* iff for each formula $\varphi(x) \in \mathrm{Frm}(\mathcal{L})$ with one free variable x there is a constant symbol $c \in \mathcal{L}$ such that $\exists x\, \varphi(x) \to \varphi(c) \in \Gamma$.

The following definition will be used in the proof of the next theorem.

Definition 8.5. Let $\mathcal{L}'$ be as in Proposition 8.3. Fix an enumeration $\varphi_0(x_0)$, $\varphi_1(x_1)$, ...of all formulas $\varphi_i(x_i)$ of $\mathcal{L}'$ in which one variable (x_i) occurs free. We define the sentences θ_n by induction on n.

Let c_0 be the first constant symbol among the d_i we added to $\mathcal{L}$ which does not occur in $\varphi_0(x_0)$. Assuming that θ_0, ..., θ_{n-1} have already been defined, let c_n be the first among the new constant symbols d_i that occurs neither in θ_0, ..., θ_{n-1} nor in $\varphi_n(x_n)$.

Now let θ_n be the formula $\exists x_n\, \varphi_n(x_n) \to \varphi_n(c_n)$.

Lemma 8.6. *Every consistent set Γ can be extended to a saturated consistent set Γ'.*

Proof. Given a consistent set of sentences Γ in a language $\mathcal{L}$, expand the language by adding a denumerable set of new constant symbols to form $\mathcal{L}'$. By Proposition 8.3, Γ is still consistent in the richer language. Further, let θ_i be as in Definition 8.5. Let

$$\Gamma_0 = \Gamma$$
$$\Gamma_{n+1} = \Gamma_n \cup \{\theta_n\}$$

i.e., $\Gamma_{n+1} = \Gamma \cup \{\theta_0, \ldots, \theta_n\}$, and let $\Gamma' = \bigcup_n \Gamma_n$. Γ' is clearly saturated.

If Γ' were inconsistent, then for some n, Γ_n would be inconsistent (Exercise: explain why). So to show that Γ' is consistent it suffices to show, by induction on n, that each set Γ_n is consistent.

The induction basis is simply the claim that $\Gamma_0 = \Gamma$ is consistent, which is the hypothesis of the theorem. For the induction step, suppose that Γ_n is consistent but $\Gamma_{n+1} = \Gamma_n \cup \{\theta_n\}$ is inconsistent. Recall that θ_n is $\exists x_n\, \varphi_n(x_n) \to \varphi_n(c_n)$, where $\varphi_n(x_n)$ is a formula of $\mathcal{L}'$ with only the variable x_n free. By the way we've chosen the c_n (see Definition 8.5), c_n does not occur in $\varphi_n(x_n)$ nor in Γ_n.

If $\Gamma_n \cup \{\theta_n\}$ is inconsistent, then $\Gamma_n \vdash \neg\theta_n$, and hence both of the following hold:

$$\Gamma_n \vdash \exists x_n\, \varphi_n(x_n) \qquad \Gamma_n \vdash \neg\varphi_n(c_n)$$

Since c_n does not occur in Γ_n or in $\varphi_n(x_n)$, Theorem 7.25 applies. From $\Gamma_n \vdash \neg\varphi_n(c_n)$, we obtain $\Gamma_n \vdash \forall x_n\, \neg\varphi_n(x_n)$. Thus we have that both $\Gamma_n \vdash \exists x_n\, \varphi_n(x_n)$ and $\Gamma_n \vdash \forall x_n\, \neg\varphi_n(x_n)$, so Γ_n itself is inconsistent. (Note that $\forall x_n\, \neg\varphi_n(x_n) \vdash \neg\exists x_n\, \varphi_n(x_n)$.) Contradiction: Γ_n was supposed to be consistent. Hence $\Gamma_n \cup \{\theta_n\}$ is consistent. □

We'll now show that *complete*, consistent sets which are saturated have the property that it contains a universally quantified sentence iff it contains all its instances and it contains an existentially quantified sentence iff it contains at least one instance. We'll use this to show that the structure we'll generate from a complete, consistent, saturated set makes all its quantified sentences true.

Proposition 8.7. *Suppose Γ is complete, consistent, and saturated.*

1. *$\exists x\, \varphi(x) \in \Gamma$ iff $\varphi(t) \in \Gamma$ for at least one closed term t.*
2. *$\forall x\, \varphi(x) \in \Gamma$ iff $\varphi(t) \in \Gamma$ for all closed terms t.*

Proof. 1. First suppose that $\exists x\, \varphi(x) \in \Gamma$. Because Γ is saturated, $(\exists x\, \varphi(x) \to \varphi(c)) \in \Gamma$ for some constant symbol c. By Proposition 7.24, item (1), and Proposition 8.2(1), $\varphi(c) \in \Gamma$.

   For the other direction, saturation is not necessary: Suppose $\varphi(t) \in \Gamma$. Then $\Gamma \vdash \exists x\, \varphi(x)$ by Proposition 7.26, item (1). By Proposition 8.2(1), $\exists x\, \varphi(x) \in \Gamma$.

2. Exercise. □

8.5 Lindenbaum's Lemma

We now prove a lemma that shows that any consistent set of sentences is contained in some set of sentences which is not just consistent, but also complete. The proof works by adding one sentence at a time, guaranteeing at each step that the set remains consistent. We do this so that for every φ, either φ or $\neg\varphi$ gets added at some stage. The union of all stages in that construction then contains either φ or its negation $\neg\varphi$ and is thus complete. It is also consistent, since we made sure at each stage not to introduce an inconsistency.

Lemma 8.8 (Lindenbaum's Lemma). *Every consistent set Γ in a language $\mathcal{L}$ can be extended to a complete and consistent set Γ^*.*

Proof. Let Γ be consistent. Let φ_0, φ_1, ... be an enumeration of all the sentences of $\mathcal{L}$. Define $\Gamma_0 = \Gamma$, and

$$\Gamma_{n+1} = \begin{cases} \Gamma_n \cup \{\varphi_n\} & \text{if } \Gamma_n \cup \{\varphi_n\} \text{ is consistent;} \\ \Gamma_n \cup \{\neg\varphi_n\} & \text{otherwise.} \end{cases}$$

Let $\Gamma^* = \bigcup_{n \geq 0} \Gamma_n$.

Each Γ_n is consistent: Γ_0 is consistent by definition. If $\Gamma_{n+1} = \Gamma_n \cup \{\varphi_n\}$, this is because the latter is consistent. If it isn't, $\Gamma_{n+1} = \Gamma_n \cup \{\neg\varphi_n\}$. We have to verify that $\Gamma_n \cup \{\neg\varphi_n\}$ is consistent. Suppose it's not. Then *both* $\Gamma_n \cup \{\varphi_n\}$ and $\Gamma_n \cup \{\neg\varphi_n\}$ are inconsistent. This means that Γ_n would be inconsistent by Proposition 7.21, contrary to the induction hypothesis.

For every n and every $i < n$, $\Gamma_i \subseteq \Gamma_n$. This follows by a simple induction on n. For $n = 0$, there are no $i < 0$, so the claim holds automatically. For the inductive step, suppose it is true for n. We have $\Gamma_{n+1} = \Gamma_n \cup \{\varphi_n\}$ or $= \Gamma_n \cup \{\neg\varphi_n\}$ by construction. So $\Gamma_n \subseteq \Gamma_{n+1}$. If $i < n$, then $\Gamma_i \subseteq \Gamma_n$ by inductive hypothesis, and so $\subseteq \Gamma_{n+1}$ by transitivity of $\subseteq$.

From this it follows that every finite subset of Γ^* is a subset of Γ_n for some n, since each $\psi \in \Gamma^*$ not already in Γ_0 is added at some stage i. If n is the last one of these, then all ψ in the finite subset are in Γ_n. So, every finite subset of Γ^* is consistent. By Proposition 7.17, Γ^* is consistent.

Every sentence of $\mathrm{Frm}(\mathcal{L})$ appears on the list used to define Γ^*. If $\varphi_n \notin \Gamma^*$, then that is because $\Gamma_n \cup \{\varphi_n\}$ was inconsistent. But then $\neg\varphi_n \in \Gamma^*$, so Γ^* is complete. □

8.6 Construction of a Model

Right now we are not concerned about $=$, i.e., we only want to show that a consistent set Γ of sentences not containing $=$ is satisfiable. We first extend Γ to a consistent, complete, and saturated set Γ^*. In this case, the definition of a model $\mathfrak{M}(\Gamma^*)$ is simple: We take the set of closed terms of $\mathcal{L}'$ as the domain. We assign every constant symbol to itself, and make sure that more generally, for every closed term t, $\mathrm{Val}^{\mathfrak{M}(\Gamma^*)}(t) = t$. The predicate symbols are assigned extensions in such a way that an atomic sentence is true in $\mathfrak{M}(\Gamma^*)$ iff it is in Γ^*. This will obviously make all the atomic sentences in Γ^* true in $\mathfrak{M}(\Gamma^*)$. The rest are true provided the Γ^* we start with is consistent, complete, and saturated.

Definition 8.9 (Term model). Let Γ^* be a complete and consistent, saturated set of sentences in a language $\mathcal{L}$. The *term model* $\mathfrak{M}(\Gamma^*)$ of Γ^* is the structure defined as follows:

1. The domain $|\mathfrak{M}(\Gamma^*)|$ is the set of all closed terms of $\mathcal{L}$.

2. The interpretation of a constant symbol c is c itself: $c^{\mathfrak{M}(\Gamma^*)} = c$.

3. The function symbol f is assigned the function which, given as arguments the closed terms $t_1, \ldots, t_n$, has as value the closed term $f(t_1, \ldots, t_n)$:

$$f^{\mathfrak{M}(\Gamma^*)}(t_1, \ldots, t_n) = f(t_1, \ldots, t_n)$$

4. If R is an n-place predicate symbol, then

$$\langle t_1, \ldots, t_n \rangle \in R^{\mathfrak{M}(\Gamma^*)} \text{ iff } R(t_1, \ldots, t_n) \in \Gamma^*.$$

We will now check that we indeed have $\mathrm{Val}^{\mathfrak{M}(\Gamma^*)}(t) = t$.

Lemma 8.10. *Let $\mathfrak{M}(\Gamma^*)$ be the term model of Definition 8.9, then $\mathrm{Val}^{\mathfrak{M}(\Gamma^*)}(t) = t$.*

Proof. The proof is by induction on t, where the base case, when t is a constant symbol, follows directly from the definition of the term model. For the induction step assume $t_1, \ldots, t_n$ are closed terms such that $\mathrm{Val}^{\mathfrak{M}(\Gamma^*)}(t_i) = t_i$ and that f is an n-ary function symbol. Then

$$\begin{aligned}
\mathrm{Val}^{\mathfrak{M}(\Gamma^*)}(f(t_1, \ldots, t_n)) &= f^{\mathfrak{M}(\Gamma^*)}(\mathrm{Val}^{\mathfrak{M}(\Gamma^*)}(t_1), \ldots, \mathrm{Val}^{\mathfrak{M}(\Gamma^*)}(t_n)) \\
&= f^{\mathfrak{M}(\Gamma^*)}(t_1, \ldots, t_n) \\
&= f(t_1, \ldots, t_n),
\end{aligned}$$

and so by induction this holds for every closed term t. □

A structure $\mathfrak{M}$ may make an existentially quantified sentence $\exists x\, \varphi(x)$ true without there being an instance $\varphi(t)$ that it makes true. A structure $\mathfrak{M}$ may make all instances $\varphi(t)$ of a universally quantified sentence $\forall x\, \varphi(x)$ true, without making $\forall x\, \varphi(x)$ true. This is because in general not every element of $|\mathfrak{M}|$ is the value of a closed term ($\mathfrak{M}$ may not be covered). This is the reason the satisfaction relation is defined via variable assignments. However, for our term model $\mathfrak{M}(\Gamma^*)$ this wouldn't be necessary—because it is covered. This is the content of the next result.

Proposition 8.11. *Let $\mathfrak{M}(\Gamma^*)$ be the term model of Definition 8.9.*

1. *$\mathfrak{M}(\Gamma^*) \vDash \exists x\, \varphi(x)$ iff $\mathfrak{M}(\Gamma^*) \vDash \varphi(t)$ for at least one term t.*
2. *$\mathfrak{M}(\Gamma^*) \vDash \forall x\, \varphi(x)$ iff $\mathfrak{M}(\Gamma^*) \vDash \varphi(t)$ for all terms t.*

Proof. 1. By Proposition 5.59, $\mathfrak{M}(\Gamma^*) \vDash \exists x\, \varphi(x)$ iff for at least one variable assignment s, $\mathfrak{M}(\Gamma^*), s \vDash \varphi(x)$. As $|\mathfrak{M}(\Gamma^*)|$ consists of the closed terms of $\mathcal{L}$, this is the case iff there is at least one closed term t such that $s(x) = t$ and $\mathfrak{M}(\Gamma^*), s \vDash \varphi(x)$. By Proposition 5.63, $\mathfrak{M}(\Gamma^*), s \vDash \varphi(x)$ iff $\mathfrak{M}(\Gamma^*), s \vDash \varphi(t)$, where $s(x) = t$. By Proposition 5.58, $\mathfrak{M}(\Gamma^*), s \vDash \varphi(t)$ iff $\mathfrak{M}(\Gamma^*) \vDash \varphi(t)$, since $\varphi(t)$ is a sentence.

2. Exercise. □

Lemma 8.12 (Truth Lemma). *Suppose φ does not contain $=$. Then $\mathfrak{M}(\Gamma^*) \vDash \varphi$ iff $\varphi \in \Gamma^*$.*

Proof. We prove both directions simultaneously, and by induction on φ.

1. $\varphi \equiv \bot$: $\mathfrak{M}(\Gamma^*) \nvDash \bot$ by definition of satisfaction. On the other hand, $\bot \notin \Gamma^*$ since Γ^* is consistent.

2. $\varphi \equiv R(t_1, \ldots, t_n)$: $\mathfrak{M}(\Gamma^*) \vDash R(t_1, \ldots, t_n)$ iff $\langle t_1, \ldots, t_n \rangle \in R^{\mathfrak{M}(\Gamma^*)}$ (by the definition of satisfaction) iff $R(t_1, \ldots, t_n) \in \Gamma^*$ (by the construction of $\mathfrak{M}(\Gamma^*)$).

3. $\varphi \equiv \neg\psi$: $\mathfrak{M}(\Gamma^*) \vDash \varphi$ iff $\mathfrak{M}(\Gamma^*) \nvDash \psi$ (by definition of satisfaction). By induction hypothesis, $\mathfrak{M}(\Gamma^*) \nvDash \psi$ iff $\psi \notin \Gamma^*$. Since Γ^* is consistent and complete, $\psi \notin \Gamma^*$ iff $\neg\psi \in \Gamma^*$.

4. $\varphi \equiv \psi \wedge \chi$: $\mathfrak{M}(\Gamma^*) \vDash \varphi$ iff we have both $\mathfrak{M}(\Gamma^*) \vDash \psi$ and $\mathfrak{M}(\Gamma^*) \vDash \chi$ (by definition of satisfaction) iff both $\psi \in \Gamma^*$ and $\chi \in \Gamma^*$ (by the induction hypothesis). By Proposition 8.2(2), this is the case iff $(\psi \wedge \chi) \in \Gamma^*$.

5. $\varphi \equiv \psi \vee \chi$: $\mathfrak{M}(\Gamma^*) \vDash \varphi$ iff $\mathfrak{M}(\Gamma^*) \vDash \psi$ or $\mathfrak{M}(\Gamma^*) \vDash \chi$ (by definition of satisfaction) iff $\psi \in \Gamma^*$ or $\chi \in \Gamma^*$ (by induction hypothesis). This is the case iff $(\psi \vee \chi) \in \Gamma^*$ (by Proposition 8.2(3)).

6. $\varphi \equiv \psi \to \chi$: exercise.

7. $\varphi \equiv \forall x\, \psi(x)$: exercise.

8. $\varphi \equiv \exists x\, \psi(x)$: $\mathfrak{M}(\Gamma^*) \vDash \varphi$ iff $\mathfrak{M}(\Gamma^*) \vDash \psi(t)$ for at least one term t (Proposition 8.11). By induction hypothesis, this is the case iff $\psi(t) \in \Gamma^*$ for at least one term t. By Proposition 8.7, this in turn is the case iff $\exists x\, \psi(x) \in \Gamma^*$. □

8.7 Identity

The construction of the term model given in the preceding section is enough to establish completeness for first-order logic for sets Γ that do not contain $=$. The term model satisfies every $\varphi \in \Gamma^*$ which does not contain $=$ (and hence all $\varphi \in \Gamma$). It does not work, however, if $=$ is present. The reason is that Γ^* then may contain a sentence $t = t'$, but in the term model the value of any term is that term itself. Hence, if t and t' are different terms, their values in the term model—i.e., t and t', respectively—are different, and so $t = t'$ is false. We can fix this, however, using a construction known as "factoring."

Definition 8.13. Let Γ^* be a consistent and complete set of sentences in $\mathcal{L}$. We define the relation $\approx$ on the set of closed terms of $\mathcal{L}$ by

$$t \approx t' \quad \text{iff} \quad t = t' \in \Gamma^*$$

Proposition 8.14. *The relation $\approx$ has the following properties:*

1. $\approx$ *is reflexive.*
2. $\approx$ *is symmetric.*
3. $\approx$ *is transitive.*
4. *If* $t \approx t'$, f *is a function symbol, and* $t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n$ *are terms, then*

$$f(t_1, \ldots, t_{i-1}, t, t_{i+1}, \ldots, t_n) \approx f(t_1, \ldots, t_{i-1}, t', t_{i+1}, \ldots, t_n).$$

5. *If* $t \approx t'$, R *is a predicate symbol, and* $t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n$ *are terms, then*

$$R(t_1, \ldots, t_{i-1}, t, t_{i+1}, \ldots, t_n) \in \Gamma^* \textit{ iff } R(t_1, \ldots, t_{i-1}, t', t_{i+1}, \ldots, t_n) \in \Gamma^*.$$

Proof. Since Γ^* is consistent and complete, $t = t' \in \Gamma^*$ iff $\Gamma^* \vdash t = t'$. Thus it is enough to show the following:

1. $\Gamma^* \vdash t = t$ for all terms t.
2. If $\Gamma^* \vdash t = t'$ then $\Gamma^* \vdash t' = t$.
3. If $\Gamma^* \vdash t = t'$ and $\Gamma^* \vdash t' = t''$, then $\Gamma^* \vdash t = t''$.
4. If $\Gamma^* \vdash t = t'$, then

$$\Gamma^* \vdash f(t_1, \ldots, t_{i-1}, t, t_{i+1}, , \ldots, t_n) = f(t_1, \ldots, t_{i-1}, t', t_{i+1}, \ldots, t_n)$$

   for every n-place function symbol f and terms $t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n$.
5. If $\Gamma^* \vdash t = t'$ and $\Gamma^* \vdash R(t_1, \ldots, t_{i-1}, t, t_{i+1}, \ldots, t_n)$, then $\Gamma^* \vdash R(t_1, \ldots, t_{i-1}, t', t_{i+1}, \ldots, t_n)$ for every n-place predicate symbol R and terms $t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n$.

□

Definition 8.15. Suppose Γ^* is a consistent and complete set in a language $\mathcal{L}$, t is a term, and $\approx$ as in the previous definition. Then:

$$[t]_\approx = \{t' : t' \in \mathrm{Trm}(\mathcal{L}), t \approx t'\}$$

and $\mathrm{Trm}(\mathcal{L})/_\approx = \{[t]_\approx : t \in \mathrm{Trm}(\mathcal{L})\}$.

Definition 8.16. Let $\mathfrak{M} = \mathfrak{M}(\Gamma^*)$ be the term model for Γ^* from Definition 8.9. Then $\mathfrak{M}/_\approx$ is the following structure:

1. $|\mathfrak{M}/_\approx| = \mathrm{Trm}(\mathcal{L})/_\approx$.
2. $c^{\mathfrak{M}/_\approx} = [c]_\approx$
3. $f^{\mathfrak{M}/_\approx}([t_1]_\approx, \ldots, [t_n]_\approx) = [f(t_1, \ldots, t_n)]_\approx$

4. $\langle [t_1]_\approx, \ldots, [t_n]_\approx \rangle \in R^{\mathfrak{M}/\approx}$ iff $\mathfrak{M} \vDash R(t_1, \ldots, t_n)$, i.e., iff $R(t_1, \ldots, t_n) \in \Gamma^*$.

Note that we have defined $f^{\mathfrak{M}/\approx}$ and $R^{\mathfrak{M}/\approx}$ for elements of $\mathrm{Trm}(\mathcal{L})/\approx$ by referring to them as $[t]_\approx$, i.e., via *representatives* $t \in [t]_\approx$. We have to make sure that these definitions do not depend on the choice of these representatives, i.e., that for some other choices t' which determine the same equivalence classes ($[t]_\approx = [t']_\approx$), the definitions yield the same result. For instance, if R is a one-place predicate symbol, the last clause of the definition says that $[t]_\approx \in R^{\mathfrak{M}/\approx}$ iff $\mathfrak{M} \vDash R(t)$. If for some other term t' with $t \approx t'$, $\mathfrak{M} \nvDash R(t)$, then the definition would require $[t']_\approx \notin R^{\mathfrak{M}/\approx}$. If $t \approx t'$, then $[t]_\approx = [t']_\approx$, but we can't have both $[t]_\approx \in R^{\mathfrak{M}/\approx}$ and $[t]_\approx \notin R^{\mathfrak{M}/\approx}$. However, Proposition 8.14 guarantees that this cannot happen.

Proposition 8.17. *$\mathfrak{M}/\approx$ is well defined, i.e., if $t_1, \ldots, t_n$, $t'_1, \ldots, t'_n$ are terms, and $t_i \approx t'_i$ then*

1. $[f(t_1, \ldots, t_n)]_\approx = [f(t'_1, \ldots, t'_n)]_\approx$, *i.e.,*
$$f(t_1, \ldots, t_n) \approx f(t'_1, \ldots, t'_n)$$
and
2. $\mathfrak{M} \vDash R(t_1, \ldots, t_n)$ *iff* $\mathfrak{M} \vDash R(t'_1, \ldots, t'_n)$, *i.e.,*
$$R(t_1, \ldots, t_n) \in \Gamma^* \text{ iff } R(t'_1, \ldots, t'_n) \in \Gamma^*.$$

Proof. Follows from Proposition 8.14 by induction on n. □

As in the case of the term model, before proving the truth lemma we need the following lemma.

Lemma 8.18. *Let $\mathfrak{M} = \mathfrak{M}(\Gamma^*)$, then $\mathrm{Val}^{\mathfrak{M}/\approx}(t) = [t]_\approx$.*

Proof. The proof is similar to that of Lemma 8.10. □

Lemma 8.19. *$\mathfrak{M}/\approx \vDash \varphi$ iff $\varphi \in \Gamma^*$ for all sentences φ.*

Proof. By induction on φ, just as in the proof of Lemma 8.12. The only case that needs additional attention is when $\varphi \equiv t = t'$.

$$\begin{aligned}\mathfrak{M}/\approx \vDash t = t' &\text{ iff } [t]_\approx = [t']_\approx \text{ (by definition of } \mathfrak{M}/\approx)\\ &\text{iff } t \approx t' \text{ (by definition of } [t]_\approx)\\ &\text{iff } t = t' \in \Gamma^* \text{ (by definition of } \approx).\end{aligned}$$

□

Note that while $\mathfrak{M}(\Gamma^*)$ is always enumerable and infinite, $\mathfrak{M}/\approx$ may be finite, since it may turn out that there are only finitely many classes $[t]_\approx$. This is to be expected, since Γ may contain sentences which require any structure in which they are true to be finite. For instance, $\forall x \forall y\, x = y$ is a consistent sentence, but is satisfied only in structures with a domain that contains exactly one element.

8.8 The Completeness Theorem

Let's combine our results: we arrive at the completeness theorem.

Theorem 8.20 (Completeness Theorem). *Let Γ be a set of sentences. If Γ is consistent, it is satisfiable.*

Proof. Suppose Γ is consistent. By Lemma 8.6, there is a saturated consistent set $\Gamma' \supseteq \Gamma$. By Lemma 8.8, there is a $\Gamma^* \supseteq \Gamma'$ which is consistent and complete. Since $\Gamma' \subseteq \Gamma^*$, for each formula $\varphi(x)$, Γ^* contains a sentence of the form $\exists x\, \varphi(x) \to \varphi(c)$ and so Γ^* is saturated. If Γ does not contain $=$, then by Lemma 8.12, $\mathfrak{M}(\Gamma^*) \vDash \varphi$ iff $\varphi \in \Gamma^*$. From this it follows in particular that for all $\varphi \in \Gamma$, $\mathfrak{M}(\Gamma^*) \vDash \varphi$, so Γ is satisfiable. If Γ does contain $=$, then by Lemma 8.19, for all sentences φ, $\mathfrak{M}/_{\approx} \vDash \varphi$ iff $\varphi \in \Gamma^*$. In particular, $\mathfrak{M}/_{\approx} \vDash \varphi$ for all $\varphi \in \Gamma$, so Γ is satisfiable. □

Corollary 8.21 (Completeness Theorem, Second Version). *For all Γ and sentences φ: if $\Gamma \vDash \varphi$ then $\Gamma \vdash \varphi$.*

Proof. Note that the Γ's in Corollary 8.21 and Theorem 8.20 are universally quantified. To make sure we do not confuse ourselves, let us restate Theorem 8.20 using a different variable: for any set of sentences Δ, if Δ is consistent, it is satisfiable. By contraposition, if Δ is not satisfiable, then Δ is inconsistent. We will use this to prove the corollary.

Suppose that $\Gamma \vDash \varphi$. Then $\Gamma \cup \{\neg\varphi\}$ is unsatisfiable by Proposition 5.68. Taking $\Gamma \cup \{\neg\varphi\}$ as our Δ, the previous version of Theorem 8.20 gives us that $\Gamma \cup \{\neg\varphi\}$ is inconsistent. By Proposition 7.19, $\Gamma \vdash \varphi$. □

8.9 The Compactness Theorem

One important consequence of the completeness theorem is the compactness theorem. The compactness theorem states that if each *finite* subset of a set of sentences is satisfiable, the entire set is satisfiable—even if the set itself is infinite. This is far from obvious. There is nothing that seems to rule out, at first glance at least, the possibility of there being infinite sets of sentences which are contradictory, but the contradiction only arises, so to speak, from the infinite number. The compactness theorem says that such a scenario can be ruled out: there are no unsatisfiable infinite sets of sentences each finite subset of which is satisfiable. Like the completeness theorem, it has a version related to entailment: if an infinite set of sentences entails something, already a finite subset does.

Definition 8.22. A set Γ of formulas is *finitely satisfiable* iff every finite $\Gamma_0 \subseteq \Gamma$ is satisfiable.

Theorem 8.23 (Compactness Theorem). *The following hold for any sentences* Γ *and* φ*:*

1. $\Gamma \models \varphi$ *iff there is a finite* $\Gamma_0 \subseteq \Gamma$ *such that* $\Gamma_0 \models \varphi$.
2. Γ *is satisfiable iff it is finitely satisfiable.*

Proof. We prove (2). If Γ is satisfiable, then there is a structure $\mathfrak{M}$ such that $\mathfrak{M} \models \varphi$ for all $\varphi \in \Gamma$. Of course, this $\mathfrak{M}$ also satisfies every finite subset of Γ, so Γ is finitely satisfiable.

Now suppose that Γ is finitely satisfiable. Then every finite subset $\Gamma_0 \subseteq \Gamma$ is satisfiable. By soundness (Corollary 7.31), every finite subset is consistent. Then Γ itself must be consistent by Proposition 7.17. By completeness (Theorem 8.20), since Γ is consistent, it is satisfiable. □

Example 8.24. In every model $\mathfrak{M}$ of a theory Γ, each term t of course picks out an element of $|\mathfrak{M}|$. Can we guarantee that it is also true that every element of $|\mathfrak{M}|$ is picked out by some term or other? In other words, are there theories Γ all models of which are covered? The compactness theorem shows that this is not the case if Γ has infinite models. Here's how to see this: Let $\mathfrak{M}$ be an infinite model of Γ, and let c be a constant symbol not in the language of Γ. Let Δ be the set of all sentences $c \neq t$ for t a term in the language $\mathcal{L}$ of Γ, i.e.,

$$\Delta = \{c \neq t : t \in \mathrm{Trm}(\mathcal{L})\}.$$

A finite subset of $\Gamma \cup \Delta$ can be written as $\Gamma' \cup \Delta'$, with $\Gamma' \subseteq \Gamma$ and $\Delta' \subseteq \Delta$. Since Δ' is finite, it can contain only finitely many terms. Let $a \in |\mathfrak{M}|$ be an element of $|\mathfrak{M}|$ not picked out by any of them, and let $\mathfrak{M}'$ be the structure that is just like $\mathfrak{M}$, but also $c^{\mathfrak{M}'} = a$. Since $a \neq \mathrm{Val}^{\mathfrak{M}}(t)$ for all t occuring in Δ', $\mathfrak{M}' \models \Delta'$. Since $\mathfrak{M} \models \Gamma$, $\Gamma' \subseteq \Gamma$, and c does not occur in Γ, also $\mathfrak{M}' \models \Gamma'$. Together, $\mathfrak{M}' \models \Gamma' \cup \Delta'$ for every finite subset $\Gamma' \cup \Delta'$ of $\Gamma \cup \Delta$. So every finite subset of $\Gamma \cup \Delta$ is satisfiable. By compactness, $\Gamma \cup \Delta$ itself is satisfiable. So there are models $\mathfrak{M} \models \Gamma \cup \Delta$. Every such $\mathfrak{M}$ is a model of Γ, but is not covered, since $\mathrm{Val}^{\mathfrak{M}}(c) \neq \mathrm{Val}^{\mathfrak{M}}(t)$ for all terms t of $\mathcal{L}$.

Example 8.25. Consider a language $\mathcal{L}$ containing the predicate symbol $<$, constant symbols 0, 1, and function symbols $+$, $\times$, $-$, $\div$. Let Γ be the set of all sentences in this language true in $\mathfrak{Q}$ with domain $\mathbb{Q}$ and the obvious interpretations. Γ is the set of all sentences of $\mathcal{L}$ true about the rational numbers. Of course, in $\mathbb{Q}$ (and even in $\mathbb{R}$), there are no numbers which are greater than 0 but less than $1/k$ for all $k \in \mathbb{Z}^+$. Such a number, if it existed, would be an *infinitesimal:* non-zero, but infinitely small. The compactness theorem shows that there are models of Γ in which infinitesimals exist: Let Δ be $\{0 < c\} \cup \{c < (1 \div \overline{k}) : k \in \mathbb{Z}^+\}$ (where $\overline{k} = (1 + (1 + \cdots + (1 + 1)\ldots))$ with k 1's). For any finite subset Δ_0 of Δ there is a K such that all the sentences $c < (1 \div \overline{k})$ in Δ_0

have $k < K$. If we expand $\mathfrak{Q}$ to $\mathfrak{Q}'$ with $c^{\mathfrak{Q}'} = 1/K$ we have that $\mathfrak{Q}' \vDash \Gamma \cup \Delta_0$, and so $\Gamma \cup \Delta$ is finitely satisfiable (Exercise: prove this in detail). By compactness, $\Gamma \cup \Delta$ is satisfiable. Any model $\mathfrak{S}$ of $\Gamma \cup \Delta$ contains an infinitesimal, namely $c^{\mathfrak{S}}$.

Example 8.26. We know that first-order logic with identity predicate can express that the size of the domain must have some minimal size: The sentence $\varphi_{\geq n}$ (which says "there are at least n distinct objects") is true only in structures where $|\mathfrak{M}|$ has at least n objects. So if we take

$$\Delta = \{\varphi_{\geq n} : n \geq 1\}$$

then any model of Δ must be infinite. Thus, we can guarantee that a theory only has infinite models by adding Δ to it: the models of $\Gamma \cup \Delta$ are all and only the infinite models of Γ.

So first-order logic can express infinitude. The compactness theorem shows that it cannot express finitude, however. For suppose some set of sentences Λ were satisfied in all and only finite structures. Then $\Delta \cup \Lambda$ is finitely satisfiable. Why? Suppose $\Delta' \cup \Lambda' \subseteq \Delta \cup \Lambda$ is finite with $\Delta' \subseteq \Delta$ and $\Lambda' \subseteq \Lambda$. Let n be the largest number such that $\varphi_{\geq n} \in \Delta'$. Λ, being satisfied in all finite structures, has a model $\mathfrak{M}$ with finitely many but $\geq n$ elements. But then $\mathfrak{M} \vDash \Delta' \cup \Lambda'$. By compactness, $\Delta \cup \Lambda$ has an infinite model, contradicting the assumption that Λ is satisfied only in finite structures.

8.10 A Direct Proof of the Compactness Theorem

We can prove the Compactness Theorem directly, without appealing to the Completeness Theorem, using the same ideas as in the proof of the completeness theorem. In the proof of the Completeness Theorem we started with a consistent set Γ of sentences, expanded it to a consistent, saturated, and complete set Γ^* of sentences, and then showed that in the term model $\mathfrak{M}(\Gamma^*)$ constructed from Γ^*, all sentences of Γ are true, so Γ is satisfiable.

We can use the same method to show that a finitely satisfiable set of sentences is satisfiable. We just have to prove the corresponding versions of the results leading to the truth lemma where we replace "consistent" with "finitely satisfiable."

Proposition 8.27. *Suppose Γ is complete and finitely satisfiable. Then:*

1. *$(\varphi \wedge \psi) \in \Gamma$ iff both $\varphi \in \Gamma$ and $\psi \in \Gamma$.*
2. *$(\varphi \vee \psi) \in \Gamma$ iff either $\varphi \in \Gamma$ or $\psi \in \Gamma$.*
3. *$(\varphi \to \psi) \in \Gamma$ iff either $\varphi \notin \Gamma$ or $\psi \in \Gamma$.*

Lemma 8.28. *Every finitely satisfiable set Γ can be extended to a saturated finitely satisfiable set Γ'.*

Proposition 8.29. *Suppose Γ is complete, finitely satisfiable, and saturated.*

1. *$\exists x\, \varphi(x) \in \Gamma$ iff $\varphi(t) \in \Gamma$ for at least one closed term t.*
2. *$\forall x\, \varphi(x) \in \Gamma$ iff $\varphi(t) \in \Gamma$ for all closed terms t.*

Lemma 8.30. *Every finitely satisfiable set Γ can be extended to a complete and finitely satisfiable set Γ^*.*

Theorem 8.31 (Compactness). *Γ is satisfiable if and only if it is finitely satisfiable.*

Proof. If Γ is satisfiable, then there is a structure $\mathfrak{M}$ such that $\mathfrak{M} \vDash \varphi$ for all $\varphi \in \Gamma$. Of course, this $\mathfrak{M}$ also satisfies every finite subset of Γ, so Γ is finitely satisfiable.

Now suppose that Γ is finitely satisfiable. By Lemma 8.28, there is a finitely satisfiable, saturated set $\Gamma' \supseteq \Gamma$. By Lemma 8.30, Γ' can be extended to a complete and finitely satisfiable set Γ^*, and Γ^* is still saturated. Construct the term model $\mathfrak{M}(\Gamma^*)$ as in Definition 8.9. Note that Proposition 8.11 did not rely on the fact that Γ^* is consistent (or complete or saturated, for that matter), but just on the fact that $\mathfrak{M}(\Gamma^*)$ is covered. The proof of the Truth Lemma (Lemma 8.12) goes through if we replace references to Proposition 8.2 and Proposition 8.7 by references to Proposition 8.27 and Proposition 8.29 □

8.11 The Löwenheim-Skolem Theorem

The Löwenheim-Skolem Theorem says that if a theory has an infinite model, then it also has a model that is at most denumerable. An immediate consequence of this fact is that first-order logic cannot express that the size of a structure is non-enumerable: any sentence or set of sentences satisfied in all non-enumerable structures is also satisfied in some enumerable structure.

Theorem 8.32. *If Γ is consistent then it has an enumerable model, i.e., it is satisfiable in a structure whose domain is either finite or denumerable.*

Proof. If Γ is consistent, the structure $\mathfrak{M}$ delivered by the proof of the completeness theorem has a domain $|\mathfrak{M}|$ that is no larger than the set of the terms of the language $\mathcal{L}$. So $\mathfrak{M}$ is at most denumerable. □

Theorem 8.33. *If Γ is a consistent set of sentences in the language of first-order logic without identity, then it has a denumerable model, i.e., it is satisfiable in a structure whose domain is infinite and enumerable.*

Proof. If Γ is consistent and contains no sentences in which identity appears, then the structure $\mathfrak{M}$ delivered by the proof of the completness theorem has a domain $|\mathfrak{M}|$ identical to the set of terms of the language $\mathcal{L}'$. So $\mathfrak{M}$ is denumerable, since $\mathrm{Trm}(\mathcal{L}')$ is. □

Example 8.34 (Skolem's Paradox). Zermelo-Fraenkel set theory **ZFC** is a very powerful framework in which practically all mathematical statements can be expressed, including facts about the sizes of sets. So for instance, **ZFC** can prove that the set $\mathbb{R}$ of real numbers is non-enumerable, it can prove Cantor's Theorem that the power set of any set is larger than the set itself, etc. If **ZFC** is consistent, its models are all infinite, and moreover, they all contain elements about which the theory says that they are non-enumerable, such as the element that makes true the theorem of **ZFC** that the power set of the natural numbers exists. By the Löwenheim-Skolem Theorem, **ZFC** also has enumerable models—models that contain "non-enumerable" sets but which themselves are enumerable.

8.12 Overspill

Theorem 8.35. *If a set Γ of sentences has arbitrarily large finite models, then it has an infinite model.*

Proof. Expand the language of Γ by adding countably many new constants c_0, c_1, ... and consider the set $\Gamma \cup \{c_i \neq c_j : i \neq j\}$. To say that Γ has arbitrarily large finite models means that for every $m > 0$ there is $n \geq m$ such that Γ has a model of cardinality n. This implies that $\Gamma \cup \{c_i \neq c_j : i \neq j\}$ is finitely satisfiable. By compactness, $\Gamma \cup \{c_i \neq c_j : i \neq j\}$ has a model $\mathfrak{M}$ whose domain must be infinite, since it satisfies all inequalities $c_i \neq c_j$. □

Proposition 8.36. *There is no sentence φ of any first-order language that is true in a structure $\mathfrak{M}$ if and only if the domain $|\mathfrak{M}|$ of the structure is infinite.*

Proof. If there were such a φ, its negation $\neg\varphi$ would be true in all and only the finite structures, and it would therefore have arbitrarily large finite models but it would lack an infinite model, contradicting Theorem 8.35. □

Part IV

Computability and Incompleteness

Chapter 9

Recursive Functions

9.1 Introduction

In order to develop a mathematical theory of computability, one has to, first of all, develop a *model* of computability. We now think of computability as the kind of thing that computers do, and computers work with symbols. But at the beginning of the development of theories of computability, the paradigmatic example of computation was *numerical* computation. Mathematicians were always interested in number-theoretic functions, i.e., functions $f : \mathbb{N}^n \to \mathbb{N}$ that can be computed. So it is not surprising that at the beginning of the theory of computability, it was such functions that were studied. The most familiar examples of computable numerical functions, such as addition, multiplication, exponentiation (of natural numbers) share an interesting feature: they can be defined *recursively*. It is thus quite natural to attempt a general definition of *computable function* on the basis of recursive definitions. Among the many possible ways to define number-theoretic functions recursively, one particularly simple pattern of definition here becomes central: so-called *primitive recursion*.

In addition to computable functions, we might be interested in computable sets and relations. A set is computable if we can compute the answer to whether or not a given number is an element of the set, and a relation is computable iff we can compute whether or not a tuple $\langle n_1, \ldots, n_k \rangle$ is an element of the relation. By considering the *characteristic function* of a set or relation, discussion of computable sets and relations can be subsumed under that of computable functions. Thus we can define primitive recursive relations as well, e.g., the relation "n evenly divides m" is a primitive recursive relation.

Primitive recursive functions—those that can be defined using just primitive recursion—are not, however, the only computable number-theoretic functions. Many generalizations of primitive recursion have been considered, but the most powerful and widely-accepted additional way of computing functions is by unbounded search. This leads to the definition of *partial recur-*

sive functions, and a related definition to *general recursive functions*. General recursive functions are computable and total, and the definition characterizes exactly the partial recursive functions that happen to be total. Recursive functions can simulate every other model of computation (Turing machines, lambda calculus, etc.) and so represent one of the many accepted models of computation.

9.2 Primitive Recursion

A characteristic of the natural numbers is that every natural number can be reached from 0 by applying the successor operation $+1$ finitely many times—any natural number is either 0 or the successor of ... the successor of 0. One way to specify a function $h\colon \mathbb{N} \to \mathbb{N}$ that makes use of this fact is this: (a) specify what the value of h is for argument 0, and (b) also specify how to, given the value of $h(x)$, compute the value of $h(x+1)$. For (a) tells us directly what $h(0)$ is, so h is defined for 0. Now, using the instruction given by (b) for $x = 0$, we can compute $h(1) = h(0+1)$ from $h(0)$. Using the same instructions for $x = 1$, we compute $h(2) = h(1+1)$ from $h(1)$, and so on. For every natural number x, we'll eventually reach the step where we define $h(x)$ from $h(x+1)$, and so $h(x)$ is defined for all $x \in \mathbb{N}$.

For instance, suppose we specify $h\colon \mathbb{N} \to \mathbb{N}$ by the following two equations:

$$\begin{aligned} h(0) &= 1 \\ h(x+1) &= 2 \cdot h(x) \end{aligned}$$

If we already know how to multiply, then these equations give us the information required for (a) and (b) above. By successively applying the second equation, we get that

$$\begin{aligned} h(1) &= 2 \cdot h(0) = 2, \\ h(2) &= 2 \cdot h(1) = 2 \cdot 2, \\ h(3) &= 2 \cdot h(2) = 2 \cdot 2 \cdot 2, \\ &\vdots \end{aligned}$$

We see that the function h we have specified is $h(x) = 2^x$.

The characteristic feature of the natural numbers guarantees that there is only one function h that meets these two criteria. A pair of equations like these is called a *definition by primitive recursion* of the function h. It is so-called because we define h "recursively," i.e., the definition, specifically the second equation, involves h itself on the right-hand-side. It is "primitive" because in defining $h(x+1)$ we only use the value $h(x)$, i.e., the immediately preceding value. This is the simplest way of defining a function on $\mathbb{N}$ recursively.

We can define even more fundamental functions like addition and multiplication by primitive recursion. In these cases, however, the functions in question are 2-place. We fix one of the argument places, and use the other for the recursion. E.g, to define $\text{add}(x, y)$ we can fix x and define the value first for $y = 0$ and then for $y + 1$ in terms of y. Since x is fixed, it will appear on the left and on the right side of the defining equations.

$$\begin{aligned}\text{add}(x, 0) &= x\\ \text{add}(x, y + 1) &= \text{add}(x, y) + 1\end{aligned}$$

These equations specify the value of add for all x *and* y. To find $\text{add}(2, 3)$, for instance, we apply the defining equations for $x = 2$, using the first to find $\text{add}(2, 0) = 2$, then using the second to successively find $\text{add}(2, 1) = 2 + 1 = 3$, $\text{add}(2, 2) = 3 + 1 = 4$, $\text{add}(2, 3) = 4 + 1 = 5$.

In the definition of add we used $+$ on the right-hand-side of the second equation, but only to add 1. In other words, we used the successor function $\text{succ}(z) = z + 1$ and applied it to the previous value $\text{add}(x, y)$ to define $\text{add}(x, y + 1)$. So we can think of the recursive definition as given in terms of a single function which we apply to the previous value. However, it doesn't hurt—and sometimes is necessary—to allow the function to depend not just on the previous value but also on x and y. Consider:

$$\begin{aligned}\text{mult}(x, 0) &= 0\\ \text{mult}(x, y + 1) &= \text{add}(\text{mult}(x, y), x)\end{aligned}$$

This is a primitive recursive definition of a function mult by applying the function add to both the preceding value $\text{mult}(x, y)$ and the first argument x. It also defines the function $\text{mult}(x, y)$ for all arguments x and y. For instance, $\text{mult}(2, 3)$ is determined by successively computing $\text{mult}(2, 0)$, $\text{mult}(2, 1)$, $\text{mult}(2, 2)$, and $\text{mult}(2, 3)$:

$$\begin{aligned}&\text{mult}(2, 0) = 0\\ &\text{mult}(2, 1) = \text{mult}(2, 0 + 1) = \text{add}(\text{mult}(2, 0), 2) = \text{add}(0, 2) = 2\\ &\text{mult}(2, 2) = \text{mult}(2, 1 + 1) = \text{add}(\text{mult}(2, 1), 2) = \text{add}(2, 2) = 4\\ &\text{mult}(2, 3) = \text{mult}(2, 2 + 1) = \text{add}(\text{mult}(2, 2), 2) = \text{add}(4, 2) = 6\end{aligned}$$

The general pattern then is this: to give a primitive recursive definition of a function $h(x_0, \ldots, x_{k-1}, y)$, we provide two equations. The first defines the value of $h(x_0, \ldots, x_{k-1}, 0)$ without reference to h. The second defines the value of $h(x_0, \ldots, x_{k-1}, y + 1)$ in terms of $h(x_0, \ldots, x_{k-1}, y)$, the other arguments x_0, $\ldots$, x_{k-1}, and y. Only the immediately preceding value of h may be used in that second equation. If we think of the operations given by the right-hand-sides of these two equations as themselves being functions f and g, then the

general pattern to define a new function h by primitive recursion is this:

$$h(x_0, \ldots, x_{k-1}, 0) = f(x_0, \ldots, x_{k-1})$$
$$h(x_0, \ldots, x_{k-1}, y+1) = g(x_0, \ldots, x_{k-1}, y, h(x_0, \ldots, x_{k-1}, y))$$

In the case of add, we have $k = 1$ and $f(x_0) = x_0$ (the identity function), and $g(x_0, y, z) = z + 1$ (the 3-place function that returns the successor of its third argument):

$$\text{add}(x_0, 0) = f(x_0) = x_0$$
$$\text{add}(x_0, y+1) = g(x_0, y, \text{add}(x_0, y)) = \text{succ}(\text{add}(x_0, y))$$

In the case of mult, we have $f(x_0) = 0$ (the constant function always returning 0) and $g(x_0, y, z) = \text{add}(z, x_0)$ (the 3-place function that returns the sum of its last and first argument):

$$\text{mult}(x_0, 0) = f(x_0) = 0$$
$$\text{mult}(x_0, y+1) = g(x_0, y, \text{mult}(x_0, y)) = \text{add}(\text{mult}(x_0, y), x_0)$$

9.3 Composition

If f and g are two one-place functions of natural numbers, we can compose them: $h(x) = g(f(x))$. The new function $h(x)$ is then defined by *composition* from the functions f and g. We'd like to generalize this to functions of more than one argument.

Here's one way of doing this: suppose f is a k-place function, and $g_0, \ldots, g_{k-1}$ are k functions which are all n-place. Then we can define a new n-place function h as follows:

$$h(x_0, \ldots, x_{n-1}) = f(g_0(x_0, \ldots, x_{n-1}), \ldots, g_{k-1}(x_0, \ldots, x_{n-1}))$$

If f and all g_i are computable, so is h: To compute $h(x_0, \ldots, x_{n-1})$, first compute the values $y_i = g_i(x_0, \ldots, x_{n-1})$ for each $i = 0, \ldots, k-1$. Then feed these values into f to compute $h(x_0, \ldots, x_{k-1}) = f(y_0, \ldots, y_{k-1})$.

This may seem like an overly restrictive characterization of what happens when we compute a new function using some existing ones. For one thing, sometimes we do not use all the arguments of a function, as when we defined $g(x, y, z) = \text{succ}(z)$ for use in the primitive recursive definition of add. Suppose we are allowed use of the following functions:

$$P_i^n(x_0, \ldots, x_{n-1}) = x_i$$

The functions P_i^k are called *projection* functions: P_i^n is an n-place function. Then g can be defined by

$$g(x, y, z) = \text{succ}(P_2^3(x, y, z)).$$

Here the role of f is played by the 1-place function succ, so $k = 1$. And we have one 3-place function P^3_2 which plays the role of g_0. The result is a 3-place function that returns the successor of the third argument.

The projection functions also allow us to define new functions by reordering or identifying arguments. For instance, the function $h(x) = \text{add}(x, x)$ can be defined by

$$h(x_0) = \text{add}(P^1_0(x_0), P^1_0(x_0)).$$

Here $k = 2$, $n = 1$, the role of $f(y_0, y_1)$ is played by add, and the roles of $g_0(x_0)$ and $g_1(x_0)$ are both played by $P^1_0(x_0)$, the one-place projection function (aka the identity function).

If $f(y_0, y_1)$ is a function we already have, we can define the function $h(x_0, x_1) = f(x_1, x_0)$ by

$$h(x_0, x_1) = f(P^2_1(x_0, x_1), P^2_0(x_0, x_1)).$$

Here $k = 2$, $n = 2$, and the roles of g_0 and g_1 are played by P^2_1 and P^2_0, respectively.

You may also worry that $g_0, \ldots, g_{k-1}$ are all required to have the same arity n. (Remember that the *arity* of a function is the number of arguments; an n-place function has arity n.) But adding the projection functions provides the desired flexibility. For example, suppose f and g are 3-place functions and h is the 2-place function defined by

$$h(x, y) = f(x, g(x, x, y), y).$$

The definition of h can be rewritten with the projection functions, as

$$h(x, y) = f(P^2_0(x, y), g(P^2_0(x, y), P^2_0(x, y), P^2_1(x, y)), P^2_1(x, y)).$$

Then h is the composition of f with P^2_0, l, and P^2_1, where

$$l(x, y) = g(P^2_0(x, y), P^2_0(x, y), P^2_1(x, y)),$$

i.e., l is the composition of g with P^2_0, P^2_0, and P^2_1.

9.4 Primitive Recursion Functions

Let us record again how we can define new functions from existing ones using primitive recursion and composition.

Definition 9.1. Suppose f is a k-place function ($k \geq 1$) and g is a $(k+2)$-place function. The function defined by *primitive recursion from f and g* is the $(k+1)$-place function h defined by the equations

$$\begin{aligned} h(x_0, \ldots, x_{k-1}, 0) &= f(x_0, \ldots, x_{k-1}) \\ h(x_0, \ldots, x_{k-1}, y+1) &= g(x_0, \ldots, x_{k-1}, y, h(x_0, \ldots, x_{k-1}, y)) \end{aligned}$$

Definition 9.2. Suppose f is a k-place function, and $g_0, \ldots, g_{k-1}$ are k functions which are all n-place. The function defined by *composition from f and $g_0, \ldots, g_{k-1}$* is the n-place function h defined by

$$h(x_0, \ldots, x_{n-1}) = f(g_0(x_0, \ldots, x_{n-1}), \ldots, g_{k-1}(x_0, \ldots, x_{n-1})).$$

In addition to succ and the projection functions

$$P_i^n(x_0, \ldots, x_{n-1}) = x_i,$$

for each natural number n and $i < n$, we will include among the primitive recursive functions the function $\text{zero}(x) = 0$.

Definition 9.3. The set of primitive recursive functions is the set of functions from $\mathbb{N}^n$ to $\mathbb{N}$, defined inductively by the following clauses:

1. zero is primitive recursive.
2. succ is primitive recursive.
3. Each projection function P_i^n is primitive recursive.
4. If f is a k-place primitive recursive function and $g_0, \ldots, g_{k-1}$ are n-place primitive recursive functions, then the composition of f with $g_0, \ldots, g_{k-1}$ is primitive recursive.
5. If f is a k-place primitive recursive function and g is a $k+2$-place primitive recursive function, then the function defined by primitive recursion from f and g is primitive recursive.

Put more concisely, the set of primitive recursive functions is the smallest set containing zero, succ, and the projection functions P_j^n, and which is closed under composition and primitive recursion.

Another way of describing the set of primitive recursive functions is by defining it in terms of "stages." Let S_0 denote the set of starting functions: zero, succ, and the projections. These are the primitive recursive functions of stage 0. Once a stage S_i has been defined, let S_{i+1} be the set of all functions you get by applying a single instance of composition or primitive recursion to functions already in S_i. Then

$$S = \bigcup_{i \in \mathbb{N}} S_i$$

is the set of all primitive recursive functions

Let us verify that add is a primitive recursive function.

Proposition 9.4. *The addition function* $\text{add}(x, y) = x + y$ *is primitive recursive.*

Proof. We already have a primitive recursive definition of add in terms of two functions f and g which matches the format of Definition 9.1:

$$\begin{aligned}\text{add}(x_0, 0) &= f(x_0) = x_0\\ \text{add}(x_0, y+1) &= g(x_0, y, \text{add}(x_0, y)) = \text{succ}(\text{add}(x_0, y))\end{aligned}$$

So add is primitive recursive provided f and g are as well. $f(x_0) = x_0 = P^1_0(x_0)$, and the projection functions count as primitive recursive, so f is primitive recursive. The function g is the three-place function $g(x_0, y, z)$ defined by

$$g(x_0, y, z) = \text{succ}(z).$$

This does not yet tell us that g is primitive recursive, since g and succ are not quite the same function: succ is one-place, and g has to be three-place. But we can define g "officially" by composition as

$$g(x_0, y, z) = \text{succ}(P^3_2(x_0, y, z))$$

Since succ and P^3_2 count as primitive recursive functions, g does as well, since it can be defined by composition from primitive recursive functions. □

Proposition 9.5. *The multiplication function* $\text{mult}(x, y) = x \cdot y$ *is primitive recursive.*

Proof. Exercise. □

Example 9.6. Here's our very first example of a primitive recursive definition:

$$\begin{aligned}h(0) &= 1\\ h(y+1) &= 2 \cdot h(y).\end{aligned}$$

This function cannot fit into the form required by Definition 9.1, since $k = 0$. The definition also involves the constants 1 and 2. To get around the first problem, let's introduce a dummy argument and define the function h':

$$\begin{aligned}h'(x_0, 0) &= f(x_0) = 1\\ h'(x_0, y+1) &= g(x_0, y, h'(x_0, y)) = 2 \cdot h'(x_0, y).\end{aligned}$$

The function $f(x_0) = 1$ can be defined from succ and zero by composition: $f(x_0) = \text{succ}(\text{zero}(x_0))$. The function g can be defined by composition from $g'(z) = 2 \cdot z$ and projections:

$$g(x_0, y, z) = g'(P^3_2(x_0, y, z))$$

and g' in turn can be defined by composition as

$$g'(z) = \text{mult}(g''(z), P^1_0(z))$$

and

$$g''(z) = \text{succ}(f(z)),$$

where f is as above: $f(z) = \text{succ}(\text{zero}(z))$. Now that we have h', we can use composition again to let $h(y) = h'(P^1_0(y), P^1_0(y))$. This shows that h can be defined from the basic functions using a sequence of compositions and primitive recursions, so h is primitive recursive.

9.5 Primitive Recursion Notations

One advantage to having the precise inductive description of the primitive recursive functions is that we can be systematic in describing them. For example, we can assign a "notation" to each such function, as follows. Use symbols zero, succ, and P^n_i for zero, successor, and the projections. Now suppose h is defined by composition from a k-place function f and n-place functions g_0, $\ldots$, g_{k-1}, and we have assigned notations F, G_0, $\ldots$, G_{k-1} to the latter functions. Then, using a new symbol $\text{Comp}_{k,n}$, we can denote the function h by $\text{Comp}_{k,n}[F, G_0, \ldots, G_{k-1}]$.

For functions defined by primitive recursion, we can use analogous notations. Suppose the $(k+1)$-ary function h is defined by primitive recursion from the k-ary function f and the $(k+2)$-ary function g, and the notations assigned to f and g are F and G, respectively. Then the notation assigned to h is $\text{Rec}_k[F, G]$.

Recall that the addition function is defined by primitive recursion as

$$\begin{aligned} \text{add}(x_0, 0) &= P^1_0(x_0) = x_0 \\ \text{add}(x_0, y+1) &= \text{succ}(P^3_2(x_0, y, \text{add}(x_0, y))) = \text{add}(x_0, y) + 1 \end{aligned}$$

Here the role of f is played by P^1_0, and the role of g is played by $\text{succ}(P^3_2(x_0, y, z))$, which is assigned the notation $\text{Comp}_{1,3}[\text{succ}, P^3_2]$ as it is the result of defining a function by composition from the 1-ary function succ and the 3-ary function P^3_2. With this setup, we can denote the addition function by

$$\text{Rec}_1[P^1_0, \text{Comp}_{1,3}[\text{succ}, P^3_2]].$$

Having these notations sometimes proves useful, e.g., when enumerating primitive recursive functions.

9.6 Primitive Recursive Functions are Computable

Suppose a function h is defined by primitive recursion

$$\begin{aligned} h(\vec{x}, 0) &= f(\vec{x}) \\ h(\vec{x}, y+1) &= g(\vec{x}, y, h(\vec{x}, y)) \end{aligned}$$

and suppose the functions f and g are computable. (We use $\vec{x}$ to abbreviate x_0, $\ldots$, x_{k-1}.) Then $h(\vec{x}, 0)$ can obviously be computed, since it is just $f(\vec{x})$ which we assume is computable. $h(\vec{x}, 1)$ can then also be computed, since $1 = 0 + 1$ and so $h(\vec{x}, 1)$ is just

$$h(\vec{x}, 1) = g(\vec{x}, 0, h(\vec{x}, 0)) = g(\vec{x}, 0, f(\vec{x})).$$

We can go on in this way and compute

$$\begin{aligned}
h(\vec{x}, 2) &= g(\vec{x}, 1, h(\vec{x}, 1)) = g(\vec{x}, 1, g(\vec{x}, 0, f(\vec{x}))) \\
h(\vec{x}, 3) &= g(\vec{x}, 2, h(\vec{x}, 2)) = g(\vec{x}, 2, g(\vec{x}, 1, g(\vec{x}, 0, f(\vec{x})))) \\
h(\vec{x}, 4) &= g(\vec{x}, 3, h(\vec{x}, 3)) = g(\vec{x}, 3, g(\vec{x}, 2, g(\vec{x}, 1, g(\vec{x}, 0, f(\vec{x})))))) \\
&\vdots
\end{aligned}$$

Thus, to compute $h(\vec{x}, y)$ in general, successively compute $h(\vec{x}, 0)$, $h(\vec{x}, 1)$, $\ldots$, until we reach $h(\vec{x}, y)$.

Thus, a primitive recursive definition yields a new computable function if the functions f and g are computable. Composition of functions also results in a computable function if the functions f and g_i are computable.

Since the basic functions zero, succ, and P_i^n are computable, and composition and primitive recursion yield computable functions from computable functions, this means that every primitive recursive function is computable.

9.7 Examples of Primitive Recursive Functions

We already have some examples of primitive recursive functions: the addition and multiplication functions add and mult. The identity function $\mathrm{id}(x) = x$ is primitive recursive, since it is just P_0^1. The constant functions $\mathrm{const}_n(x) = n$ are primitive recursive since they can be defined from zero and succ by successive composition. This is useful when we want to use constants in primitive recursive definitions, e.g., if we want to define the function $f(x) = 2 \cdot x$ can obtain it by composition from $\mathrm{const}_n(x)$ and multiplication as $f(x) = \mathrm{mult}(\mathrm{const}_2(x), P_0^1(x))$. We'll make use of this trick from now on.

Proposition 9.7. *The exponentiation function* $\exp(x, y) = x^y$ *is primitive recursive.*

Proof. We can define exp primitive recursively as

$$\begin{aligned}
\exp(x, 0) &= 1 \\
\exp(x, y + 1) &= \mathrm{mult}(x, \exp(x, y)).
\end{aligned}$$

Strictly speaking, this is not a recursive definition from primitive recursive functions. Officially, though, we have:

$$\begin{aligned}\exp(x, 0) &= f(x)\\ \exp(x, y+1) &= g(x, y, \exp(x, y)).\end{aligned}$$

where

$$\begin{aligned}f(x) &= \text{succ}(\text{zero}(x)) = 1\\ g(x, y, z) &= \text{mult}(P^3_0(x, y, z), P^3_2(x, y, z)) = x \cdot z\end{aligned}$$

and so f and g are defined from primitive recursive functions by composition. □

Proposition 9.8. *The predecessor function* $\text{pred}(y)$ *defined by*

$$\text{pred}(y) = \begin{cases} 0 & \text{if } y = 0\\ y - 1 & \text{otherwise}\end{cases}$$

is primitive recursive.

Proof. Note that

$$\begin{aligned}\text{pred}(0) &= 0 \text{ and}\\ \text{pred}(y+1) &= y.\end{aligned}$$

This is almost a primitive recursive definition. It does not, strictly speaking, fit into the pattern of definition by primitive recursion, since that pattern requires at least one extra argument x. It is also odd in that it does not actually use $\text{pred}(y)$ in the definition of $\text{pred}(y+1)$. But we can first define $\text{pred}'(x, y)$ by

$$\begin{aligned}\text{pred}'(x, 0) &= \text{zero}(x) = 0,\\ \text{pred}'(x, y+1) &= P^3_1(x, y, \text{pred}'(x, y)) = y.\end{aligned}$$

and then define pred from it by composition, e.g., as $\text{pred}(x) = \text{pred}'(\text{zero}(x), P^1_0(x))$.□

Proposition 9.9. *The factorial function* $\text{fac}(x) = x! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot x$ *is primitive recursive.*

Proof. The obvious primitive recursive definition is

$$\begin{aligned}\text{fac}(0) &= 1\\ \text{fac}(y+1) &= \text{fac}(y) \cdot (y+1).\end{aligned}$$

Officially, we have to first define a two-place function h

$$\begin{aligned} h(x,0) &= \text{const}_1(x) \\ h(x,y+1) &= g(x,y,h(x,y)) \end{aligned}$$

where $g(x,y,z) = \text{mult}(P^3_2(x,y,z), \text{succ}(P^3_1(x,y,z)))$ and then let

$$\text{fac}(y) = h(P^1_0(y), P^1_0(y)) = h(y,y).$$

From now on we'll be a bit more laissez-faire and not give the official definitions by composition and primitive recursion. □

Proposition 9.10. *Truncated subtraction, $x \dot{-} y$, defined by*

$$x \dot{-} y = \begin{cases} 0 & \text{if } x < y \\ x - y & \text{otherwise} \end{cases}$$

is primitive recursive.

Proof. We have:

$$\begin{aligned} x \dot{-} 0 &= x \\ x \dot{-} (y+1) &= \text{pred}(x \dot{-} y) \end{aligned}$$ □

Proposition 9.11. *The distance between x and y, $|x-y|$, is primitive recursive.*

Proof. We have $|x-y| = (x \dot{-} y) + (y \dot{-} x)$, so the distance can be defined by composition from $+$ and $\dot{-}$, which are primitive recursive. □

Proposition 9.12. *The maximum of x and y, $\max(x,y)$, is primitive recursive.*

Proof. We can define $\max(x,y)$ by composition from $+$ and $\dot{-}$ by

$$\max(x,y) = x + (y \dot{-} x).$$

If x is the maximum, i.e., $x \geq y$, then $y \dot{-} x = 0$, so $x + (y \dot{-} x) = x + 0 = x$. If y is the maximum, then $y \dot{-} x = y - x$, and so $x + (y \dot{-} x) = x + (y - x) = y$.□

Proposition 9.13. *The minimum of x and y, $\min(x,y)$, is primitive recursive.*

Proof. Exercise. □

Proposition 9.14. *The set of primitive recursive functions is closed under the following two operations:*

1. *Finite sums: if $f(\vec{x}, z)$ is primitive recursive, then so is the function*

$$g(\vec{x}, y) = \sum_{z=0}^{y} f(\vec{x}, z).$$

2. *Finite products: if $f(\vec{x}, z)$ is primitive recursive, then so is the function*

$$h(\vec{x}, y) = \prod_{z=0}^{y} f(\vec{x}, z).$$

Proof. For example, finite sums are defined recursively by the equations

$$\begin{aligned} g(\vec{x}, 0) &= f(\vec{x}, 0) \\ g(\vec{x}, y+1) &= g(\vec{x}, y) + f(\vec{x}, y+1). \end{aligned}$$

□

9.8 Primitive Recursive Relations

Definition 9.15. A relation $R(\vec{x})$ is said to be primitive recursive if its characteristic function,

$$\chi_R(\vec{x}) = \begin{cases} 1 & \text{if } R(\vec{x}) \\ 0 & \text{otherwise} \end{cases}$$

is primitive recursive.

In other words, when one speaks of a primitive recursive relation $R(\vec{x})$, one is referring to a relation of the form $\chi_R(\vec{x}) = 1$, where χ_R is a primitive recursive function which, on any input, returns either 1 or 0. For example, the relation IsZero(x), which holds if and only if $x = 0$, corresponds to the function χ_{IsZero}, defined using primitive recursion by

$$\begin{aligned} \chi_{\text{IsZero}}(0) &= 1, \\ \chi_{\text{IsZero}}(x+1) &= 0. \end{aligned}$$

It should be clear that one can compose relations with other primitive recursive functions. So the following are also primitive recursive:

1. The equality relation, $x = y$, defined by IsZero$(|x - y|)$
2. The less-than relation, $x \leq y$, defined by IsZero$(x \dot{-} y)$

Proposition 9.16. *The set of primitive recursive relations is closed under Boolean operations, that is, if $P(\vec{x})$ and $Q(\vec{x})$ are primitive recursive, so are*

1. $\neg P(\vec{x})$
2. $P(\vec{x}) \wedge Q(\vec{x})$

3. $P(\vec{x}) \vee Q(\vec{x})$
4. $P(\vec{x}) \to Q(\vec{x})$

Proof. Suppose $P(\vec{x})$ and $Q(\vec{x})$ are primitive recursive, i.e., their characteristic functions χ_P and χ_Q are. We have to show that the characteristic functions of $\neg P(\vec{x})$, etc., are also primitive recursive.

$$\chi_{\neg P}(\vec{x}) = \begin{cases} 0 & \text{if } \chi_P(\vec{x}) = 1 \\ 1 & \text{otherwise} \end{cases}$$

We can define $\chi_{\neg P}(\vec{x})$ as $1 \dot{-} \chi_P(\vec{x})$.

$$\chi_{P \wedge Q}(\vec{x}) = \begin{cases} 1 & \text{if } \chi_P(\vec{x}) = \chi_Q(\vec{x}) = 1 \\ 0 & \text{otherwise} \end{cases}$$

We can define $\chi_{P \wedge Q}(\vec{x})$ as $\chi_P(\vec{x}) \cdot \chi_Q(\vec{x})$ or as $\min(\chi_P(\vec{x}), \chi_Q(\vec{x}))$. Similarly,

$$\begin{aligned} \chi_{P \vee Q}(\vec{x}) &= \max(\chi_P(\vec{x}), \chi_Q(\vec{x}))) \text{ and} \\ \chi_{P \to Q}(\vec{x}) &= \max(1 \dot{-} \chi_P(\vec{x}), \chi_Q(\vec{x})). \end{aligned}$$

□

Proposition 9.17. *The set of primitive recursive relations is closed under bounded quantification, i.e., if $R(\vec{x}, z)$ is a primitive recursive relation, then so are the relations*

$$\begin{aligned} &(\forall z < y)\ R(\vec{x}, z) \text{ and} \\ &(\exists z < y)\ R(\vec{x}, z). \end{aligned}$$

$(\forall z < y)\ R(\vec{x}, z)$ holds of $\vec{x}$ and y if and only if $R(\vec{x}, z)$ holds for every z less than y, and similarly for $(\exists z < y)\ R(\vec{x}, z)$.

Proof. By convention, we take $(\forall z < 0)\ R(\vec{x}, z)$ to be true (for the trivial reason that there are no z less than 0) and $(\exists z < 0)\ R(\vec{x}, z)$ to be false. A bounded universal quantifier functions just like a finite product or iterated minimum, i.e., if $P(\vec{x}, y) \Leftrightarrow (\forall z < y)\ R(\vec{x}, z)$ then $\chi_P(\vec{x}, y)$ can be defined by

$$\begin{aligned} \chi_P(\vec{x}, 0) &= 1 \\ \chi_P(\vec{x}, y+1) &= \min(\chi_P(\vec{x}, y), \chi_R(\vec{x}, y))). \end{aligned}$$

Bounded existential quantification can similarly be defined using max. Alternatively, it can be defined from bounded universal quantification, using the equivalence $(\exists z < y)\ R(\vec{x}, z) \leftrightarrow \neg(\forall z < y)\ \neg R(\vec{x}, z)$. Note that, for example, a bounded quantifier of the form $(\exists x \leq y)\ \ldots x \ldots$ is equivalent to $(\exists x < y+1)\ \ldots x \ldots$. □

Another useful primitive recursive function is the conditional function, $\text{cond}(x, y, z)$, defined by

$$\text{cond}(x, y, z) = \begin{cases} y & \text{if } x = 0 \\ z & \text{otherwise.} \end{cases}$$

This is defined recursively by

$$\begin{aligned} \text{cond}(0, y, z) &= y, \\ \text{cond}(x + 1, y, z) &= z. \end{aligned}$$

One can use this to justify definitions of primitive recursive functions by cases from primitive recursive relations:

Proposition 9.18. *If $g_0(\vec{x}), \ldots, g_m(\vec{x})$ are primitive recursive functions, and $R_0(\vec{x})$, $\ldots$, $R_{m-1}(\vec{x})$ are primitive recursive relations, then the function f defined by*

$$f(\vec{x}) = \begin{cases} g_0(\vec{x}) & \textit{if } R_0(\vec{x}) \\ g_1(\vec{x}) & \textit{if } R_1(\vec{x}) \textit{ and not } R_0(\vec{x}) \\ \vdots & \\ g_{m-1}(\vec{x}) & \textit{if } R_{m-1}(\vec{x}) \textit{ and none of the previous hold} \\ g_m(\vec{x}) & \textit{otherwise} \end{cases}$$

is also primitive recursive.

Proof. When $m = 1$, this is just the function defined by

$$f(\vec{x}) = \text{cond}(\chi_{\neg R_0}(\vec{x}), g_0(\vec{x}), g_1(\vec{x})).$$

For m greater than 1, one can just compose definitions of this form. □

9.9 Bounded Minimization

It is often useful to define a function as the least number satisfying some property or relation P. If P is decidable, we can compute this function simply by trying out all the possible numbers, $0, 1, 2, \ldots$, until we find the least one satisfying P. This kind of unbounded search takes us out of the realm of primitive recursive functions. However, if we're only interested in the least number *less than some independently given bound*, we stay primitive recursive. In other words, and a bit more generally, suppose we have a primitive recursive relation $R(x, z)$. Consider the function that maps x and y to the least $z < y$ such that $R(x, z)$. It, too, can be computed, by testing whether $R(x, 0)$, $R(x, 1)$, $\ldots$, $R(x, y - 1)$. But why is it primitive recursive?

Proposition 9.19. *If $R(\vec{x}, z)$ is primitive recursive, so is the function $m_R(\vec{x}, y)$ which returns the least z less than y such that $R(\vec{x}, z)$ holds, if there is one, and y otherwise. We will write the function m_R as*

$$(\min z < y)\, R(\vec{x}, z),$$

Proof. Note than there can be no $z < 0$ such that $R(\vec{x}, z)$ since there is no $z < 0$ at all. So $m_R(\vec{x}, 0) = 0$.

In case the bound is of the form $y + 1$ we have three cases:

1. There is a $z < y$ such that $R(\vec{x}, z)$, in which case $m_R(\vec{x}, y + 1) = m_R(\vec{x}, y)$.
2. There is no such $z < y$ but $R(\vec{x}, y)$ holds, then $m_R(\vec{x}, y + 1) = y$.
3. There is no $z < y + 1$ such that $R(\vec{x}, z)$, then $m_R(\vec{z}, y + 1) = y + 1$.

So we can define $m_R(\vec{x}, 0)$ by primitive recursion as follows:

$$m_R(\vec{x}, 0) = 0$$
$$m_R(\vec{x}, y+1) = \begin{cases} m_R(\vec{x}, y) & \text{if } m_R(\vec{x}, y) \neq y \\ y & \text{if } m_R(\vec{x}, y) = y \text{ and } R(\vec{x}, y) \\ y + 1 & \text{otherwise.} \end{cases}$$

Note that there is a $z < y$ such that $R(\vec{x}, z)$ iff $m_R(\vec{x}, y) \neq y$. □

9.10 Primes

Bounded quantification and bounded minimization provide us with a good deal of machinery to show that natural functions and relations are primitive recursive. For example, consider the relation "x divides y", written $x \mid y$. The relation $x \mid y$ holds if division of y by x is possible without remainder, i.e., if y is an integer multiple of x. (If it doesn't hold, i.e., the remainder when dividing x by y is > 0, we write $x \nmid y$.) In other words, $x \mid y$ iff for some z, $x \cdot z = y$. Obviously, any such z, if it exists, must be $\leq y$. So, we have that $x \mid y$ iff for some $z \leq y$, $x \cdot z = y$. We can define the relation $x \mid y$ by bounded existential quantification from $=$ and multiplication by

$$x \mid y \Leftrightarrow (\exists z \leq y)\ (x \cdot z) = y.$$

We've thus shown that $x \mid y$ is primitive recursive.

A natural number x is *prime* if it is neither 0 nor 1 and is only divisible by 1 and itself. In other words, prime numbers are such that, whenever $y \mid x$, either $y = 1$ or $y = x$. To test if x is prime, we only have to check if $y \mid x$ for all $y \leq x$, since if $y > x$, then automatically $y \nmid x$. So, the relation $\mathrm{Prime}(x)$, which holds iff x is prime, can be defined by

$$\mathrm{Prime}(x) \Leftrightarrow x \geq 2 \wedge (\forall y \leq x)\ (y \mid x \rightarrow y = 1 \vee y = x)$$

and is thus primitive recursive.

The primes are 2, 3, 5, 7, 11, etc. Consider the function $p(x)$ which returns the xth prime in that sequence, i.e., $p(0) = 2$, $p(1) = 3$, $p(2) = 5$, etc. (For convenience we will often write $p(x)$ as p_x ($p_0 = 2$, $p_1 = 3$, etc.)

If we had a function nextPrime(x), which returns the first prime number larger than x, p can be easily defined using primitive recursion:

$$\begin{aligned} p(0) &= 2 \\ p(x+1) &= \text{nextPrime}(p(x)) \end{aligned}$$

Since nextPrime(x) is the least y such that $y > x$ and y is prime, it can be easily computed by unbounded search. But it can also be defined by bounded minimization, thanks to a result due to Euclid: there is always a prime number between x and $x! + 1$.

$$\text{nextPrime(x)} = (\min y \leq x! + 1)\,(y > x \wedge \text{Prime}(y)).$$

This shows, that nextPrime(x) and hence $p(x)$ are (not just computable but) primitive recursive.

(If you're curious, here's a quick proof of Euclid's theorem. Suppose p_n is the largest prime $\leq x$ and consider the product $p = p_0 \cdot p_1 \cdot \cdots \cdot p_n$ of all primes $\leq x$. Either $p + 1$ is prime or there is a prime between x and $p + 1$. Why? Suppose $p + 1$ is not prime. Then some prime number $q \mid p + 1$ where $q < p + 1$. None of the primes $\leq x$ divide $p + 1$. (By definition of p, each of the primes $p_i \leq x$ divides p, i.e., with remainder 0. So, each of the primes $p_i \leq x$ divides $p + 1$ with remainder 1, and so $p_i \nmid p + 1$.) Hence, q is a prime $> x$ and $< p + 1$. And $p \leq x!$, so there is a prime $> x$ and $\leq x! + 1$.)

9.11 Sequences

The set of primitive recursive functions is remarkably robust. But we will be able to do even more once we have developed a adequate means of handling *sequences*. We will identify finite sequences of natural numbers with natural numbers in the following way: the sequence $\langle a_0, a_1, a_2, \ldots, a_k \rangle$ corresponds to the number

$$p_0^{a_0+1} \cdot p_1^{a_1+1} \cdot p_2^{a_2+1} \cdot \cdots \cdot p_k^{a_k+1}.$$

We add one to the exponents to guarantee that, for example, the sequences $\langle 2, 7, 3 \rangle$ and $\langle 2, 7, 3, 0, 0 \rangle$ have distinct numeric codes. We can take both 0 and 1 to code the empty sequence; for concreteness, let Λ denote 0.

The reason that this coding of sequences works is the so-called Fundamental Theorem of Arithmetic: every natural number $n \geq 2$ can be written in one and only one way in the form

$$n = p_0^{a_0} \cdot p_1^{a_1} \cdot \cdots \cdot p_k^{a_k}$$

with $a_k \geq 1$. This guarantees that the mapping $\langle\rangle(a_0, \ldots, a_k) = \langle a_0, \ldots, a_k\rangle$ is injective: different sequences are mapped to different numbers; to each number only at most one sequence corresponds.

We'll now show that the operations of determining the length of a sequence, determining its ith element, appending an element to a sequence, and concatenating two sequences, are all primitive recursive.

Proposition 9.20. *The function* $\text{len}(s)$*, which returns the length of the sequence* s*, is primitive recursive.*

Proof. Let $R(i, s)$ be the relation defined by

$$R(i, s) \text{ iff } p_i \mid s \wedge p_{i+1} \nmid s.$$

R is clearly primitive recursive. Whenever s is the code of a non-empty sequence, i.e.,

$$s = p_0^{a_0+1} \cdot \cdots \cdot p_k^{a_k+1},$$

$R(i, s)$ holds if p_i is the largest prime such that $p_i \mid s$, i.e., $i = k$. The length of s thus is $i + 1$ iff p_i is the largest prime that divides s, so we can let

$$\text{len}(s) = \begin{cases} 0 & \text{if } s = 0 \text{ or } s = 1 \\ 1 + (\min i < s)\, R(i, s) & \text{otherwise} \end{cases}$$

We can use bounded minimization, since there is only one i that satisfies $R(s, i)$ when s is a code of a sequence, and if i exists it is less than s itself. □

Proposition 9.21. *The function* $\text{append}(s, a)$*, which returns the result of appending* a *to the sequence* s*, is primitive recursive.*

Proof. append can be defined by:

$$\text{append}(s, a) = \begin{cases} 2^{a+1} & \text{if } s = 0 \text{ or } s = 1 \\ s \cdot p_{\text{len}(s)}^{a+1} & \text{otherwise.} \end{cases}$$

□

Proposition 9.22. *The function* $\text{element}(s, i)$*, which returns the* i*th element of* s *(where the initial element is called the* 0*th), or* 0 *if* i *is greater than or equal to the length of* s*, is primitive recursive.*

Proof. Note that a is the ith element of s iff p_i^{a+1} is the largest power of p_i that divides s, i.e., $p_i^{a+1} \mid s$ but $p_i^{a+2} \nmid s$. So:

$$\text{element}(s, i) = \begin{cases} 0 & \text{if } i \geq \text{len}(s) \\ (\min a < s)\, (p_i^{a+2} \nmid s) & \text{otherwise.} \end{cases}$$

□

Instead of using the official names for the functions defined above, we introduce a more compact notation. We will use $(s)_i$ instead of element(s, i), and $\langle s_0, \ldots, s_k \rangle$ to abbreviate

$$\text{append}(\text{append}(\ldots \text{append}(\Lambda, s_0) \ldots), s_k).$$

Note that if s has length k, the elements of s are $(s)_0, \ldots, (s)_{k-1}$.

Proposition 9.23. *The function* concat(s, t)*, which concatenates two sequences, is primitive recursive.*

Proof. We want a function concat with the property that

$$\text{concat}(\langle a_0, \ldots, a_k \rangle, \langle b_0, \ldots, b_l \rangle) = \langle a_0, \ldots, a_k, b_0, \ldots, b_l \rangle.$$

We'll use a "helper" function hconcat(s, t, n) which concatenates the first n symbols of t to s. This function can be defined by primitive recursion as follows:

$$\begin{aligned} \text{hconcat}(s, t, 0) &= s \\ \text{hconcat}(s, t, n+1) &= \text{append}(\text{hconcat}(s, t, n), (t)_n) \end{aligned}$$

Then we can define concat by

$$\text{concat}(s, t) = \text{hconcat}(s, t, \text{len}(t)). \qquad \square$$

We will write $s \frown t$ instead of concat(s, t).

It will be useful for us to be able to bound the numeric code of a sequence in terms of its length and its largest element. Suppose s is a sequence of length k, each element of which is less than or equal to some number x. Then s has at most k prime factors, each at most p_{k-1}, and each raised to at most $x + 1$ in the prime factorization of s. In other words, if we define

$$\text{sequenceBound}(x, k) = p_{k-1}^{k \cdot (x+1)},$$

then the numeric code of the sequence s described above is at most sequenceBound(x, k).

Having such a bound on sequences gives us a way of defining new functions using bounded search. For example, we can define concat using bounded search. All we need to do is write down a primitive recursive *specification* of the object (number of the concatenated sequence) we are looking for, and a bound on how far to look. The following works:

$$\begin{aligned} \text{concat}(s, t) = (&\min v < \text{sequenceBound}(s + t, \text{len}(s) + \text{len}(t))) \\ &(\text{len}(v) = \text{len}(s) + \text{len}(t) \wedge \\ &\quad (\forall i < \text{len}(s))\ ((v)_i = (s)_i) \wedge \\ &\quad (\forall j < \text{len}(t))\ ((v)_{\text{len}(s)+j} = (t)_j)) \end{aligned}$$

Proposition 9.24. *The function* subseq(s, i, n) *which returns the subsequence of s of length n beginning at the ith element, is primitive recursive.*

Proof. Exercise. □

9.12 Trees

Sometimes it is useful to represent trees as natural numbers, just like we can represent sequences by numbers and properties of and operations on them by primitive recursive relations and functions on their codes. We'll use sequences and their codes to do this. A tree can be either a single node (possibly with a label) or else a node (possibly with a label) connected to a number of subtrees. The node is called the *root* of the tree, and the subtrees it is connected to its *immediate subtrees*.

We code trees recursively as a sequence $\langle k, d_1, \ldots, d_k \rangle$, where k is the number of immediate subtrees and $d_1, \ldots, d_k$ the codes of the immediate subtrees. If the nodes have labels, they can be included after the immediate subtrees. So a tree consisting just of a single node with label l would be coded by $\langle 0, l \rangle$, and a tree consisting of a root (labelled l_1) connected to two single nodes (labelled l_2, l_3) would be coded by $\langle 2, \langle 0, l_2 \rangle, \langle 0, l_3 \rangle, l_1 \rangle$.

Proposition 9.25. *The function* SubtreeSeq(t)*, which returns the code of a sequence the elements of which are the codes of all subtrees of the tree with code t, is primitive recursive.*

Proof. First note that ISubtrees$(t) = \text{subseq}(t, 1, (t)_0)$ is primitive recursive and returns the codes of the immediate subtrees of a tree t. Now we can define a helper function hSubtreeSeq(t, n) which computes the sequence of all subtrees which are n nodes removed from the root. The sequence of subtrees of t which is 0 nodes removed from the root—in other words, begins at the root of t—is the sequence consisting just of t. To obtain a sequence of all level $n+1$ subtrees of t, we concatenate the level n subtrees with a sequence consisting of all immediate subtrees of the level n subtrees. To get a list of all these, note that if $f(x)$ is a primitive recursive function returning codes of sequences, then $g_f(s, k) = f((s)_0) \frown \ldots \frown f((s)_k)$ is also primitive recursive:

$$\begin{aligned} g(s, 0) &= f((s)_0) \\ g(s, k+1) &= g(s, k) \frown f((s)_{k+1}) \end{aligned}$$

For instance, if s is a sequence of trees, then $h(s) = g_{\text{ISubtrees}}(s, \text{len}(s))$ gives the sequence of the immediate subtrees of the elements of s. We can use it to define hSubtreeSeq by

$$\begin{aligned} \text{hSubtreeSeq}(t, 0) &= \langle t \rangle \\ \text{hSubtreeSeq}(t, n+1) &= \text{hSubtreeSeq}(t, n) \frown h(\text{hSubtreeSeq}(t, n)). \end{aligned}$$

The maximum level of subtrees in a tree coded by t, i.e., the maximum distance between the root and a leaf node, is bounded by the code t. So a sequence of codes of all subtrees of the tree coded by t is given by hSubtreeSeq(t, t).□

9.13 Other Recursions

Using pairing and sequencing, we can justify more exotic (and useful) forms of primitive recursion. For example, it is often useful to define two functions simultaneously, such as in the following definition:

$$\begin{aligned} h_0(\vec{x}, 0) &= f_0(\vec{x}) \\ h_1(\vec{x}, 0) &= f_1(\vec{x}) \\ h_0(\vec{x}, y+1) &= g_0(\vec{x}, y, h_0(\vec{x}, y), h_1(\vec{x}, y)) \\ h_1(\vec{x}, y+1) &= g_1(\vec{x}, y, h_0(\vec{x}, y), h_1(\vec{x}, y)) \end{aligned}$$

This is an instance of *simultaneous recursion*. Another useful way of defining functions is to give the value of $h(\vec{x}, y+1)$ in terms of *all* the values $h(\vec{x}, 0)$, $\ldots$, $h(\vec{x}, y)$, as in the following definition:

$$\begin{aligned} h(\vec{x}, 0) &= f(\vec{x}) \\ h(\vec{x}, y+1) &= g(\vec{x}, y, \langle h(\vec{x}, 0), \ldots, h(\vec{x}, y)\rangle). \end{aligned}$$

The following schema captures this idea more succinctly:

$$h(\vec{x}, y) = g(\vec{x}, y, \langle h(\vec{x}, 0), \ldots, h(\vec{x}, y-1)\rangle)$$

with the understanding that the last argument to g is just the empty sequence when y is 0. In either formulation, the idea is that in computing the "successor step," the function h can make use of the entire sequence of values computed so far. This is known as a *course-of-values* recursion. For a particular example, it can be used to justify the following type of definition:

$$h(\vec{x}, y) = \begin{cases} g(\vec{x}, y, h(\vec{x}, k(\vec{x}, y))) & \text{if } k(\vec{x}, y) < y \\ f(\vec{x}) & \text{otherwise} \end{cases}$$

In other words, the value of h at y can be computed in terms of the value of h at *any* previous value, given by k.

You should think about how to obtain these functions using ordinary primitive recursion. One final version of primitive recursion is more flexible in that one is allowed to change the *parameters* (side values) along the way:

$$\begin{aligned} h(\vec{x}, 0) &= f(\vec{x}) \\ h(\vec{x}, y+1) &= g(\vec{x}, y, h(k(\vec{x}), y)) \end{aligned}$$

This, too, can be simulated with ordinary primitive recursion. (Doing so is tricky. For a hint, try unwinding the computation by hand.)

9.14 Non-Primitive Recursive Functions

The primitive recursive functions do not exhaust the intuitively computable functions. It should be intuitively clear that we can make a list of all the unary primitive recursive functions, f_0, f_1, f_2, ... such that we can effectively compute the value of f_x on input y; in other words, the function $g(x, y)$, defined by

$$g(x, y) = f_x(y)$$

is computable. But then so is the function

$$\begin{aligned} h(x) &= g(x, x) + 1 \\ &= f_x(x) + 1. \end{aligned}$$

For each primitive recursive function f_i, the value of h and f_i differ at i. So h is computable, but not primitive recursive; and one can say the same about g. This is an "effective" version of Cantor's diagonalization argument.

One can provide more explicit examples of computable functions that are not primitive recursive. For example, let the notation $g^n(x)$ denote $g(g(\ldots g(x)))$, with n g's in all; and define a sequence $g_0, g_1, \ldots$ of functions by

$$\begin{aligned} g_0(x) &= x + 1 \\ g_{n+1}(x) &= g_n^x(x) \end{aligned}$$

You can confirm that each function g_n is primitive recursive. Each successive function grows much faster than the one before; $g_1(x)$ is equal to $2x$, $g_2(x)$ is equal to $2^x \cdot x$, and $g_3(x)$ grows roughly like an exponential stack of x 2's. The Ackermann–Péter function is essentially the function $G(x) = g_x(x)$, and one can show that this grows faster than any primitive recursive function.

Let us return to the issue of enumerating the primitive recursive functions. Remember that we have assigned symbolic notations to each primitive recursive function; so it suffices to enumerate notations. We can assign a natural number $\#(F)$ to each notation F, recursively, as follows:

$$\begin{aligned} \#(0) &= \langle 0 \rangle \\ \#(S) &= \langle 1 \rangle \\ \#(P_i^n) &= \langle 2, n, i \rangle \\ \#(\mathrm{Comp}_{k,l}[H, G_0, \ldots, G_{k-1}]) &= \langle 3, k, l, \#(H), \#(G_0), \ldots, \#(G_{k-1}) \rangle \\ \#(\mathrm{Rec}_l[G, H]) &= \langle 4, l, \#(G), \#(H) \rangle \end{aligned}$$

Here we are using the fact that every sequence of numbers can be viewed as a natural number, using the codes from the last section. The upshot is that every code is assigned a natural number. Of course, some sequences (and hence some numbers) do not correspond to notations; but we can let f_i be the unary primitive recursive function with notation coded as i, if i codes such a

notation; and the constant 0 function otherwise. The net result is that we have an explicit way of enumerating the unary primitive recursive functions.

(In fact, some functions, like the constant zero function, will appear more than once on the list. This is not just an artifact of our coding, but also a result of the fact that the constant zero function has more than one notation. We will later see that one can not computably avoid these repetitions; for example, there is no computable function that decides whether or not a given notation represents the constant zero function.)

We can now take the function $g(x, y)$ to be given by $f_x(y)$, where f_x refers to the enumeration we have just described. How do we know that $g(x, y)$ is computable? Intuitively, this is clear: to compute $g(x, y)$, first "unpack" x, and see if it is a notation for a unary function. If it is, compute the value of that function on input y.

You may already be convinced that (with some work!) one can write a program (say, in Java or C++) that does this; and now we can appeal to the Church-Turing thesis, which says that anything that, intuitively, is computable can be computed by a Turing machine.

Of course, a more direct way to show that $g(x, y)$ is computable is to describe a Turing machine that computes it, explicitly. This would, in particular, avoid the Church-Turing thesis and appeals to intuition. Soon we will have built up enough machinery to show that $g(x, y)$ is computable, appealing to a model of computation that can be *simulated* on a Turing machine: namely, the recursive functions.

9.15 Partial Recursive Functions

To motivate the definition of the recursive functions, note that our proof that there are computable functions that are not primitive recursive actually establishes much more. The argument was simple: all we used was the fact that it is possible to enumerate functions $f_0, f_1, \ldots$ such that, as a function of x and y, $f_x(y)$ is computable. So the argument applies to *any class of functions that can be enumerated in such a way*. This puts us in a bind: we would like to describe the computable functions explicitly; but any explicit description of a collection of computable functions cannot be exhaustive!

The way out is to allow *partial* functions to come into play. We will see that it *is* possible to enumerate the partial computable functions. In fact, we already pretty much know that this is the case, since it is possible to enumerate Turing machines in a systematic way. We will come back to our diagonal argument later, and explore why it does not go through when partial functions are included.

The question is now this: what do we need to add to the primitive recursive functions to obtain all the partial recursive functions? We need to do two things:

1. Modify our definition of the primitive recursive functions to allow for partial functions as well.

2. *Add* something to the definition, so that some new partial functions are included.

The first is easy. As before, we will start with zero, successor, and projections, and close under composition and primitive recursion. The only difference is that we have to modify the definitions of composition and primitive recursion to allow for the possibility that some of the terms in the definition are not defined. If f and g are partial functions, we will write $f(x) \downarrow$ to mean that f is defined at x, i.e., x is in the domain of f; and $f(x) \uparrow$ to mean the opposite, i.e., that f is not defined at x. We will use $f(x) \simeq g(x)$ to mean that either $f(x)$ and $g(x)$ are both undefined, or they are both defined and equal. We will use these notations for more complicated terms as well. We will adopt the convention that if h and $g_0, \ldots, g_k$ all are partial functions, then

$$h(g_0(\vec{x}), \ldots, g_k(\vec{x}))$$

is defined if and only if each g_i is defined at $\vec{x}$, and h is defined at $g_0(\vec{x})$, $\ldots$, $g_k(\vec{x})$. With this understanding, the definitions of composition and primitive recursion for partial functions is just as above, except that we have to replace "$=$" by "$\simeq$".

What we will add to the definition of the primitive recursive functions to obtain partial functions is the *unbounded search operator*. If $f(x, \vec{z})$ is any partial function on the natural numbers, define $\mu x\ f(x, \vec{z})$ to be

> the least x such that $f(0, \vec{z}), f(1, \vec{z}), \ldots, f(x, \vec{z})$ are all defined, and $f(x, \vec{z}) = 0$, if such an x exists

with the understanding that $\mu x\ f(x, \vec{z})$ is undefined otherwise. This defines $\mu x\ f(x, \vec{z})$ uniquely.

Note that our definition makes no reference to Turing machines, or algorithms, or any specific computational model. But like composition and primitive recursion, there is an operational, computational intuition behind unbounded search. When it comes to the computability of a partial function, arguments where the function is undefined correspond to inputs for which the computation does not halt. The procedure for computing $\mu x\ f(x, \vec{z})$ will amount to this: compute $f(0, \vec{z}), f(1, \vec{z}), f(2, \vec{z})$ until a value of 0 is returned. If any of the intermediate computations do not halt, however, neither does the computation of $\mu x\ f(x, \vec{z})$.

If $R(x, \vec{z})$ is any relation, $\mu x\ R(x, \vec{z})$ is defined to be $\mu x\ (1 \dot{-} \chi_R(x, \vec{z}))$. In other words, $\mu x\ R(x, \vec{z})$ returns the least value of x such that $R(x, \vec{z})$ holds. So, if $f(x, \vec{z})$ is a total function, $\mu x\ f(x, \vec{z})$ is the same as $\mu x\ (f(x, \vec{z}) = 0)$. But note that our original definition is more general, since it allows for the possibility

that $f(x, \vec{z})$ is not everywhere defined (whereas, in contrast, the characteristic function of a relation is always total).

Definition 9.26. The set of *partial recursive functions* is the smallest set of partial functions from the natural numbers to the natural numbers (of various arities) containing zero, successor, and projections, and closed under composition, primitive recursion, and unbounded search.

Of course, some of the partial recursive functions will happen to be total, i.e., defined for every argument.

Definition 9.27. The set of *recursive functions* is the set of partial recursive functions that are total.

A recursive function is sometimes called "total recursive" to emphasize that it is defined everywhere.

9.16 The Normal Form Theorem

Theorem 9.28 (Kleene's Normal Form Theorem). *There is a primitive recursive relation $T(e, x, s)$ and a primitive recursive function $U(s)$, with the following property: if f is any partial recursive function, then for some e,*

$$f(x) \simeq U(\mu s\ T(e, x, s))$$

for every x.

The proof of the normal form theorem is involved, but the basic idea is simple. Every partial recursive function has an *index* e, intuitively, a number coding its program or definition. If $f(x) \downarrow$, the computation can be recorded systematically and coded by some number s, and the fact that s codes the computation of f on input x can be checked primitive recursively using only x and the definition e. Consequently, the relation T, "the function with index e has a computation for input x, and s codes this computation," is primitive recursive. Given the full record of the computation s, the "upshot" of s is the value of $f(x)$, and it can be obtained from s primitive recursively as well.

The normal form theorem shows that only a single unbounded search is required for the definition of any partial recursive function. Basically, we can search through all numbers until we find one that codes a computation of the function with index e for input x. We can use the numbers e as "names" of partial recursive functions, and write φ_e for the function f defined by the equation in the theorem. Note that any partial recursive function can have more than one index—in fact, every partial recursive function has infinitely many indices.

9.17 The Halting Problem

The *halting problem* in general is the problem of deciding, given the specification e (e.g., program) of a computable function and a number n, whether the computation of the function on input n halts, i.e., produces a result. Famously, Alan Turing proved that this problem itself cannot be solved by a computable function, i.e., the function

$$h(e,n) = \begin{cases} 1 & \text{if computation } e \text{ halts on input } n \\ 0 & \text{otherwise,} \end{cases}$$

is not computable.

In the context of partial recursive functions, the role of the specification of a program may be played by the index e given in Kleene's normal form theorem. If f is a partial recursive function, any e for which the equation in the normal form theorem holds, is an index of f. Given a number e, the normal form theorem states that

$$\varphi_e(x) \simeq U(\mu s\ T(e,x,s))$$

is partial recursive, and for every partial recursive $f\colon \mathbb{N} \to \mathbb{N}$, there is an $e \in \mathbb{N}$ such that $\varphi_e(x) \simeq f(x)$ for all $x \in \mathbb{N}$. In fact, for each such f there is not just one, but infinitely many such e. The *halting function* h is defined by

$$h(e,x) = \begin{cases} 1 & \text{if } \varphi_e(x) \downarrow \\ 0 & \text{otherwise.} \end{cases}$$

Note that $h(e,x) = 0$ if $\varphi_e(x) \uparrow$, but also when e is not the index of a partial recursive function at all.

Theorem 9.29. *The halting function h is not partial recursive.*

Proof. If h were partial recursive, we could define

$$d(y) = \begin{cases} 1 & \text{if } h(y,y) = 0 \\ \mu x\ x \neq x & \text{otherwise.} \end{cases}$$

Since no number x satisfies $x \neq x$, there is no $\mu x\ x \neq x$, and so $d(y) \uparrow$ iff $h(y,y) \neq 0$. From this definition it follows that

1. $d(y) \downarrow$ iff $\varphi_y(y) \uparrow$ or y is not the index of a partial recursive function.
2. $d(y) \uparrow$ iff $\varphi_y(y) \downarrow$.

If h were partial recursive, then d would be partial recursive as well. Thus, by the Kleene normal form theorem, it has an index e_d. Consider the value of $h(e_d, e_d)$. There are two possible cases, 0 and 1.

1. If $h(e_d, e_d) = 1$ then $\varphi_{e_d}(e_d) \downarrow$. But $\varphi_{e_d} \simeq d$, and $d(e_d)$ is defined iff $h(e_d, e_d) = 0$. So $h(e_d, e_d) \neq 1$.

2. If $h(e_d, e_d) = 0$ then either e_d is not the index of a partial recursive function, or it is and $\varphi_{e_d}(e_d) \uparrow$. But again, $\varphi_{e_d} \simeq d$, and $d(e_d)$ is undefined iff $\varphi_{e_d}(e_d) \downarrow$.

The upshot is that e_d cannot, after all, be the index of a partial recursive function. But if h were partial recursive, d would be too, and so our definition of e_d as an index of it would be admissible. We must conclude that h cannot be partial recursive. □

9.18 General Recursive Functions

There is another way to obtain a set of total functions. Say a total function $f(x, \vec{z})$ is *regular* if for every sequence of natural numbers $\vec{z}$, there is an x such that $f(x, \vec{z}) = 0$. In other words, the regular functions are exactly those functions to which one can apply unbounded search, and end up with a total function. One can, conservatively, restrict unbounded search to regular functions:

Definition 9.30. The set of *general recursive functions* is the smallest set of functions from the natural numbers to the natural numbers (of various arities) containing zero, successor, and projections, and closed under composition, primitive recursion, and unbounded search applied to *regular* functions.

Clearly every general recursive function is total. The difference between Definition 9.30 and Definition 9.27 is that in the latter one is allowed to use partial recursive functions along the way; the only requirement is that the function you end up with at the end is total. So the word "general," a historic relic, is a misnomer; on the surface, Definition 9.30 is *less* general than Definition 9.27. But, fortunately, the difference is illusory; though the definitions are different, the set of general recursive functions and the set of recursive functions are one and the same.

Chapter 10

Arithmetization of Syntax

10.1 Introduction

In order to connect computability and logic, we need a way to talk about the objects of logic (symbols, terms, formulas, derivations), operations on them, and their properties and relations, in a way amenable to computational treatment. We can do this directly, by considering computable functions and relations on symbols, sequences of symbols, and other objects built from them. Since the objects of logical syntax are all finite and built from an enumerable sets of symbols, this is possible for some models of computation. But other models of computation—such as the recursive functions—-are restricted to numbers, their relations and functions. Moreover, ultimately we also want to be able to deal with syntax within certain theories, specifically, in theories formulated in the language of arithmetic. In these cases it is necessary to *arithmetize* syntax, i.e., to represent syntactic objects, operations on them, and their relations, as numbers, arithmetical functions, and arithmetical relations, respectively. The idea, which goes back to Leibniz, is to assign numbers to syntactic objects.

It is relatively straightforward to assign numbers to symbols as their "codes." Some symbols pose a bit of a challenge, since, e.g., there are infinitely many variables, and even infinitely many function symbols of each arity n. But of course it's possible to assign numbers to symbols systematically in such a way that, say, v_2 and v_3 are assigned different codes. Sequences of symbols (such as terms and formulas) are a bigger challenge. But if we can deal with sequences of numbers purely arithmetically (e.g., by the powers-of-primes coding of sequences), we can extend the coding of individual symbols to coding of sequences of symbols, and then further to sequences or other arrangements of formulas, such as derivations. This extended coding is called "Gödel numbering." Every term, formula, and derivation is assigned a Gödel number.

By coding sequences of symbols as sequences of their codes, and by chosing a system of coding sequences that can be dealt with using computable

functions, we can then also deal with Gödel numbers using computable functions. In practice, all the relevant functions will be primitive recursive. For instance, computing the length of a sequence and computing the i-th element of a sequence from the code of the sequence are both primitive recursive. If the number coding the sequence is, e.g., the Gödel number of a formula φ, we immediately see that the length of a formula and the (code of the) i-th symbol in a formula can also be computed from the Gödel number of φ. It is a bit harder to prove that, e.g., the property of being the Gödel number of a correctly formed term or of a correct derivation is primitive recursive. It is nevertheless possible, because the sequences of interest (terms, formulas, derivations) are inductively defined.

As an example, consider the operation of substitution. If φ is a formula, x a variable, and t a term, then $\varphi[t/x]$ is the result of replacing every free occurrence of x in φ by t. Now suppose we have assigned Gödel numbers to φ, x, t—say, k, l, and m, respectively. The same scheme assigns a Gödel number to $\varphi[t/x]$, say, n. This mapping—of k, l, and m to n—is the arithmetical analog of the substitution operation. When the substitution operation maps φ, x, t to $\varphi[t/x]$, the arithmetized substitution functions maps the Gödel numbers k, l, m to the Gödel number n. We will see that this function is primitive recursive.

Arithmetization of syntax is not just of abstract interest, although it was originally a non-trivial insight that languages like the language of arithmetic, which do not come with mechanisms for "talking about" languages can, after all, formalize complex properties of expressions. It is then just a small step to ask what a theory in this language, such as Peano arithmetic, can *prove* about its own language (including, e.g., whether sentences are provable or true). This leads us to the famous limitative theorems of Gödel (about unprovability) and Tarski (the undefinability of truth). But the trick of arithmetizing syntax is also important in order to prove some important results in computability theory, e.g., about the computational power of theories or the relationship between different models of computability. The arithmetization of syntax serves as a model for arithmetizing other objects and properties. For instance, it is similarly possible to arithmetize configurations and computations (say, of Turing machines). This makes it possible to simulate computations in one model (e.g., Turing machines) in another (e.g., recursive functions).

10.2 Coding Symbols

The basic language $\mathcal{L}$ of first order logic makes use of the symbols

$$\bot \quad \neg \quad \vee \quad \wedge \quad \rightarrow \quad \forall \quad \exists \quad = \quad (\quad) \quad ,$$

together with enumerable sets of variables and constant symbols, and enumerable sets of function symbols and predicate symbols of arbitrary arity. We can assign *codes* to each of these symbols in such a way that every symbol is

assigned a unique number as its code, and no two different symbols are assigned the same number. We know that this is possible since the set of all symbols is enumerable and so there is a bijection between it and the set of natural numbers. But we want to make sure that we can recover the symbol (as well as some information about it, e.g., the arity of a function symbol) from its code in a computable way. There are many possible ways of doing this, of course. Here is one such way, which uses primitive recursive functions. (Recall that $\langle n_0, \ldots, n_k \rangle$ is the number coding the sequence of numbers $n_0, \ldots, n_k$.)

Definition 10.1. If s is a symbol of $\mathcal{L}$, let the *symbol code* c_s be defined as follows:

1. If s is among the logical symbols, c_s is given by the following table:

$$\begin{array}{cccccc} \bot & \neg & \vee & \wedge & \to & \forall \\ \langle 0,0\rangle & \langle 0,1\rangle & \langle 0,2\rangle & \langle 0,3\rangle & \langle 0,4\rangle & \langle 0,5\rangle \\ \exists & = & (&) & , & \\ \langle 0,6\rangle & \langle 0,7\rangle & \langle 0,8\rangle & \langle 0,9\rangle & \langle 0,10\rangle & \end{array}$$

2. If s is the i-th variable v_i, then $\mathrm{c}_s = \langle 1, i\rangle$.
3. If s is the i-th constant symbol c_i, then $\mathrm{c}_s = \langle 2, i\rangle$.
4. If s is the i-th n-ary function symbol f_i^n, then $\mathrm{c}_s = \langle 3, n, i\rangle$.
5. If s is the i-th n-ary predicate symbol P_i^n, then $\mathrm{c}_s = \langle 4, n, i\rangle$.

Proposition 10.2. *The following relations are primitive recursive:*

1. $\mathrm{Fn}(x, n)$ *iff* x *is the code of* f_i^n *for some* i*, i.e.,* x *is the code of an* n*-ary function symbol.*
2. $\mathrm{Pred}(x, n)$ *iff* x *is the code of* P_i^n *for some* i *or* x *is the code of* $=$ *and* $n = 2$*, i.e.,* x *is the code of an* n*-ary predicate symbol.*

Definition 10.3. If $s_0, \ldots, s_{n-1}$ is a sequence of symbols, its *Gödel number* is $\langle \mathrm{c}_{s_0}, \ldots, \mathrm{c}_{s_{n-1}} \rangle$.

Note that *codes* and *Gödel numbers* are different things. For instance, the variable v_5 has a code $\mathrm{c}_{v_5} = \langle 1, 5\rangle = 2^2 \cdot 3^6$. But the variable v_5 considered as a term is also a sequence of symbols (of length 1). The *Gödel number* $\ulcorner v_5 \urcorner$ of the *term* v_5 is $\langle \mathrm{c}_{v_5} \rangle = 2^{\mathrm{c}_{v_5}+1} = 2^{2^2 \cdot 3^6 + 1}$.

Example 10.4. Recall that if $k_0, \ldots, k_{n-1}$ is a sequence of numbers, then the code of the sequence $\langle k_0, \ldots, k_{n-1}\rangle$ in the power-of-primes coding is

$$2^{k_0+1} \cdot 3^{k_1+1} \cdot \ldots \cdot p_{n-1}^{k_{n-1}},$$

where p_i is the i-th prime (starting with $p_0 = 2$). So for instance, the formula $v_0 = 0$, or, more explicitly, $=(v_0, c_0)$, has the Gödel number

$$\langle \mathrm{c}_=, \mathrm{c}_(, \mathrm{c}_{v_0}, \mathrm{c}_,, \mathrm{c}_{c_0}, \mathrm{c}_) \rangle.$$

Here, $\mathrm{c}_=$ is $\langle 0, 7 \rangle = 2^{0+1} \cdot 3^{7=1}$, c_{v_0} is $\langle 1, 0 \rangle = 2^{1+1} \cdot 3^{0+1}$, etc. So ${}^{\ulcorner}{=}(v_0, c_0)^{\urcorner}$ is

$$\begin{aligned} &2^{\mathrm{c}_=+1} \cdot 3^{\mathrm{c}_(+1} \cdot 5^{\mathrm{c}_{v_0}+1} \cdot 7^{\mathrm{c}_,+1} \cdot 11^{\mathrm{c}_{c_0}+1} \cdot 13^{\mathrm{c}_)+1} = \\ &\qquad 2^{2^1 \cdot 3^8+1} \cdot 3^{2^1 \cdot 3^9+1} \cdot 5^{2^2 \cdot 3^1+1} \cdot 7^{2^1 \cdot 3^{11}+1} \cdot 11^{2^3 \cdot 3^1+1} \cdot 13^{2^1 \cdot 3^{10}+1} = \\ &\qquad\qquad 2^{13\,123} \cdot 3^{39\,367} \cdot 5^{13} \cdot 7^{354\,295} \cdot 11^{25} \cdot 13^{118\,099}. \end{aligned}$$

10.3 Coding Terms

A term is simply a certain kind of sequence of symbols: it is built up inductively from constants and variables according to the formation rules for terms. Since sequences of symbols can be coded as numbers—using a coding scheme for the symbols plus a way to code sequences of numbers—assigning Gödel numbers to terms is not difficult. The challenge is rather to show that the property a number has if it is the Gödel number of a correctly formed term is computable, or in fact primitive recursive.

Variables and constant symbols are the simplest terms, and testing whether x is the Gödel number of such a term is easy: $\mathrm{Var}(x)$ holds if x is ${}^{\ulcorner}v_i{}^{\urcorner}$ for some i. In other words, x is a sequence of length 1 and its single element $(x)_0$ is the code of some variable v_i, i.e., x is $\langle\langle 1, i \rangle\rangle$ for some i. Similarly, $\mathrm{Const}(x)$ holds if x is ${}^{\ulcorner}c_i{}^{\urcorner}$ for some i. Both of these relations are primitive recursive, since if such an i exists, it must be $< x$:

$$\begin{aligned} \mathrm{Var}(x) &\Leftrightarrow (\exists i < x)\ x = \langle\langle 1, i \rangle\rangle \\ \mathrm{Const}(x) &\Leftrightarrow (\exists i < x)\ x = \langle\langle 2, i \rangle\rangle \end{aligned}$$

Proposition 10.5. *The relations* $\mathrm{Term}(x)$ *and* $\mathrm{ClTerm}(x)$ *which hold iff x is the Gödel number of a term or a closed term, respectively, are primitive recursive.*

Proof. A sequence of symbols s is a term iff there is a sequence $s_0, \ldots, s_{k-1} = s$ of terms which records how the term s was formed from constant symbols and variables according to the formation rules for terms. To express that such a putative formation sequence follows the formation rules it has to be the case that, for each $i < k$, either

1. s_i is a variable v_j, or
2. s_i is a constant symbol c_j, or
3. s_i is built from n terms $t_1, \ldots, t_n$ occurring prior to place i using an n-place function symbol f^n_j.

To show that the corresponding relation on Gödel numbers is primitive recursive, we have to express this condition primitive recursively, i.e., using primitive recursive functions, relations, and bounded quantification.

Suppose y is the number that codes the sequence $s_0, \ldots, s_{k-1}$, i.e., $y = \langle {}^{\ulcorner}s_0{}^{\urcorner}, \ldots, {}^{\ulcorner}s_{k-1}{}^{\urcorner}\rangle$. It codes a formation sequence for the term with Gödel number x iff for all $i < k$:

1. $\mathrm{Var}((y)_i)$, or
2. $\mathrm{Const}((y)_i)$, or
3. there is an n and a number $z = \langle z_1, \ldots, z_n\rangle$ such that each z_l is equal to some $(y)_{i'}$ for $i' < i$ and

$$(y)_i = {}^{\ulcorner}f_j^n({}^{\urcorner} \frown \mathrm{flatten}(z) \frown {}^{\ulcorner})^{\urcorner},$$

and moreover $(y)_{k-1} = x$. (The function $\mathrm{flatten}(z)$ turns the sequence $\langle {}^{\ulcorner}t_1{}^{\urcorner}, \ldots, {}^{\ulcorner}t_n{}^{\urcorner}\rangle$ into ${}^{\ulcorner}t_1, \ldots, t_n{}^{\urcorner}$ and is primitive recursive.)

The indices j, n, the Gödel numbers z_l of the terms t_l, and the code z of the sequence $\langle z_1, \ldots, z_n\rangle$, in (3) are all less than y. We can replace k above with $\mathrm{len}(y)$. Hence we can express "y is the code of a formation sequence of the term with Gödel number x" in a way that shows that this relation is primitive recursive.

We now just have to convince ourselves that there is a primitive recursive bound on y. But if x is the Gödel number of a term, it must have a formation sequence with at most $\mathrm{len}(x)$ terms (since every term in the formation sequence of s must start at some place in s, and no two subterms can start at the same place). The Gödel number of each subterm of s is of course $\leq x$. Hence, there always is a formation sequence with code $\leq p_{k-1}^{k(x+1)}$, where $k = \mathrm{len}(x)$.

For ClTerm, simply leave out the clause for variables. □

Proposition 10.6. *The function* $\mathrm{num}(n) = {}^{\ulcorner}\overline{n}{}^{\urcorner}$ *is primitive recursive.*

Proof. We define $\mathrm{num}(n)$ by primitive recursion:

$$\begin{aligned}\mathrm{num}(0) &= {}^{\ulcorner}\mathsf{o}{}^{\urcorner}\\ \mathrm{num}(n+1) &= {}^{\ulcorner}\mathsf{\prime}({}^{\urcorner} \frown \mathrm{num}(n) \frown {}^{\ulcorner})^{\urcorner}.\end{aligned}$$

□

10.4 Coding Formulas

Proposition 10.7. *The relation* $\mathrm{Atom}(x)$ *which holds iff* x *is the Gödel number of an atomic formula, is primitive recursive.*

Proof. The number x is the Gödel number of an atomic formula iff one of the following holds:

1. There are n, $j < x$, and $z < x$ such that for each $i < n$, $\mathrm{Term}((z)_i)$ and $x =$
$$^{\#}P^n_j(^{\#} \frown \mathrm{flatten}(z) \frown {}^{\#})^{\#}.$$
2. There are $z_1, z_2 < x$ such that $\mathrm{Term}(z_1)$, $\mathrm{Term}(z_2)$, and $x =$
$$^{\#}{=}(^{\#} \frown z_1 \frown {}^{\#},{}^{\#} \frown z_2 \frown {}^{\#})^{\#}.$$
3. $x = {}^{\#}\bot{}^{\#}$. □

Proposition 10.8. *The relation* $\mathrm{Frm}(x)$ *which holds iff x is the Gödel number of a formula is primitive recursive.*

Proof. A sequence of symbols s is a formula iff there is formation sequence s_0, $\ldots$, $s_{k-1} = s$ of formula which records how s was formed from atomic formulas according to the formation rules. The code for each s_i (and indeed of the code of the sequence $\langle s_0, \ldots, s_{k-1}\rangle$) is less than the code x of s. □

Proposition 10.9. *The relation* $\mathrm{FreeOcc}(x, z, i)$*, which holds iff the i-th symbol of the formula with Gödel number x is a free occurrence of the variable with Gödel number z, is primitive recursive.*

Proof. Exercise. □

Proposition 10.10. *The property* $\mathrm{Sent}(x)$ *which holds iff x is the Gödel number of a sentence is primitive recursive.*

Proof. A sentence is a formula without free occurrences of variables. So $\mathrm{Sent}(x)$ holds iff

$$(\forall i < \mathrm{len}(x))\ (\forall z < x)\ ((\exists j < z)\ z = {}^{\#}v_j{}^{\#} \to \neg\mathrm{FreeOcc}(x, z, i)).$$ □

10.5 Substitution

Recall that substitution is the operation of replacing all free occurrences of a variable u in a formula φ by a term t, written $\varphi[t/u]$. This operation, when carried out on Gödel numbers of variables, formulas, and terms, is primitive recursive.

Proposition 10.11. *There is a primitive recursive function* $\mathrm{Subst}(x, y, z)$ *with the property that*
$$\mathrm{Subst}({}^{\#}\varphi^{\#}, {}^{\#}t^{\#}, {}^{\#}u^{\#}) = {}^{\#}\varphi[t/u]^{\#}.$$

Proof. We can then define a function hSubst by primitive recursion as follows:

$$\begin{aligned}\mathrm{hSubst}(x, y, z, 0) &= \Lambda \\ \mathrm{hSubst}(x, y, z, i+1) &= \\ &\begin{cases} \mathrm{hSubst}(x, y, z, i) \frown y & \text{if } \mathrm{FreeOcc}(x, z, i) \\ \mathrm{append}(\mathrm{hSubst}(x, y, z, i), (x)_i) & \text{otherwise.} \end{cases}\end{aligned}$$

$\mathrm{Subst}(x, y, z)$ can now be defined as $\mathrm{hSubst}(x, y, z, \mathrm{len}(x))$. □

Proposition 10.12. *The relation* $\mathrm{FreeFor}(x, y, z)$*, which holds iff the term with Gödel number* y *is free for the variable with Gödel number* z *in the formula with Gödel number* x*, is primitive recursive.*

Proof. Exercise. □

10.6 Derivations in LK

In order to arithmetize derivations, we must represent derivations as numbers. Since derivations are trees of sequents where each inference carries also a label, a recursive representation is the most obvious approach: we represent a derivation as a tuple, the components of which are the end-sequent, the label, and the representations of the sub-derivations leading to the premises of the last inference.

Definition 10.13. If Γ is a finite sequence of sentences, $\Gamma = \langle \varphi_1, \ldots, \varphi_n \rangle$, then ${}^{\#}\Gamma^{\#} = \langle {}^{\#}\varphi_1{}^{\#}, \ldots, {}^{\#}\varphi_n{}^{\#} \rangle$.

If $\Gamma \Rightarrow \Delta$ is a sequent, then a Gödel number of $\Gamma \Rightarrow \Delta$ is

$${}^{\#}\Gamma \Rightarrow \Delta^{\#} = \langle {}^{\#}\Gamma^{\#}, {}^{\#}\Delta^{\#} \rangle$$

If π is a derivation in **LK**, then ${}^{\#}\pi^{\#}$ is defined as follows:

1. If π consists only of the initial sequent $\Gamma \Rightarrow \Delta$, then ${}^{\#}\pi^{\#}$ is

   $$\langle 0, {}^{\#}\Gamma \Rightarrow \Delta^{\#} \rangle.$$

2. If π ends in an inference with one or two premises, has $\Gamma \Rightarrow \Delta$ as its conclusion, and π_1 and π_2 are the immediate subproof ending in the premise of the last inference, then ${}^{\#}\pi^{\#}$ is

   $$\begin{aligned}&\langle 1, {}^{\#}\pi_1{}^{\#}, {}^{\#}\Gamma \Rightarrow \Delta^{\#}, k \rangle \text{ or} \\ &\langle 2, {}^{\#}\pi_1{}^{\#}, {}^{\#}\pi_2{}^{\#}, {}^{\#}\Gamma \Rightarrow \Delta^{\#}, k \rangle,\end{aligned}$$

   respectively, where k is given by the following table according to which rule was used in the last inference:

Rule:	WL	WR	CL	CR	XL	XR
k:	1	2	3	4	5	6

Rule:	$\neg$L	$\neg$R	$\wedge$L	$\wedge$R	$\vee$L	$\vee$R
k:	7	8	9	10	11	12

Rule:	$\to$L	$\to$R	$\forall$L	$\forall$R	$\exists$L	$\exists$R
k:	13	14	15	16	17	18

Rule:	Cut	=
k:	19	20

Example 10.14. Consider the very simple derivation

$$\dfrac{\dfrac{\varphi \Rightarrow \varphi}{\varphi \wedge \psi \Rightarrow \varphi}\,\wedge\text{L}}{\Rightarrow (\varphi \wedge \psi) \to \varphi}\,\to\text{R}$$

The Gödel number of the initial sequent would be $p_0 = \langle 0, {}^{\ulcorner}\varphi \Rightarrow \varphi^{\#}\rangle$. The Gödel number of the derivation ending in the conclusion of $\wedge$L would be $p_1 = \langle 1, p_0, {}^{\ulcorner}\varphi \wedge \psi \Rightarrow \varphi^{\#}, 9\rangle$ (1 since $\wedge$L has one premise, the Gödel number of the conclusion $\varphi \wedge \psi \Rightarrow \varphi$, and 9 is the number coding $\wedge$L). The Gödel number of the entire derivation then is $\langle 1, p_1, {}^{\ulcorner}\Rightarrow (\varphi \wedge \psi) \to \varphi)^{\#}, 14\rangle$, i.e.,

$$\langle 1, \langle 1, \langle 0, {}^{\ulcorner}\varphi \Rightarrow \varphi)^{\#}\rangle, {}^{\ulcorner}\varphi \wedge \psi \Rightarrow \varphi^{\#}, 9\rangle, {}^{\ulcorner}\Rightarrow (\varphi \wedge \psi) \to \varphi^{\#}, 14\rangle.$$

Having settled on a representation of derivations, we must also show that we can manipulate such derivations primitive recursively, and express their essential properties and relations so. Some operations are simple: e.g., given a Gödel number p of a derivation, $\text{EndSeq}(p) = (p)_{(p)_0+1}$ gives us the Gödel number of its end-sequent and $\text{LastRule}(p) = (p)_{(p)_0+2}$ the code of its last rule. The property $\text{Sequent}(s)$ defined by

$$\text{len}(s) = 2 \wedge (\forall i < \text{len}((s)_0) + \text{len}((s)_1))\ \text{Sent}(((s)_0 \frown (s)_1)_i)$$

holds of s iff s is the Gödel number of a sequent consisting of sentences. Some are much harder. We'll at least sketch how to do this. The goal is to show that the relation "π is a derivation of φ from Γ" is a primitive recursive relation of the Gödel numbers of π and φ.

Proposition 10.15. *The property* $\text{Correct}(p)$ *which holds iff the last inference in the derivation* π *with Gödel number* p *is correct, is primitive recursive.*

Proof. $\Gamma \Rightarrow \Delta$ is an initial sequent if either there is a sentence φ such that $\Gamma \Rightarrow \Delta$ is $\varphi \Rightarrow \varphi$, or there is a term t such that $\Gamma \Rightarrow \Delta$ is $\emptyset \Rightarrow t = t$. In terms of Gödel numbers, $\text{InitSeq}(s)$ holds iff

$$\begin{aligned}&(\exists x < s)\ (\text{Sent}(x) \wedge s = \langle\langle x\rangle, \langle x\rangle\rangle) \vee \\ &(\exists t < s)\ (\text{Term}(t) \wedge s = \langle 0, \langle {}^{\ulcorner}=({}^{\#} \frown t \frown {}^{\ulcorner}, {}^{\#} \frown t \frown {}^{\ulcorner})^{\#}\rangle\rangle).\end{aligned}$$

We also have to show that for each rule of inference R the relation $\text{FollowsBy}_R(p)$ is primitive recursive, where $\text{FollowsBy}_R(p)$ holds iff p is the Gödel number of derivation π, and the end-sequent of π follows by a correct application of R from the immediate sub-derivations of π.

A simple case is that of the $\wedge$R rule. If π ends in a correct $\wedge$R inference, it looks like this:

$$\dfrac{\begin{array}{cc}\vdots\,\pi_1 & \vdots\,\pi_2\\ \Gamma \Rightarrow \Delta, \varphi & \Gamma \Rightarrow \Delta, \psi\end{array}}{\Gamma \Rightarrow \Delta, \varphi \wedge \psi}\ \wedge\text{R}$$

So, the last inference in the derivation π is a correct application of $\wedge$R iff there are sequences of sentences Γ and Δ as well as two sentences φ and ψ such that the end-sequent of π_1 is $\Gamma \Rightarrow \Delta, \varphi$, the end-sequent of π_2 is $\Gamma \Rightarrow \Delta, \psi$, and the end-sequent of π is $\Gamma \Rightarrow \Delta, \varphi \wedge \psi$. We just have to translate this into Gödel numbers. If $s = {}^{\ulcorner}\Gamma \Rightarrow \Delta^{\urcorner}$ then $(s)_0 = {}^{\ulcorner}\Gamma^{\urcorner}$ and $(s)_1 = {}^{\ulcorner}\Delta^{\urcorner}$. So, $\text{FollowsBy}_{\wedge\text{R}}(p)$ holds iff

$$\begin{aligned}&(\exists g < p)\ (\exists d < p)\ (\exists a < p)\ (\exists b < p)\\ &\quad \text{EndSequent}(p) = \langle g, d \frown \langle {}^{\ulcorner}({}^{\urcorner} \frown a \frown {}^{\ulcorner}\wedge{}^{\urcorner} \frown b \frown {}^{\ulcorner}){}^{\urcorner}\rangle\rangle \wedge\\ &\quad \text{EndSequent}((p)_1) = \langle g, d \frown \langle a\rangle\rangle \wedge\\ &\quad \text{EndSequent}((p)_2) = \langle g, d \frown \langle b\rangle\rangle \wedge\\ &\quad (p)_0 = 2 \wedge \text{LastRule}(p) = 10.\end{aligned}$$

The individual lines express, respectively, "there is a sequence (Γ) with Gödel number g, there is a sequence (Δ) with Gödel number d, a formula (φ) with Gödel number a, and a formula (ψ) with Gödel number b," such that "the end-sequent of π is $\Gamma \Rightarrow \Delta, \varphi \wedge \psi$," "the end-sequent of π_1 is $\Gamma \Rightarrow \Delta, \varphi$," "the end-sequent of π_2 is $\Gamma \Rightarrow \Delta, \psi$," and "$\pi$ has two immediate subderivations and the last inference rule is $\wedge$R (with number 10)."

The last inference in π is a correct application of $\exists$R iff there are sequences Γ and Δ, a formula φ, a variable x, and a term t, such that the end-sequent of π is $\Gamma \Rightarrow \Delta, \exists x\, \varphi$ and the end-sequent of π_1 is $\Gamma \Rightarrow \Delta, \varphi[t/x]$. So in terms of Gödel numbers, we have $\text{FollowsBy}_{\exists\text{R}}(p)$ iff

$$\begin{aligned}&(\exists g < p)\ (\exists d < p)\ (\exists a < p)\ (\exists x < p)\ (\exists t < p)\\ &\quad \text{EndSequent}(p) = \langle g, d \frown \langle {}^{\ulcorner}\exists{}^{\urcorner} \frown x \frown a\rangle\rangle \wedge\\ &\quad \text{EndSequent}((p)_1) = \langle g, d \frown \langle \text{Subst}(a, t, x)\rangle\rangle \wedge\\ &\quad (p)_0 = 1 \wedge \text{LastRule}(p) = 18.\end{aligned}$$

We then define $\text{Correct}(p)$ as

$$\begin{aligned}&\text{Sequent}(\text{EndSequent}(p)) \wedge \\ &\qquad [(\text{LastRule}(p) = 1 \wedge \text{FollowsBy}_{\text{WL}}(p)) \vee \cdots \vee \\ &\qquad\quad (\text{LastRule}(p) = 20 \wedge \text{FollowsBy}_{=}(p)) \vee \\ &\qquad\qquad (p)_0 = 0 \wedge \text{InitialSeq}(\text{EndSequent}(p))]\end{aligned}$$

The first line ensures that the end-sequent of d is actually a sequent consisting of sentences. The last line covers the case where p is just an initial sequent. □

Proposition 10.16. *The relation* $\text{Deriv}(p)$ *which holds if p is the Gödel number of a correct derivation π, is primitive recursive.*

Proof. A derivation π is correct if every one of its inferences is a correct application of a rule, i.e., if every one of its sub-derivations ends in a correct inference. So, $\text{Deriv}(d)$ iff

$$(\forall i < \text{len}(\text{SubtreeSeq}(p)))\ \text{Correct}((\text{SubtreeSeq}(p))_i.$$ □

Proposition 10.17. *Suppose Γ is a primitive recursive set of sentences. Then the relation $\text{Prf}_\Gamma(x, y)$ expressing "x is the code of a derivation π of $\Gamma_0 \Rightarrow \varphi$ for some finite $\Gamma_0 \subseteq \Gamma$ and y is the Gödel number of φ" is primitive recursive.*

Proof. Suppose "$y \in \Gamma$" is given by the primitive recursive predicate $R_\Gamma(y)$. We have to show that $\text{Prf}_\Gamma(x, y)$ which holds iff y is the Gödel number of a sentence φ and x is the code of an **LK**-derivation with end-sequent $\Gamma_0 \Rightarrow \varphi$ is primitive recursive.

By the previous proposition, the property $\text{Deriv}(x)$ which holds iff x is the code of a correct derivation π in **LK** is primitive recursive. If x is such a code, then $\text{EndSequent}(x)$ is the code of the end-sequent of π, and so $(\text{EndSequent}(x))_0$ is the code of the left side of the end sequent and $(\text{EndSequent}(x))_1$ the right side. So we can express "the right side of the end-sequent of π is φ" as $\text{len}((\text{EndSequent}(x))_1) = 1 \wedge ((\text{EndSequent}(x))_1)_0 = x$. The left side of the end-sequent of π is of course automatically finite, we just have to express that every sentence in it is in Γ. Thus we can define $\text{Prf}_\Gamma(x, y)$ by

$$\begin{aligned}\text{Prf}_\Gamma(x, y) \Leftrightarrow\ &\text{Deriv}(x) \wedge \\ &(\forall i < \text{len}((\text{EndSequent}(x))_0))\ R_\Gamma(((\text{EndSequent}(x))_0)_i) \wedge \\ &\text{len}((\text{EndSequent}(x))_1) = 1 \wedge ((\text{EndSequent}(x))_1)_0 = y.\end{aligned}$$ □

Chapter 11

Representability in Q

11.1 Introduction

The incompleteness theorems apply to theories in which basic facts about computable functions can be expressed and proved. We will describe a very minimal such theory called "**Q**" (or, sometimes, "Robinson's Q," after Raphael Robinson). We will say what it means for a function to be *representable* in **Q**, and then we will prove the following:

> A function is representable in **Q** if and only if it is computable.

For one thing, this provides us with another model of computability. But we will also use it to show that the set $\{\varphi : \mathbf{Q} \vdash \varphi\}$ is not decidable, by reducing the halting problem to it. By the time we are done, we will have proved much stronger things than this.

The language of **Q** is the language of arithmetic; **Q** consists of the following axioms (to be used in conjunction with the other axioms and rules of first-order logic with identity predicate):

$$\forall x \, \forall y \, (x' = y' \to x = y) \tag{Q_1}$$
$$\forall x \, \mathsf{o} \neq x' \tag{Q_2}$$
$$\forall x \, (x = \mathsf{o} \vee \exists y \, x = y') \tag{Q_3}$$
$$\forall x \, (x + \mathsf{o}) = x \tag{Q_4}$$
$$\forall x \, \forall y \, (x + y') = (x + y)' \tag{Q_5}$$
$$\forall x \, (x \times \mathsf{o}) = \mathsf{o} \tag{Q_6}$$
$$\forall x \, \forall y \, (x \times y') = ((x \times y) + x) \tag{Q_7}$$
$$\forall x \, \forall y \, (x < y \leftrightarrow \exists z \, (z' + x) = y) \tag{Q_8}$$

For each natural number n, define the numeral $\overline{n}$ to be the term $0^{\prime\prime\cdots\prime}$ where there are n tick marks in all. So, $\overline{0}$ is the constant symbol o by itself, $\overline{1}$ is o', $\overline{2}$ is o'', etc.

As a theory of arithmetic, **Q** is *extremely* weak; for example, you can't even prove very simple facts like $\forall x\, x \neq x'$ or $\forall x \forall y\, (x + y) = (y + x)$. But we will see that much of the reason that **Q** is so interesting is *because* it is so weak. In fact, it is just barely strong enough for the incompleteness theorem to hold. Another reason **Q** is interesting is because it has a *finite* set of axioms.

A stronger theory than **Q** (called *Peano arithmetic* **PA**) is obtained by adding a schema of induction to **Q**:

$$(\varphi(0) \wedge \forall x\, (\varphi(x) \to \varphi(x'))) \to \forall x\, \varphi(x)$$

where $\varphi(x)$ is any formula. If $\varphi(x)$ contains free variables other than x, we add universal quantifiers to the front to bind all of them (so that the corresponding instance of the induction schema is a sentence). For instance, if $\varphi(x, y)$ also contains the variable y free, the corresponding instance is

$$\forall y\, ((\varphi(0) \wedge \forall x\, (\varphi(x) \to \varphi(x'))) \to \forall x\, \varphi(x))$$

Using instances of the induction schema, one can prove much more from the axioms of **PA** than from those of **Q**. In fact, it takes a good deal of work to find "natural" statements about the natural numbers that can't be proved in Peano arithmetic!

Definition 11.1. A function $f(x_0, \ldots, x_k)$ from the natural numbers to the natural numbers is said to be *representable in* **Q** if there is a formula $\varphi_f(x_0, \ldots, x_k, y)$ such that whenever $f(n_0, \ldots, n_k) = m$, **Q** proves

1. $\varphi_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{m})$
2. $\forall y\, (\varphi_f(\overline{n_0}, \ldots, \overline{n_k}, y) \to \overline{m} = y)$.

There are other ways of stating the definition; for example, we could equivalently require that **Q** proves $\forall y\, (\varphi_f(\overline{n_0}, \ldots, \overline{n_k}, y) \leftrightarrow y = \overline{m})$.

Theorem 11.2. *A function is representable in* **Q** *if and only if it is computable.*

There are two directions to proving the theorem. The left-to-right direction is fairly straightforward once arithmetization of syntax is in place. The other direction requires more work. Here is the basic idea: we pick "general recursive" as a way of making "computable" precise, and show that every general recursive function is representable in **Q**. Recall that a function is general recursive if it can be defined from zero, the successor function succ, and the projection functions P^n_i, using composition, primitive recursion, and regular minimization. So one way of showing that every general recursive function is representable in **Q** is to show that the basic functions are representable, and whenever some functions are representable, then so are the functions defined from them using composition, primitive recursion, and regular minimization.

In other words, we might show that the basic functions are representable, and that the representable functions are "closed under" composition, primitive recursion, and regular minimization. This guarantees that every general recursive function is representable.

It turns out that the step where we would show that representable functions are closed under primitive recursion is hard. In order to avoid this step, we show first that in fact we can do without primitive recursion. That is, we show that every general recursive function can be defined from basic functions using composition and regular minimization alone. To do this, we show that primitive recursion can actually be done by a specific regular minimization. However, for this to work, we have to add some additional basic functions: addition, multiplication, and the characteristic function of the identity relation $\chi_=$. Then, we can prove the theorem by showing that all of *these* basic functions are representable in **Q**, and the representable functions are closed under composition and regular minimization.

11.2 Functions Representable in Q are Computable

We'll prove that every function that is representable in **Q** is computable. We first have to establish a lemma about functions representable in **Q**.

Lemma 11.3. *If $f(x_0, \ldots, x_k)$ is representable in* **Q**, *there is a formula $\varphi(x_0, \ldots, x_k, y)$ such that*

$$\mathbf{Q} \vdash \varphi_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{m}) \quad \textit{iff} \quad m = f(n_0, \ldots, n_k).$$

Proof. The "if" part is Definition 11.1(1). The "only if" part is seen as follows: Suppose $\mathbf{Q} \vdash \varphi_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{m})$ but $m \neq f(n_0, \ldots, n_k)$. Let $l = f(n_0, \ldots, n_k)$. By Definition 11.1(1), $\mathbf{Q} \vdash \varphi_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{l})$. By Definition 11.1(2), $\forall y\, (\varphi_f(\overline{n_0}, \ldots, \overline{n_k}, y) \to \overline{l} = y)$. Using logic and the assumption that $\mathbf{Q} \vdash \varphi_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{m})$, we get that $\mathbf{Q} \vdash \overline{l} = \overline{m}$. On the other hand, by Lemma 11.14, $\mathbf{Q} \vdash \overline{l} \neq \overline{m}$. So **Q** is inconsistent. But that is impossible, since **Q** is satisfied by the standard model (see **??**), $\mathfrak{N} \vDash \mathbf{Q}$, and satisfiable theories are always consistent by the Soundness Theorem (Corollary 7.31). □

Lemma 11.4. *Every function that is representable in* **Q** *is computable.*

Proof. Let's first give the intuitive idea for why this is true. To compute f, we do the following. List all the possible derivations δ in the language of arithmetic. This is possible to do mechanically. For each one, check if it is a derivation of a formula of the form $\varphi_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{m})$ (the formula representing f in **Q** from Lemma 11.3). If it is, $m = f(n_0, \ldots, n_k)$ by Lemma 11.3, and we've found the value of f. The search terminates because $\mathbf{Q} \vdash \varphi_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{f(n_0, \ldots, n_k)})$, so eventually we find a δ of the right sort.

This is not quite precise because our procedure operates on derivations and formulas instead of just on numbers, and we haven't explained exactly why "listing all possible derivations" is mechanically possible. But as we've seen, it is possible to code terms, formulas, and derivations by Gödel numbers. We've also introduced a precise model of computation, the general recursive functions. And we've seen that the relation $\mathrm{Prf}_{\mathbf{Q}}(d, y)$, which holds iff d is the Gödel number of a derivation of the formula with Gödel number y from the axioms of **Q**, is (primitive) recursive. Other primitive recursive functions we'll need are num (Proposition 10.6) and Subst (Proposition 10.11). From these, it is possible to define f by minimization; thus, f is recursive.

First, define

$$\begin{aligned} A(n_0, \ldots, n_k, m) = \\ \mathrm{Subst}(\mathrm{Subst}(\ldots \mathrm{Subst}(&{}^{\ulcorner}\varphi_f{}^{\urcorner}, \mathrm{num}(n_0), {}^{\ulcorner}x_0{}^{\urcorner}), \\ &\ldots), \mathrm{num}(n_k), {}^{\ulcorner}x_k{}^{\urcorner}), \mathrm{num}(m), {}^{\ulcorner}y{}^{\urcorner}) \end{aligned}$$

This looks complicated, but it's just the function $A(n_0, \ldots, n_k, m) = {}^{\ulcorner}\varphi_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{m})^{\urcorner}$.

Now, consider the relation $R(n_0, \ldots, n_k, s)$ which holds if $(s)_0$ is the Gödel number of a derivation from **Q** of $\varphi_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{(s)_1})$:

$$R(n_0, \ldots, n_k, s) \quad \text{iff} \quad \mathrm{Prf}_{\mathbf{Q}}((s)_0, A(n_0, \ldots, n_k, (s)_1))$$

If we can find an s such that $R(n_0, \ldots, n_k, s)$ hold, we have found a pair of numbers—$(s)_0$ and (s_1)—such that $(s)_0$ is the Gödel number of a derivation of $A_f(\overline{n_0}, \ldots, \overline{n_k}, (s)_1)$. So looking for s is like looking for the pair d and m in the informal proof. And a computable function that "looks for" such an s can be defined by regular minimization. Note that R is regular: for every n_0, ..., n_k, there is a derivation δ of $\mathbf{Q} \vdash \varphi_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{f(n_0, \ldots, n_k)})$, so $R(n_0, \ldots, n_k, s)$ holds for $s = \langle {}^{\ulcorner}\delta^{\urcorner}, f(n_0, \ldots, n_k)\rangle$. So, we can write f as

$$f(n_0, \ldots, n_k) = (\mu s\, R(n_0, \ldots, n_k, s))_1. \qquad \square$$

11.3 The Beta Function Lemma

In order to show that we can carry out primitive recursion if addition, multiplication, and $\chi_=$ are available, we need to develop functions that handle sequences. (If we had exponentiation as well, our task would be easier.) When we had primitive recursion, we could define things like the "n-th prime," and pick a fairly straightforward coding. But here we do not have primitive recursion—in fact we want to show that we can do primitive recursion using minimization—so we need to be more clever.

Lemma 11.5. *There is a function $\beta(d, i)$ such that for every sequence $a_0, \ldots, a_n$ there is a number d, such that for every $i \leq n$, $\beta(d, i) = a_i$. Moreover, β can be defined from the basic functions using just composition and regular minimization.*

Think of d as coding the sequence $\langle a_0, \ldots, a_n \rangle$, and $\beta(d, i)$ returning the i-th element. (Note that this "coding" does *not* use the power-of-primes coding we're already familiar with!). The lemma is fairly minimal; it doesn't say we can concatenate sequences or append elements, or even that we can *compute* d from $a_0, \ldots, a_n$ using functions definable by composition and regular minimization. All it says is that there is a "decoding" function such that every sequence is "coded."

The use of the notation β is Gödel's. To repeat, the hard part of proving the lemma is defining a suitable β using the seemingly restricted resources, i.e., using just composition and minimization—however, we're allowed to use addition, multiplication, and $\chi_=$. There are various ways to prove this lemma, but one of the cleanest is still Gödel's original method, which used a number-theoretic fact called Sunzi's Theorem (traditionally, the "Chinese Remainder Theorem").

Definition 11.6. Two natural numbers a and b are *relatively prime* iff their greatest common divisor is 1; in other words, they have no other divisors in common.

Definition 11.7. Natural numbers a and b are *congruent modulo* c, $a \equiv b \mod c$, iff $c \mid (a - b)$, i.e., a and b have the same remainder when divided by c.

Here is Sunzi's Theorem:

Theorem 11.8. *Suppose $x_0, \ldots, x_n$ are (pairwise) relatively prime. Let $y_0, \ldots, y_n$ be any numbers. Then there is a number z such that*

$$\begin{aligned} z &\equiv y_0 \mod x_0 \\ z &\equiv y_1 \mod x_1 \\ &\vdots \\ z &\equiv y_n \mod x_n. \end{aligned}$$

Here is how we will use Sunzi's Theorem: if $x_0, \ldots, x_n$ are bigger than $y_0, \ldots, y_n$ respectively, then we can take z to code the sequence $\langle y_0, \ldots, y_n \rangle$. To recover y_i, we need only divide z by x_i and take the remainder. To use this coding, we will need to find suitable values for $x_0, \ldots, x_n$.

A couple of observations will help us in this regard. Given $y_0, \ldots, y_n$, let

$$j = \max(n, y_0, \ldots, y_n) + 1,$$

and let

$$\begin{aligned} x_0 &= 1 + j! \\ x_1 &= 1 + 2 \cdot j! \\ x_2 &= 1 + 3 \cdot j! \\ &\vdots \\ x_n &= 1 + (n+1) \cdot j! \end{aligned}$$

Then two things are true:

1. $x_0, \ldots, x_n$ are relatively prime.
2. For each i, $y_i < x_i$.

To see that (1) is true, note that if p is a prime number and $p \mid x_i$ and $p \mid x_k$, then $p \mid 1 + (i+1)j!$ and $p \mid 1 + (k+1)j!$. But then p divides their difference,

$$(1 + (i+1)j!) - (1 + (k+1)j!) = (i-k)j!.$$

Since p divides $1 + (i+1)j!$, it can't divide $j!$ as well (otherwise, the first division would leave a remainder of 1). So p divides $i - k$, since p divides $(i-k)j!$. But $|i-k|$ is at most n, and we have chosen $j > n$, so this implies that $p \mid j!$, again a contradiction. So there is no prime number dividing both x_i and x_k. Clause (2) is easy: we have $y_i < j < j! < x_i$.

Now let us prove the β function lemma. Remember that we can use 0, successor, plus, times, $\chi_=$, projections, and any function defined from them using composition and minimization applied to regular functions. We can also use a relation if its characteristic function is so definable. As before we can show that these relations are closed under Boolean combinations and bounded quantification; for example:

$$\begin{aligned} \text{not}(x) &= \chi_=(x, 0) \\ (\min x \leq z)\, R(x, y) &= \mu x\, (R(x, y) \vee x = z) \\ (\exists x \leq z)\, R(x, y) &\Leftrightarrow R((\min x \leq z)\, R(x, y), y) \end{aligned}$$

We can then show that all of the following are also definable without primitive recursion:

1. The pairing function, $J(x, y) = \frac{1}{2}[(x + y)(x + y + 1)] + x$;
2. the projection functions

$$\begin{aligned} K(z) &= (\min x \leq z)\, (\exists y \leq z)\, z = J(x, y), \\ L(z) &= (\min y \leq z)\, (\exists x \leq z)\, z = J(x, y); \end{aligned}$$

3. the less-than relation $x < y$;
4. the divisibility relation $x \mid y$;
5. the function $\mathrm{rem}(x, y)$ which returns the remainder when y is divided by x.

Now define

$$\begin{aligned}\beta^*(d_0, d_1, i) &= \mathrm{rem}(1 + (i+1)d_1, d_0) \text{ and}\\ \beta(d, i) &= \beta^*(K(d), L(d), i).\end{aligned}$$

This is the function we want. Given $a_0, \ldots, a_n$ as above, let

$$j = \max(n, a_0, \ldots, a_n) + 1,$$

and let $d_1 = j!$. By (1) above, we know that $1 + d_1, 1 + 2d_1, \ldots, 1 + (n+1)d_1$ are relatively prime, and by (2) that all are greater than $a_0, \ldots, a_n$. By Sunzi's Theorem there is a value d_0 such that for each i,

$$d_0 \equiv a_i \mod (1 + (i+1)d_1)$$

and so (because d_1 is greater than a_i),

$$a_i = \mathrm{rem}(1 + (i+1)d_1, d_0).$$

Let $d = J(d_0, d_1)$. Then for each $i \leq n$, we have

$$\begin{aligned}\beta(d, i) &= \beta^*(d_0, d_1, i)\\ &= \mathrm{rem}(1 + (i+1)d_1, d_0)\\ &= a_i\end{aligned}$$

which is what we need. This completes the proof of the β-function lemma.

11.4 Simulating Primitive Recursion

Now we can show that definition by primitive recursion can be "simulated" by regular minimization using the beta function. Suppose we have $f(\vec{x})$ and $g(\vec{x}, y, z)$. Then the function $h(x, \vec{z})$ defined from f and g by primitive recursion is

$$\begin{aligned}h(\vec{x}, 0) &= f(\vec{x})\\ h(\vec{x}, y+1) &= g(\vec{x}, y, h(\vec{x}, y)).\end{aligned}$$

We need to show that h can be defined from f and g using just composition and regular minimization, using the basic functions and functions defined from them using composition and regular minimization (such as β).

Lemma 11.9. *If h can be defined from f and g using primitive recursion, it can be defined from f, g, the functions* zero, succ, P^n_i, add, mult, $\chi_=$, *using composition and regular minimization.*

Proof. First, define an auxiliary function $\hat{h}(\vec{x}, y)$ which returns the least number d such that d codes a sequence which satisfies

1. $(d)_0 = f(\vec{x})$, and
2. for each $i < y$, $(d)_{i+1} = g(\vec{x}, i, (d)_i)$,

where now $(d)_i$ is short for $\beta(d, i)$. In other words, $\hat{h}$ returns the sequence $\langle h(\vec{x}, 0), h(\vec{x}, 1), \ldots, h(\vec{x}, y)\rangle$. We can write $\hat{h}$ as

$$\hat{h}(\vec{x}, y) = \mu d\ (\beta(d, 0) = f(\vec{x}) \wedge (\forall i < y)\ \beta(d, i+1) = g(\vec{x}, i, \beta(d, i))).$$

Note: no primitive recursion is needed here, just minimization. The function we minimize is regular because of the beta function lemma Lemma 11.5.

But now we have

$$h(\vec{x}, y) = \beta(\hat{h}(\vec{x}, y), y),$$

so h can be defined from the basic functions using just composition and regular minimization. □

11.5 Basic Functions are Representable in Q

First we have to show that all the basic functions are representable in **Q**. In the end, we need to show how to assign to each k-ary basic function $f(x_0, \ldots, x_{k-1})$ a formula $\varphi_f(x_0, \ldots, x_{k-1}, y)$ that represents it.

We will be able to represent zero, successor, plus, times, the characteristic function for equality, and projections. In each case, the appropriate representing function is entirely straightforward; for example, zero is represented by the formula $y = 0$, successor is represented by the formula $x'_0 = y$, and addition is represented by the formula $(x_0 + x_1) = y$. The work involves showing that **Q** can prove the relevant sentences; for example, saying that addition is represented by the formula above involves showing that for every pair of natural numbers m and n, **Q** proves

$$\overline{n} + \overline{m} = \overline{n+m} \text{ and}$$
$$\forall y\, ((\overline{n} + \overline{m}) = y \to y = \overline{n+m}).$$

Proposition 11.10. *The zero function* $\mathrm{zero}(x) = 0$ *is represented in* **Q** *by* $\varphi_{\mathrm{zero}}(x, y) \equiv y = 0$.

Proposition 11.11. *The successor function* $\mathrm{succ}(x) = x + 1$ *is represented in* **Q** *by* $\varphi_{\mathrm{succ}}(x, y) \equiv y = x'$.

Proposition 11.12. *The projection function* $P^n_i(x_0, \ldots, x_{n-1}) = x_i$ *is represented in* **Q** *by*

$$\varphi_{P^n_i}(x_0, \ldots, x_{n-1}, y) \equiv y = x_i.$$

Proposition 11.13. *The characteristic function of* $=$*,*

$$\chi_=(x_0, x_1) = \begin{cases} 1 & \text{if } x_0 = x_1 \\ 0 & \text{otherwise} \end{cases}$$

is represented in **Q** *by*

$$\varphi_{\chi_=}(x_0, x_1, y) \equiv (x_0 = x_1 \wedge y = \overline{1}) \vee (x_0 \neq x_1 \wedge y = \overline{0}).$$

The proof requires the following lemma.

Lemma 11.14. *Given natural numbers* n *and* m*, if* $n \neq m$*, then* $\mathbf{Q} \vdash \overline{n} \neq \overline{m}$*.*

Proof. Use induction on n to show that for every m, if $n \neq m$, then $Q \vdash \overline{n} \neq \overline{m}$.

In the base case, $n = 0$. If m is not equal to 0, then $m = k + 1$ for some natural number k. We have an axiom that says $\forall x\, 0 \neq x'$. By a quantifier axiom, replacing x by $\overline{k}$, we can conclude $0 \neq \overline{k}'$. But $\overline{k}'$ is just $\overline{m}$.

In the induction step, we can assume the claim is true for n, and consider $n + 1$. Let m be any natural number. There are two possibilities: either $m = 0$ or for some k we have $m = k + 1$. The first case is handled as above. In the second case, suppose $n + 1 \neq k + 1$. Then $n \neq k$. By the induction hypothesis for n we have $\mathbf{Q} \vdash \overline{n} \neq \overline{k}$. We have an axiom that says $\forall x\, \forall y\, x' = y' \rightarrow x = y$. Using a quantifier axiom, we have $\overline{n}' = \overline{k}' \rightarrow \overline{n} = \overline{k}$. Using propositional logic, we can conclude, in **Q**, $\overline{n} \neq \overline{k} \rightarrow \overline{n}' \neq \overline{k}'$. Using modus ponens, we can conclude $\overline{n}' \neq \overline{k}'$, which is what we want, since $\overline{k}'$ is $\overline{m}$. □

Note that the lemma does not say much: in essence it says that **Q** can prove that different numerals denote different objects. For example, **Q** proves $0'' \neq 0'''$. But showing that this holds in general requires some care. Note also that although we are using induction, it is induction *outside* of **Q**.

Proof of Proposition 11.13. If $n = m$, then $\overline{n}$ and $\overline{m}$ are the same term, and $\chi_=(n, m) = 1$. But $\mathbf{Q} \vdash (\overline{n} = \overline{m} \wedge \overline{1} = \overline{1})$, so it proves $\varphi_=(\overline{n}, \overline{m}, \overline{1})$. If $n \neq m$, then $\chi_=(n, m) = 0$. By Lemma 11.14, $\mathbf{Q} \vdash \overline{n} \neq \overline{m}$ and so also $(\overline{n} \neq \overline{m} \wedge 0 = 0)$. Thus $\mathbf{Q} \vdash \varphi_=(\overline{n}, \overline{m}, \overline{0})$.

For the second part, we also have two cases. If $n = m$, we have to show that $\mathbf{Q} \vdash \forall y\, (\varphi_=(\overline{n}, \overline{m}, y) \rightarrow y = \overline{1})$. Arguing informally, suppose $\varphi_=(\overline{n}, \overline{m}, y)$, i.e.,

$$(\overline{n} = \overline{n} \wedge y = \overline{1}) \vee (\overline{n} \neq \overline{n} \wedge y = \overline{0})$$

The left disjunct implies $y = \overline{1}$ by logic; the right contradicts $\overline{n} = \overline{n}$ which is provable by logic.

Suppose, on the other hand, that $n \neq m$. Then $\varphi_=(\overline{n}, \overline{m}, y)$ is

$$(\overline{n} = \overline{m} \wedge y = \overline{1}) \vee (\overline{n} \neq \overline{m} \wedge y = \overline{0})$$

Here, the left disjunct contradicts $\overline{n} \neq \overline{m}$, which is provable in **Q** by Lemma 11.14; the right disjunct entails $y = \overline{0}$. □

Proposition 11.15. *The addition function* $\mathrm{add}(x_0, x_1) = x_0 + x_1$ *is represented in* **Q** *by*

$$\varphi_{\mathrm{add}}(x_0, x_1, y) \equiv y = (x_0 + x_1).$$

Lemma 11.16. $\mathbf{Q} \vdash (\overline{n} + \overline{m}) = \overline{n + m}$

Proof. We prove this by induction on m. If $m = 0$, the claim is that $\mathbf{Q} \vdash (\overline{n} + \mathsf{0}) = \overline{n}$. This follows by axiom Q_4. Now suppose the claim for m; let's prove the claim for $m + 1$, i.e., prove that $\mathbf{Q} \vdash (\overline{n} + \overline{m + 1}) = \overline{n + m + 1}$. Note that $\overline{m + 1}$ is just $\overline{m}'$, and $\overline{n + m + 1}$ is just $\overline{n + m}'$. By axiom Q_5, $\mathbf{Q} \vdash (\overline{n} + \overline{m}') = (\overline{n} + \overline{m})'$. By induction hypothesis, $\mathbf{Q} \vdash (\overline{n} + \overline{m}) = \overline{n + m}$. So $\mathbf{Q} \vdash (\overline{n} + \overline{m}') = \overline{n + m}'$. □

Proof of Proposition 11.15. The formula $\varphi_{\mathrm{add}}(x_0, x_1, y)$ representing add is $y = (x_0 + x_1)$. First we show that if $\mathrm{add}(n, m) = k$, then $\mathbf{Q} \vdash \varphi_{\mathrm{add}}(\overline{n}, \overline{m}, \overline{k})$, i.e., $\mathbf{Q} \vdash \overline{k} = (\overline{n} + \overline{m})$. But since $k = n + m$, $\overline{k}$ just is $\overline{n + m}$, and we've shown in Lemma 11.16 that $\mathbf{Q} \vdash (\overline{n} + \overline{m}) = \overline{n + m}$.

We also have to show that if $\mathrm{add}(n, m) = k$, then

$$\mathbf{Q} \vdash \forall y\, (\varphi_{\mathrm{add}}(\overline{n}, \overline{m}, y) \to y = \overline{k}).$$

Suppose we have $(\overline{n} + \overline{m}) = y$. Since

$$\mathbf{Q} \vdash (\overline{n} + \overline{m}) = \overline{n + m},$$

we can replace the left side with $\overline{n + m}$ and get $\overline{n + m} = y$, for arbitrary y. □

Proposition 11.17. *The multiplication function* $\mathrm{mult}(x_0, x_1) = x_0 \cdot x_1$ *is represented in* **Q** *by*

$$\varphi_{\mathrm{mult}}(x_0, x_1, y) \equiv y = (x_0 \times x_1).$$

Proof. Exercise. □

Lemma 11.18. $\mathbf{Q} \vdash (\overline{n} \times \overline{m}) = \overline{n \cdot m}$

Proof. Exercise. □

Recall that we use $\times$ for the function symbol of the language of arithmetic, and $\cdot$ for the ordinary multiplication operation on numbers. So $\cdot$ can appear between expressions for numbers (such as in $m \cdot n$) while $\times$ appears only between terms of the language of arithmetic (such as in $(\overline{m} \times \overline{n})$). Even more confusingly, $+$ is used for both the function symbol and the addition operation. When it appears between terms—e.g., in $(\overline{n} + \overline{m})$—it is the 2-place function symbol of the language of arithmetic, and when it appears between numbers—e.g., in $n + m$—it is the addition operation. This includes the case $\overline{n + m}$: this is the standard numeral corresponding to the number $n + m$.

11.6 Composition is Representable in Q

Suppose h is defined by

$$h(x_0, \ldots, x_{l-1}) = f(g_0(x_0, \ldots, x_{l-1}), \ldots, g_{k-1}(x_0, \ldots, x_{l-1})).$$

where we have already found formulas $\varphi_f, \varphi_{g_0}, \ldots, \varphi_{g_{k-1}}$ representing the functions f, and $g_0, \ldots, g_{k-1}$, respectively. We have to find a formula φ_h representing h.

Let's start with a simple case, where all functions are 1-place, i.e., consider $h(x) = f(g(x))$. If $\varphi_f(y, z)$ represents f, and $\varphi_g(x, y)$ represents g, we need a formula $\varphi_h(x, z)$ that represents h. Note that $h(x) = z$ iff there is a y such that both $z = f(y)$ and $y = g(x)$. (If $h(x) = z$, then $g(x)$ is such a y; if such a y exists, then since $y = g(x)$ and $z = f(y)$, $z = f(g(x))$.) This suggests that $\exists y\, (\varphi_g(x, y) \wedge \varphi_f(y, z))$ is a good candidate for $\varphi_h(x, z)$. We just have to verify that **Q** proves the relevant formulas.

Proposition 11.19. *If $h(n) = m$, then $\mathbf{Q} \vdash \varphi_h(\overline{n}, \overline{m})$.*

Proof. Suppose $h(n) = m$, i.e., $f(g(n)) = m$. Let $k = g(n)$. Then

$$\mathbf{Q} \vdash \varphi_g(\overline{n}, \overline{k})$$

since φ_g represents g, and

$$\mathbf{Q} \vdash \varphi_f(\overline{k}, \overline{m})$$

since φ_f represents f. Thus,

$$\mathbf{Q} \vdash \varphi_g(\overline{n}, \overline{k}) \wedge \varphi_f(\overline{k}, \overline{m})$$

and consequently also

$$\mathbf{Q} \vdash \exists y\, (\varphi_g(\overline{n}, y) \wedge \varphi_f(y, \overline{m})),$$

i.e., $\mathbf{Q} \vdash \varphi_h(\overline{n}, \overline{m})$. □

Proposition 11.20. *If $h(n) = m$, then $\mathbf{Q} \vdash \forall z\,(\varphi_h(\overline{n}, z) \to z = \overline{m})$.*

Proof. Suppose $h(n) = m$, i.e., $f(g(n)) = m$. Let $k = g(n)$. Then

$$\mathbf{Q} \vdash \forall y\,(\varphi_g(\overline{n}, y) \to y = \overline{k})$$

since φ_g represents g, and

$$\mathbf{Q} \vdash \forall z\,(\varphi_f(\overline{k}, z) \to z = \overline{m})$$

since φ_f represents f. Using just a little bit of logic, we can show that also

$$\mathbf{Q} \vdash \forall z\,(\exists y\,(\varphi_g(\overline{n}, y) \wedge \varphi_f(y, z)) \to z = \overline{m}).$$

i.e., $\mathbf{Q} \vdash \forall y\,(\varphi_h(\overline{n}, y) \to y = \overline{m})$. □

The same idea works in the more complex case where f and g_i have arity greater than 1.

Proposition 11.21. *If $\varphi_f(y_0, \ldots, y_{k-1}, z)$ represents $f(y_0, \ldots, y_{k-1})$ in $\mathbf{Q}$, and $\varphi_{g_i}(x_0, \ldots, x_{l-1}, y)$ represents $g_i(x_0, \ldots, x_{l-1})$ in $\mathbf{Q}$, then*

$$\exists y_0 \ldots \exists y_{k-1}\,(\varphi_{g_0}(x_0, \ldots, x_{l-1}, y_0) \wedge \cdots \wedge \\ !A_{g_{k-1}}(x_0, \ldots, x_{l-1}, y_{k-1}) \wedge \varphi_f(y_0, \ldots, y_{k-1}, z))$$

represents

$$h(x_0, \ldots, x_{l-1}) = f(g_0(x_0, \ldots, x_{l-1}), \ldots, g_{k-1}(x_0, \ldots, x_{l-1})).$$

Proof. Exercise. □

11.7 Regular Minimization is Representable in Q

Let's consider unbounded search. Suppose $g(x, z)$ is regular and representable in $\mathbf{Q}$, say by the formula $\varphi_g(x, z, y)$. Let f be defined by $f(z) = \mu x\,[g(x, z) = 0]$. We would like to find a formula $\varphi_f(z, y)$ representing f. The value of $f(z)$ is that number x which (a) satisfies $g(x, z) = 0$ and (b) is the least such, i.e., for any $w < x$, $g(w, z) \neq 0$. So the following is a natural choice:

$$\varphi_f(z, y) \equiv \varphi_g(y, z, 0) \wedge \forall w\,(w < y \to \neg\varphi_g(w, z, 0)).$$

In the general case, of course, we would have to replace z with $z_0, \ldots, z_k$.

The proof, again, will involve some lemmas about things **Q** is strong enough to prove.

Lemma 11.22. *For every constant symbol a and every natural number n,*

$$\mathbf{Q} \vdash (a' + \overline{n}) = (a + \overline{n})'.$$

Proof. The proof is, as usual, by induction on n. In the base case, $n = 0$, we need to show that **Q** proves $(a' + \mathsf{o}) = (a + \mathsf{o})'$. But we have:

$$\mathbf{Q} \vdash (a' + \mathsf{o}) = a' \quad \text{by axiom } Q_4 \tag{11.1}$$
$$\mathbf{Q} \vdash (a + \mathsf{o}) = a \quad \text{by axiom } Q_4 \tag{11.2}$$
$$\mathbf{Q} \vdash (a + \mathsf{o})' = a' \quad \text{by eq. (11.2)} \tag{11.3}$$
$$\mathbf{Q} \vdash (a' + \mathsf{o}) = (a + \mathsf{o})' \quad \text{by eq. (11.1) and eq. (11.3)}$$

In the induction step, we can assume that we have shown that $\mathbf{Q} \vdash (a' + \overline{n}) = (a + \overline{n})'$. Since $\overline{n+1}$ is $\overline{n}'$, we need to show that **Q** proves $(a' + \overline{n}') = (a + \overline{n}')'$. We have:

$$\mathbf{Q} \vdash (a' + \overline{n}') = (a' + \overline{n})' \quad \text{by axiom } Q_5 \tag{11.4}$$
$$\mathbf{Q} \vdash (a' + \overline{n}') = (a + \overline{n}')' \quad \text{inductive hypothesis} \tag{11.5}$$
$$\mathbf{Q} \vdash (a' + \overline{n})' = (a + \overline{n}')' \quad \text{by eq. (11.4) and eq. (11.5).}$$ □

It is again worth mentioning that this is weaker than saying that **Q** proves $\forall x \forall y\, (x' + y) = (x + y)'$. Although this sentence is true in $\mathfrak{N}$, **Q** does not prove it.

Lemma 11.23. $\mathbf{Q} \vdash \forall x \neg x < \mathsf{o}$.

Proof. We give the proof informally (i.e., only giving hints as to how to construct the formal derivation).

We have to prove $\neg a < \mathsf{o}$ for an arbitrary a. By the definition of $<$, we need to prove $\neg \exists y\, (y' + a) = \mathsf{o}$ in **Q**. We'll assume $\exists y\, (y' + a) = \mathsf{o}$ and prove a contradiction. Suppose $(b' + a) = \mathsf{o}$. Using Q_3, we have that $a = \mathsf{o} \vee \exists y\, a = y'$. We distinguish cases.

Case 1: $a = \mathsf{o}$ holds. From $(b' + a) = \mathsf{o}$, we have $(b' + \mathsf{o}) = \mathsf{o}$. By axiom Q_4 of **Q**, we have $(b' + \mathsf{o}) = b'$, and hence $b' = \mathsf{o}$. But by axiom Q_2 we also have $b' \neq \mathsf{o}$, a contradiction.

Case 2: For some c, $a = c'$. But then we have $(b' + c') = \mathsf{o}$. By axiom Q_5, we have $(b' + c)' = \mathsf{o}$, again contradicting axiom Q_2. □

Lemma 11.24. *For every natural number n,*

$$\mathbf{Q} \vdash \forall x\, (x < \overline{n+1} \to (x = \mathsf{o} \vee \cdots \vee x = \overline{n})).$$

Proof. We use induction on n. Let us consider the base case, when $n = 0$. In that case, we need to show $a < \overline{1} \to a = \mathsf{o}$, for arbitrary a. Suppose $a < \overline{1}$. Then by the defining axiom for $<$, we have $\exists y\, (y' + a) = \mathsf{o}'$ (since $\overline{1} \equiv \mathsf{o}'$).

Suppose b has that property, i.e., we have $(b' + a) = 0'$. We need to show $a = 0$. By axiom Q_3, we have either $a = 0$ or that there is a c such that $a = c'$. In the former case, there is nothing to show. So suppose $a = c'$. Then we have $(b' + c') = 0'$. By axiom Q_5 of **Q**, we have $(b' + c)' = 0'$. By axiom Q_1, we have $(b' + c) = 0$. But this means, by axiom Q_8, that $c < 0$, contradicting Lemma 11.23.

Now for the inductive step. We prove the case for $n + 1$, assuming the case for n. So suppose $a < \overline{n+2}$. Again using Q_3 we can distinguish two cases: $a = 0$ and for some b, $a = c'$. In the first case, $a = 0 \vee \cdots \vee a = \overline{n+1}$ follows trivially. In the second case, we have $c' < \overline{n+2}$, i.e., $c' < \overline{n+1}'$. By axiom Q_8, for some d, $(d' + c') = \overline{n+1}'$. By axiom Q_5, $(d' + c)' = \overline{n+1}'$. By axiom Q_1, $(d' + c) = \overline{n+1}$, and so $c < \overline{n+1}$ by axiom Q_8. By inductive hypothesis, $c = 0 \vee \cdots \vee c = \overline{n}$. From this, we get $c' = 0' \vee \cdots \vee c' = \overline{n}'$ by logic, and so $a = \overline{1} \vee \cdots \vee a = \overline{n+1}$ since $a = c'$. □

Lemma 11.25. *For every natural number m,*

$$\mathbf{Q} \vdash \forall y\,((y < \overline{m} \vee \overline{m} < y) \vee y = \overline{m}).$$

Proof. By induction on m. First, consider the case $m = 0$. $\mathbf{Q} \vdash \forall y\,(y = 0 \vee \exists z\, y = z')$ by Q_3. Let a be arbitrary. Then either $a = 0$ or for some b, $a = b'$. In the former case, we also have $(a < 0 \vee 0 < a) \vee a = 0$. But if $a = b'$, then $(b' + 0) = (a + 0)$ by the logic of $=$. By Q_4, $(a + 0) = a$, so we have $(b' + 0) = a$, and hence $\exists z\,(z' + 0) = a$. By the definition of $<$ in Q_8, $0 < a$. If $0 < a$, then also $(0 < a \vee a < 0) \vee a = 0$.

Now suppose we have

$$\mathbf{Q} \vdash \forall y\,((y < \overline{m} \vee \overline{m} < y) \vee y = \overline{m})$$

and we want to show

$$\mathbf{Q} \vdash \forall y\,((y < \overline{m+1} \vee \overline{m+1} < y) \vee y = \overline{m+1})$$

Let a be arbitrary. By Q_3, either $a = 0$ or for some b, $a = b'$. In the first case, we have $\overline{m}' + a = \overline{m+1}$ by Q_4, and so $a < \overline{m+1}$ by Q_8.

Now consider the second case, $a = b'$. By the induction hypothesis, $(b < \overline{m} \vee \overline{m} < b) \vee b = \overline{m}$.

The first disjunct $b < \overline{m}$ is equivalent (by Q_8) to $\exists z\,(z' + b) = \overline{m}$. Suppose c has this property. If $(c' + b) = \overline{m}$, then also $(c' + b)' = \overline{m}'$. By Q_5, $(c' + b)' = (c' + b')$. Hence, $(c' + b') = \overline{m}'$. We get $\exists u\,(u' + b') = \overline{m+1}$ by existentially generalizing on c' and keeping in mind that $\overline{m}' \equiv \overline{m+1}$. Hence, if $b < \overline{m}$ then $b' < \overline{m+1}$ and so $a < \overline{m+1}$.

Now suppose $\overline{m} < b$, i.e., $\exists z\,(z' + \overline{m}) = b$. Suppose c is such a z, i.e., $(c' + \overline{m}) = b$. By logic, $(c' + \overline{m})' = b'$. By Q_5, $(c' + \overline{m}') = b'$. Since $a = b'$ and $\overline{m}' \equiv \overline{m+1}$, $(c' + \overline{m+1}) = a$. By Q_8, $\overline{m+1} < a$.

Finally, assume $b = \overline{m}$. Then, by logic, $b' = \overline{m}'$, and so $a = \overline{m+1}$.

Hence, from each disjunct of the case for m and b, we can obtain the corresponding disjunct for for $m+1$ and a. □

Proposition 11.26. *If $\varphi_g(x,z,y)$ represents $g(x,z)$ in* **Q**, *then*

$$\varphi_f(z,y) \equiv \varphi_g(y,z,0) \wedge \forall w\,(w < y \to \neg\varphi_g(w,z,0))$$

represents $f(z) = \mu x\,[g(x,z) = 0]$.

Proof. First we show that if $f(n) = m$, then $\mathbf{Q} \vdash \varphi_f(\overline{n},\overline{m})$, i.e.,

$$\mathbf{Q} \vdash \varphi_g(\overline{m},\overline{n},0) \wedge \forall w\,(w < \overline{m} \to \neg\varphi_g(w,\overline{n},0)).$$

Since $\varphi_g(x,z,y)$ represents $g(x,z)$ and $g(m,n) = 0$ if $f(n) = m$, we have

$$\mathbf{Q} \vdash \varphi_g(\overline{m},\overline{n},0).$$

If $f(n) = m$, then for every $k < m$, $g(k,n) \neq 0$. So

$$\mathbf{Q} \vdash \neg\varphi_g(\overline{k},\overline{n},0).$$

We get that

$$\mathbf{Q} \vdash \forall w\,(w < \overline{m} \to \neg\varphi_g(w,\overline{n},0)). \tag{11.6}$$

by Lemma 11.23 in case $m = 0$ and by Lemma 11.24 otherwise.

Now let's show that if $f(n) = m$, then $\mathbf{Q} \vdash \forall y\,(\varphi_f(\overline{n},y) \to y = \overline{m})$. We again sketch the argument informally, leaving the formalization to the reader.

Suppose $\varphi_f(\overline{n},b)$. From this we get (a) $\varphi_g(b,\overline{n},0)$ and (b) $\forall w\,(w < b \to \neg\varphi_g(w,\overline{n},0))$. By Lemma 11.25, $(b < \overline{m} \vee \overline{m} < b) \vee b = \overline{m}$. We'll show that both $b < \overline{m}$ and $\overline{m} < b$ leads to a contradiction.

If $\overline{m} < b$, then $\neg\varphi_g(\overline{m},\overline{n},0)$ from (b). But $m = f(n)$, so $g(m,n) = 0$, and so $\mathbf{Q} \vdash \varphi_g(\overline{m},\overline{n},0)$ since φ_g represents g. So we have a contradiction.

Now suppose $b < \overline{m}$. Then since $\mathbf{Q} \vdash \forall w\,(w < \overline{m} \to \neg\varphi_g(w,\overline{n},0))$ by eq. (11.6), we get $\neg\varphi_g(b,\overline{n},0)$. This again contradicts (a). □

11.8 Computable Functions are Representable in Q

Theorem 11.27. *Every computable function is representable in* **Q**.

Proof. For definiteness, and using the Church-Turing Thesis, let's say that a function is computable iff it is general recursive. The general recursive functions are those which can be defined from the zero function zero, the successor

function succ, and the projection function P_i^n using composition, primitive recursion, and regular minimization. By Lemma 11.9, any function h that can be defined from f and g can also be defined using composition and regular minimization from f, g, and zero, succ, P_i^n, add, mult, $\chi_=$. Consequently, a function is general recursive iff it can be defined from zero, succ, P_i^n, add, mult, $\chi_=$ using composition and regular minimization.

We've furthermore shown that the basic functions in question are representable in **Q** (Propositions 11.10 to 11.13, 11.15 and 11.17), and that any function defined from representable functions by composition or regular minimization (Proposition 11.21, Proposition 11.26) is also representable. Thus every general recursive function is representable in **Q**. □

We have shown that the set of computable functions can be characterized as the set of functions representable in **Q**. In fact, the proof is more general. From the definition of representability, it is not hard to see that any theory extending **Q** (or in which one can interpret **Q**) can represent the computable functions. But, conversely, in any derivation system in which the notion of derivation is computable, every representable function is computable. So, for example, the set of computable functions can be characterized as the set of functions representable in Peano arithmetic, or even Zermelo-Fraenkel set theory. As Gödel noted, this is somewhat surprising. We will see that when it comes to provability, questions are very sensitive to which theory you consider; roughly, the stronger the axioms, the more you can prove. But across a wide range of axiomatic theories, the representable functions are exactly the computable ones; stronger theories do not represent more functions as long as they are axiomatizable.

11.9 Representing Relations

Let us say what it means for a *relation* to be representable.

Definition 11.28. A relation $R(x_0, \ldots, x_k)$ on the natural numbers is *representable in* **Q** if there is a formula $\varphi_R(x_0, \ldots, x_k)$ such that whenever $R(n_0, \ldots, n_k)$ is true, **Q** proves $\varphi_R(\overline{n_0}, \ldots, \overline{n_k})$, and whenever $R(n_0, \ldots, n_k)$ is false, **Q** proves $\neg\varphi_R(\overline{n_0}, \ldots, \overline{n_k})$.

Theorem 11.29. *A relation is representable in* **Q** *if and only if it is computable.*

Proof. For the forwards direction, suppose $R(x_0, \ldots, x_k)$ is represented by the formula $\varphi_R(x_0, \ldots, x_k)$. Here is an algorithm for computing R: on input n_0, $\ldots$, n_k, simultaneously search for a proof of $\varphi_R(\overline{n_0}, \ldots, \overline{n_k})$ and a proof of $\neg\varphi_R(\overline{n_0}, \ldots, \overline{n_k})$. By our hypothesis, the search is bound to find one or the other; if it is the first, report "yes," and otherwise, report "no."

In the other direction, suppose $R(x_0, \ldots, x_k)$ is computable. By definition, this means that the function $\chi_R(x_0, \ldots, x_k)$ is computable. By Theorem 11.2,

χ_R is represented by a formula, say $\varphi_{\chi_R}(x_0, \ldots, x_k, y)$. Let $\varphi_R(x_0, \ldots, x_k)$ be the formula $\varphi_{\chi_R}(x_0, \ldots, x_k, \overline{1})$. Then for any $n_0, \ldots, n_k$, if $R(n_0, \ldots, n_k)$ is true, then $\chi_R(n_0, \ldots, n_k) = 1$, in which case $\mathbf{Q}$ proves $\varphi_{\chi_R}(\overline{n_0}, \ldots, \overline{n_k}, \overline{1})$, and so $\mathbf{Q}$ proves $\varphi_R(\overline{n_0}, \ldots, \overline{n_k})$. On the other hand, if $R(n_0, \ldots, n_k)$ is false, then $\chi_R(n_0, \ldots, n_k) = 0$. This means that $\mathbf{Q}$ proves

$$\forall y\, (\varphi_{\chi_R}(\overline{n_0}, \ldots, \overline{n_k}, y) \to y = \overline{0}).$$

Since $\mathbf{Q}$ proves $\overline{0} \neq \overline{1}$, $\mathbf{Q}$ proves $\neg\varphi_{\chi_R}(\overline{n_0}, \ldots, \overline{n_k}, \overline{1})$, and so it proves $\neg\varphi_R(\overline{n_0}, \ldots, \overline{n_k})$.□

11.10 Undecidability

We call a theory $\mathbf{T}$ *undecidable* if there is no computational procedure which, after finitely many steps and unfailingly, provides a correct answer to the question "does $\mathbf{T}$ prove φ?" for any sentence φ in the language of $\mathbf{T}$. So $\mathbf{Q}$ would be decidable iff there were a computational procedure which decides, given a sentence φ in the language of arithmetic, whether $\mathbf{Q} \vdash \varphi$ or not. We can make this more precise by asking: Is the relation $\mathrm{Prov}_{\mathbf{Q}}(y)$, which holds of y iff y is the Gödel number of a sentence provable in $\mathbf{Q}$, recursive? The answer is: no.

Theorem 11.30. $\mathbf{Q}$ *is undecidable, i.e., the relation*

$$\mathrm{Prov}_{\mathbf{Q}}(y) \Leftrightarrow \mathrm{Sent}(y) \wedge \exists x\, \mathrm{Prf}_{\mathbf{Q}}(x, y)$$

is not recursive.

Proof. Suppose it were. Then we could solve the halting problem as follows: Given e and n, we know that $\varphi_e(n) \downarrow$ iff there is an s such that $T(e, n, s)$, where T is Kleene's predicate from Theorem 9.28. Since T is primitive recursive it is representable in $\mathbf{Q}$ by a formula ψ_T, that is, $\mathbf{Q} \vdash \psi_T(\overline{e}, \overline{n}, \overline{s})$ iff $T(e, n, s)$. If $\mathbf{Q} \vdash \psi_T(\overline{e}, \overline{n}, \overline{s})$ then also $\mathbf{Q} \vdash \exists y\, \psi_T(\overline{e}, \overline{n}, y)$. If no such s exists, then $\mathbf{Q} \vdash \neg\psi_T(\overline{e}, \overline{n}, \overline{s})$ for every s. But $\mathbf{Q}$ is ω-consistent, i.e., if $\mathbf{Q} \vdash \neg\varphi(\overline{n})$ for every $n \in \mathbb{N}$, then $\mathbf{Q} \nvdash \exists y\, \varphi(y)$. We know this because the axioms of $\mathbf{Q}$ are true in the standard model $\mathfrak{N}$. So, $\mathbf{Q} \nvdash \exists y\, \psi_T(\overline{e}, \overline{n}, y)$. In other words, $\mathbf{Q} \vdash \exists y\, \psi_T(\overline{e}, \overline{n}, y)$ iff there is an s such that $T(e, n, s)$, i.e., iff $\varphi_e(n) \downarrow$. From e and n we can compute ${}^{\#}\exists y\, \psi_T(\overline{e}, \overline{n}, y)^{\#}$, let $g(e, n)$ be the primitive recursive function which does that. So

$$h(e, n) = \begin{cases} 1 & \text{if } \mathrm{Prov}_{\mathbf{Q}}(g(e, n)) \\ 0 & \text{otherwise.} \end{cases}$$

This would show that h is recursive if $\mathrm{Prov}_{\mathbf{Q}}$ is. But h is not recursive, by Theorem 9.29, so $\mathrm{Prov}_{\mathbf{Q}}$ cannot be either. □

Corollary 11.31. *First-order logic is undecidable.*

Proof. If first-order logic were decidable, provability in $\mathbf{Q}$ would be as well, since $\mathbf{Q} \vdash \varphi$ iff $\vdash \omega \to \varphi$, where ω is the conjunction of the axioms of $\mathbf{Q}$. □

Chapter 12

Incompleteness and Provability

12.1 Introduction

Hilbert thought that a system of axioms for a mathematical structure, such as the natural numbers, is inadequate unless it allows one to derive all true statements about the structure. Combined with his later interest in formal systems of deduction, this suggests that he thought that we should guarantee that, say, the formal systems we are using to reason about the natural numbers is not only consistent, but also *complete*, i.e., every statement in its language is either derivable or its negation is. Gödel's first incompleteness theorem shows that no such system of axioms exists: there is no complete, consistent, axiomatizable formal system for arithmetic. In fact, no "sufficiently strong," consistent, axiomatizable mathematical theory is complete.

A more important goal of Hilbert's, the centerpiece of his program for the justification of modern ("classical") mathematics, was to find finitary consistency proofs for formal systems representing classical reasoning. With regard to Hilbert's program, then, Gödel's second incompleteness theorem was a much bigger blow. The second incompleteness theorem can be stated in vague terms, like the first incompleteness theorem. Roughly speaking, it says that no sufficiently strong theory of arithmetic can prove its own consistency. We will have to take "sufficiently strong" to include a little bit more than **Q**.

The idea behind Gödel's original proof of the incompleteness theorem can be found in the Epimenides paradox. Epimenides, a Cretan, asserted that all Cretans are liars; a more direct form of the paradox is the assertion "this sentence is false." Essentially, by replacing truth with derivability, Gödel was able to formalize a sentence which, in a roundabout way, asserts that it itself is not derivable. If that sentence were derivable, the theory would then be inconsistent. Gödel showed that the negation of that sentence is also not derivable from the system of axioms he was considering. (For this second part, Gödel had to assume that the theory **T** is what's called "ω-consistent."

ω-Consistency is related to consistency, but is a stronger property.[1] A few years after Gödel, Rosser showed that assuming simple consistency of **T** is enough.)

The first challenge is to understand how one can construct a sentence that refers to itself. For every formula φ in the language of **Q**, let $\ulcorner\varphi\urcorner$ denote the numeral corresponding to ${}^{\#}\varphi^{\#}$. Think about what this means: φ is a formula in the language of **Q**, ${}^{\#}\varphi^{\#}$ is a natural number, and $\ulcorner\varphi\urcorner$ is a *term* in the language of **Q**. So every formula φ in the language of **Q** has a *name*, $\ulcorner\varphi\urcorner$, which is a term in the language of **Q**; this provides us with a conceptual framework in which formulas in the language of **Q** can "say" things about other formulas. The following lemma is known as the fixed-point lemma.

Lemma 12.1. *Let* **T** *be any theory extending* **Q**, *and let* $\psi(x)$ *be any formula with only the variable* x *free. Then there is a sentence* φ *such that* $\mathbf{T} \vdash \varphi \leftrightarrow \psi(\ulcorner\varphi\urcorner)$.

The lemma asserts that given any property $\psi(x)$, there is a sentence φ that asserts "$\psi(x)$ is true of me," and **T** "knows" this.

How can we construct such a sentence? Consider the following version of the Epimenides paradox, due to Quine:

> "Yields falsehood when preceded by its quotation" yields falsehood when preceded by its quotation.

This sentence is not directly self-referential. It simply makes an assertion about the syntactic objects between quotes, and, in doing so, it is on par with sentences like

1. "Robert" is a nice name.
2. "I ran." is a short sentence.
3. "Has three words" has three words.

But what happens when one takes the phrase "yields falsehood when preceded by its quotation," and precedes it with a quoted version of itself? Then one has the original sentence! In short, the sentence asserts that it is false.

12.2 The Fixed-Point Lemma

The fixed-point lemma says that for any formula $\psi(x)$, there is a sentence φ such that $\mathbf{T} \vdash \varphi \leftrightarrow \psi(\ulcorner\varphi\urcorner)$, provided **T** extends **Q**. In the case of the liar sentence, we'd want φ to be equivalent (provably in **T**) to "$\ulcorner\varphi\urcorner$ is false," i.e., the statement that ${}^{\#}\varphi^{\#}$ is the Gödel number of a false sentence. To understand the idea of the proof, it will be useful to compare it with Quine's informal gloss

[1] That is, any ω-consistent theory is consistent, but not vice versa.

of φ as, "'yields a falsehood when preceded by its own quotation' yields a falsehood when preceded by its own quotation." The operation of taking an expression, and then forming a sentence by preceding this expression by its own quotation may be called *diagonalizing* the expression, and the result its diagonalization. So, the diagonalization of 'yields a falsehood when preceded by its own quotation' is "'yields a falsehood when preceded by its own quotation' yields a falsehood when preceded by its own quotation." Now note that Quine's liar sentence is not the diagonalization of 'yields a falsehood' but of 'yields a falsehood when preceded by its own quotation.' So the property being diagonalized to yield the liar sentence itself involves diagonalization!

In the language of arithmetic, we form quotations of a formula with one free variable by computing its Gödel numbers and then substituting the standard numeral for that Gödel number into the free variable. The diagonalization of $\alpha(x)$ is $\alpha(\overline{n})$, where $n = {}^{\ulcorner}\alpha(x)^{\urcorner}$. (From now on, let's abbreviate $\overline{{}^{\ulcorner}\alpha(x)^{\urcorner}}$ as $\ulcorner\alpha(x)\urcorner$.) So if $\psi(x)$ is "is a falsehood," then "yields a falsehood if preceded by its own quotation," would be "yields a falsehood when applied to the Gödel number of its diagonalization." If we had a symbol $diag$ for the function $\mathrm{diag}(n)$ which computes the Gödel number of the diagonalization of the formula with Gödel number n, we could write $\alpha(x)$ as $\psi(diag(x))$. And Quine's version of the liar sentence would then be the diagonalization of it, i.e., $\alpha(\ulcorner\alpha(x)\urcorner)$ or $\psi(diag(\ulcorner\psi(diag(x))\urcorner))$. Of course, $\psi(x)$ could now be any other property, and the same construction would work. For the incompleteness theorem, we'll take $\psi(x)$ to be "x is not derivable in **T**." Then $\alpha(x)$ would be "yields a sentence not derivable in **T** when applied to the Gödel number of its diagonalization."

To formalize this in **T**, we have to find a way to formalize diag. The function $\mathrm{diag}(n)$ is computable, in fact, it is primitive recursive: if n is the Gödel number of a formula $\alpha(x)$, $\mathrm{diag}(n)$ returns the Gödel number of $\alpha(\ulcorner\alpha(x)\urcorner)$. (Recall, $\ulcorner\alpha(x)\urcorner$ is the standard numeral of the Gödel number of $\alpha(x)$, i.e., $\overline{{}^{\ulcorner}\alpha(x)^{\urcorner}}$). If $diag$ were a function symbol in **T** representing the function diag, we could take φ to be the formula $\psi(diag(\ulcorner\psi(diag(x))\urcorner))$. Notice that

$$\begin{aligned}\mathrm{diag}({}^{\ulcorner}\psi(diag(x))^{\urcorner}) &= {}^{\ulcorner}\psi(diag(\ulcorner\psi(diag(x))\urcorner))^{\urcorner}\\ &= {}^{\ulcorner}\varphi^{\urcorner}.\end{aligned}$$

Assuming **T** can derive

$$diag(\ulcorner\psi(diag(x))\urcorner) = \ulcorner\varphi\urcorner,$$

it can derive $\psi(diag(\ulcorner\psi(diag(x))\urcorner)) \leftrightarrow \psi(\ulcorner\varphi\urcorner)$. But the left hand side is, by definition, φ.

Of course, $diag$ will in general not be a function symbol of **T**, and certainly is not one of **Q**. But, since diag is computable, it is *representable* in **Q** by some formula $\theta_{\mathrm{diag}}(x, y)$. So instead of writing $\psi(diag(x))$ we can write

$\exists y\,(\theta_{\text{diag}}(x,y) \wedge \psi(y))$. Otherwise, the proof sketched above goes through, and in fact, it goes through already in **Q**.

Lemma 12.2. *Let $\psi(x)$ be any formula with one free variable x. Then there is a sentence φ such that $\mathbf{Q} \vdash \varphi \leftrightarrow \psi(\ulcorner \varphi \urcorner)$.*

Proof. Given $\psi(x)$, let $\alpha(x)$ be the formula $\exists y\,(\theta_{\text{diag}}(x,y) \wedge \psi(y))$ and let φ be its diagonalization, i.e., the formula $\alpha(\ulcorner \alpha(x) \urcorner)$.

Since θ_{diag} represents diag, and $\text{diag}(\ulcorner \alpha(x) \urcorner) = \ulcorner \varphi \urcorner$, **Q** can derive

$$!D_{\text{diag}}(\ulcorner \alpha(x) \urcorner, \ulcorner \varphi \urcorner) \tag{12.1}$$

$$\forall y\,(\theta_{\text{diag}}(\ulcorner \alpha(x) \urcorner, y) \to y = \ulcorner \varphi \urcorner). \tag{12.2}$$

Now we show that $\mathbf{Q} \vdash \varphi \leftrightarrow \psi(\ulcorner \varphi \urcorner)$. We argue informally, using just logic and facts derivable in **Q**.

First, suppose φ, i.e., $\alpha(\ulcorner \alpha(x) \urcorner)$. Going back to the definition of $\alpha(x)$, we see that $\alpha(\ulcorner \alpha(x) \urcorner)$ just is

$$\exists y\,(\theta_{\text{diag}}(\ulcorner \alpha(x) \urcorner, y) \wedge \psi(y)).$$

Consider such a y. Since $\theta_{\text{diag}}(\ulcorner \alpha(x) \urcorner, y)$, by eq. (12.2), $y = \ulcorner \varphi \urcorner$. So, from $\psi(y)$ we have $\psi(\ulcorner \varphi \urcorner)$.

Now suppose $\psi(\ulcorner \varphi \urcorner)$. By eq. (12.1), we have

$$!D_{\text{diag}}(\ulcorner \alpha(x) \urcorner, \ulcorner \varphi \urcorner) \wedge \psi(\ulcorner \varphi \urcorner).$$

It follows that

$$\exists y\,(\theta_{\text{diag}}(\ulcorner \alpha(x) \urcorner, y) \wedge \psi(y)).$$

But that's just $\alpha(\ulcorner \alpha(x) \urcorner)$, i.e., φ. □

You should compare this to the proof of the fixed-point lemma in computability theory. The difference is that here we want to define a *statement* in terms of itself, whereas there we wanted to define a *function* in terms of itself; this difference aside, it is really the same idea.

12.3 The First Incompleteness Theorem

We can now describe Gödel's original proof of the first incompleteness theorem. Let **T** be any computably axiomatized theory in a language extending the language of arithmetic, such that **T** includes the axioms of **Q**. This means that, in particular, **T** represents computable functions and relations.

We have argued that, given a reasonable coding of formulas and proofs as numbers, the relation $\text{Prf}_T(x,y)$ is computable, where $\text{Prf}_T(x,y)$ holds if

and only if x is the Gödel number of a derivation of the formula with Gödel number y in $\mathbf{T}$. In fact, for the particular theory that Gödel had in mind, Gödel was able to show that this relation is primitive recursive, using the list of 45 functions and relations in his paper. The 45th relation, xBy, is just $\mathsf{Prf}_T(x, y)$ for his particular choice of $\mathbf{T}$. Remember that where Gödel uses the word "recursive" in his paper, we would now use the phrase "primitive recursive."

Since $\mathsf{Prf}_T(x, y)$ is computable, it is representable in $\mathbf{T}$. We will use $\mathsf{Prf}_T(x, y)$ to refer to the formula that represents it. Let $\mathsf{Prov}_T(y)$ be the formula $\exists x\, \mathsf{Prf}_T(x, y)$. This describes the 46th relation, $\mathrm{Bew}(y)$, on Gödel's list. As Gödel notes, this is the only relation that "cannot be asserted to be recursive." What he probably meant is this: from the definition, it is not clear that it is computable; and later developments, in fact, show that it isn't.

Let $\mathbf{T}$ be an axiomatizable theory containing $\mathbf{Q}$. Then $\mathrm{Prf}_T(x, y)$ is decidable, hence representable in $\mathbf{Q}$ by a formula $\mathsf{Prf}_T(x, y)$. Let $\mathsf{Prov}_T(y)$ be the formula we described above. By the fixed-point lemma, there is a formula $\gamma_{\mathbf{T}}$ such that $\mathbf{Q}$ (and hence $\mathbf{T}$) derives

$$\gamma_{\mathbf{T}} \leftrightarrow \neg \mathsf{Prov}_T(\ulcorner \gamma_{\mathbf{T}} \urcorner). \tag{12.3}$$

Note that $\gamma_{\mathbf{T}}$ says, in essence, "$\gamma_{\mathbf{T}}$ is not derivable in $\mathbf{T}$."

Lemma 12.3. *If $\mathbf{T}$ is a consistent, axiomatizable theory extending $\mathbf{Q}$, then $\mathbf{T} \nvdash \gamma_{\mathbf{T}}$.*

Proof. Suppose $\mathbf{T}$ derives $\gamma_{\mathbf{T}}$. Then there *is* a derivation, and so, for some number m, the relation $\mathrm{Prf}_T(m, {}^{\#}\gamma_{\mathbf{T}}{}^{\#})$ holds. But then $\mathbf{Q}$ derives the sentence $\mathsf{Prf}_T(\overline{m}, \ulcorner \gamma_{\mathbf{T}} \urcorner)$. So $\mathbf{Q}$ derives $\exists x\, \mathsf{Prf}_T(x, \ulcorner \gamma_{\mathbf{T}} \urcorner)$, which is, by definition, $\mathsf{Prov}_T(\ulcorner \gamma_{\mathbf{T}} \urcorner)$. By eq. (12.3), $\mathbf{Q}$ derives $\neg \gamma_{\mathbf{T}}$, and since $\mathbf{T}$ extends $\mathbf{Q}$, so does $\mathbf{T}$. We have shown that if $\mathbf{T}$ derives $\gamma_{\mathbf{T}}$, then it also derives $\neg \gamma_{\mathbf{T}}$, and hence it would be inconsistent. □

Definition 12.4. A theory $\mathbf{T}$ is *ω-consistent* if the following holds: if $\exists x\, \varphi(x)$ is any sentence and $\mathbf{T}$ derives $\neg \varphi(\overline{0})$, $\neg \varphi(\overline{1})$, $\neg \varphi(\overline{2})$, ... then $\mathbf{T}$ does not prove $\exists x\, \varphi(x)$.

Note that every ω-consistent theory is also consistent. This follows simply from the fact that if $\mathbf{T}$ is inconsistent, then $\mathbf{T} \vdash \varphi$ for every φ. In particular, if $\mathbf{T}$ is inconsistent, it derives both $\neg \varphi(\overline{n})$ for every n and also derives $\exists x\, \varphi(x)$. So, if $\mathbf{T}$ is inconsistent, it is ω-inconsistent. By contraposition, if $\mathbf{T}$ is ω-consistent, it must be consistent.

Lemma 12.5. *If $\mathbf{T}$ is an ω-consistent, axiomatizable theory extending $\mathbf{Q}$, then $\mathbf{T} \nvdash \neg \gamma_{\mathbf{T}}$.*

Proof. We show that if $\mathbf{T}$ derives $\neg \gamma_{\mathbf{T}}$, then it is ω-inconsistent. Suppose $\mathbf{T}$ derives $\neg \gamma_{\mathbf{T}}$. If $\mathbf{T}$ is inconsistent, it is ω-inconsistent, and we are done. Otherwise, $\mathbf{T}$ is consistent, so it does not derive $\gamma_{\mathbf{T}}$ by Lemma 12.3. Since there is

no derivation of $\gamma_{\mathbf{T}}$ in $\mathbf{T}$, $\mathbf{Q}$ derives

$$\neg \mathsf{Prf}_T(\overline{0}, \ulcorner \gamma_{\mathbf{T}} \urcorner), \neg \mathsf{Prf}_T(\overline{1}, \ulcorner \gamma_{\mathbf{T}} \urcorner), \neg \mathsf{Prf}_T(\overline{2}, \ulcorner \gamma_{\mathbf{T}} \urcorner), \ldots$$

and so does $\mathbf{T}$. On the other hand, by eq. (12.3), $\neg\gamma_{\mathbf{T}}$ is equivalent to $\exists x\, \mathsf{Prf}_T(x, \ulcorner \gamma_{\mathbf{T}} \urcorner)$. So $\mathbf{T}$ is ω-inconsistent. □

Theorem 12.6. *Let* $\mathbf{T}$ *be any* ω*-consistent, axiomatizable theory extending* $\mathbf{Q}$*. Then* $\mathbf{T}$ *is not complete.*

Proof. If $\mathbf{T}$ is ω-consistent, it is consistent, so $\mathbf{T} \nvdash \gamma_{\mathbf{T}}$ by Lemma 12.3. By Lemma 12.5, $\mathbf{T} \nvdash \neg\gamma_{\mathbf{T}}$. This means that $\mathbf{T}$ is incomplete, since it derives neither $\gamma_{\mathbf{T}}$ nor $\neg\gamma_{\mathbf{T}}$. □

12.4 Rosser's Theorem

Can we modify Gödel's proof to get a stronger result, replacing "ω-consistent" with simply "consistent"? The answer is "yes," using a trick discovered by Rosser. Rosser's trick is to use a "modified" derivability predicate $\mathsf{RProv}_T(y)$ instead of $\mathsf{Prov}_T(y)$.

Theorem 12.7. *Let* $\mathbf{T}$ *be any consistent, axiomatizable theory extending* $\mathbf{Q}$*. Then* $\mathbf{T}$ *is not complete.*

Proof. Recall that $\mathsf{Prov}_T(y)$ is defined as $\exists x\, \mathsf{Prf}_T(x, y)$, where $\mathsf{Prf}_T(x, y)$ represents the decidable relation which holds iff x is the Gödel number of a derivation of the sentence with Gödel number y. The relation that holds between x and y if x is the Gödel number of a *refutation* of the sentence with Gödel number y is also decidable. Let $\mathrm{not}(x)$ be the primitive recursive function which does the following: if x is the code of a formula φ, $\mathrm{not}(x)$ is a code of $\neg\varphi$. Then $\mathsf{Ref}_T(x, y)$ holds iff $\mathsf{Prf}_T(x, \mathrm{not}(y))$. Let $\mathsf{Ref}_T(x, y)$ represent it. Then, if $\mathbf{T} \vdash \neg\varphi$ and δ is a corresponding derivation, $\mathbf{Q} \vdash \mathsf{Ref}_T(\ulcorner \delta \urcorner, \ulcorner \varphi \urcorner)$. We define $\mathsf{RProv}_T(y)$ as

$$\exists x\, (\mathsf{Prf}_T(x, y) \wedge \forall z\, (z < x \to \neg \mathsf{Ref}_T(z, y))).$$

Roughly, $\mathsf{RProv}_T(y)$ says "there is a proof of y in $\mathbf{T}$, and there is no shorter refutation of y." Assuming $\mathbf{T}$ is consistent, $\mathsf{RProv}_T(y)$ is true of the same numbers as $\mathsf{Prov}_T(y)$; but from the point of view of *provability* in $\mathbf{T}$ (and we now know that there is a difference between truth and provability!) the two have different properties. If $\mathbf{T}$ is *in*consistent, then the two do *not* hold of the same numbers! ($\mathsf{RProv}_T(y)$ is often read as "y is Rosser provable." Since, as just discussed, Rosser provability is not some special kind of provability—in inconsistent theories, there are sentences that are provable but not Rosser provable—this may be confusing. To avoid the confusion, you could instead read it as "y is shmovable.")

By the fixed-point lemma, there is a formula $\rho_{\mathbf{T}}$ such that

$$\mathbf{Q} \vdash \rho_{\mathbf{T}} \leftrightarrow \neg \mathsf{RProv}_T(\ulcorner \rho_{\mathbf{T}} \urcorner). \tag{12.4}$$

In contrast to the proof of Theorem 12.6, here we claim that if $\mathbf{T}$ is consistent, $\mathbf{T}$ doesn't derive $\rho_{\mathbf{T}}$, and $\mathbf{T}$ also doesn't derive $\neg\rho_{\mathbf{T}}$. (In other words, we don't need the assumption of ω-consistency.)

First, let's show that $\mathbf{T} \nvdash \rho_T$. Suppose it did, so there is a derivation of ρ_T from T; let n be its Gödel number. Then $\mathbf{Q} \vdash \mathsf{Prf}_T(\overline{n}, \ulcorner \rho_T \urcorner)$, since Prf_T represents Prf_T in $\mathbf{Q}$. Also, for each $k < n$, k is not the Gödel number of a derivation of $\neg\rho_T$, since $\mathbf{T}$ is consistent. So for each $k < n$, $\mathbf{Q} \vdash \neg\mathsf{Ref}_T(\overline{k}, \ulcorner \rho_T \urcorner)$. By Lemma 11.24, $\mathbf{Q} \vdash \forall z\,(z < \overline{n} \to \neg\mathsf{Ref}_T(z, \ulcorner \rho_T \urcorner))$. Thus,

$$\mathbf{Q} \vdash \exists x\,(\mathsf{Prf}_T(x, \ulcorner \rho_T \urcorner) \wedge \forall z\,(z < x \to \neg\mathsf{Ref}_T(z, \ulcorner \rho_T \urcorner))),$$

but that's just $\mathsf{RProv}_T(\ulcorner \rho_T \urcorner)$. By eq. (12.4), $\mathbf{Q} \vdash \neg\rho_T$. Since $\mathbf{T}$ extends $\mathbf{Q}$, also $\mathbf{T} \vdash \neg\rho_T$. We've assumed that $\mathbf{T} \vdash \rho_T$, so $\mathbf{T}$ would be inconsistent, contrary to the assumption of the theorem.

Now, let's show that $\mathbf{T} \nvdash \neg\rho_T$. Again, suppose it did, and suppose n is the Gödel number of a derivation of $\neg\rho_T$. Then $\mathrm{Ref}_T(n, {}^{\#}\rho_T{}^{\#})$ holds, and since Ref_T represents Ref_T in $\mathbf{Q}$, $\mathbf{Q} \vdash \mathsf{Ref}_T(\overline{n}, \ulcorner \rho_T \urcorner)$. We'll again show that $\mathbf{T}$ would then be inconsistent because it would also derive ρ_T. Since

$$\mathbf{Q} \vdash \rho_T \leftrightarrow \neg\mathsf{RProv}_T(\ulcorner \rho_T \urcorner),$$

and since $\mathbf{T}$ extends $\mathbf{Q}$, it suffices to show that

$$\mathbf{Q} \vdash \neg\mathsf{RProv}_T(\ulcorner \rho_T \urcorner).$$

The sentence $\neg\mathsf{RProv}_T(\ulcorner \rho_T \urcorner)$, i.e.,

$$\neg\exists x\,(\mathsf{Prf}_T(x, \ulcorner \rho_T \urcorner) \wedge \forall z\,(z < x \to \neg\mathsf{Ref}_T(z, \ulcorner \rho_T \urcorner))),$$

is logically equivalent to

$$\forall x\,(\mathsf{Prf}_T(x, \ulcorner \rho_T \urcorner) \to \exists z\,(z < x \wedge \mathsf{Ref}_T(z, \ulcorner \rho_T \urcorner))).$$

We argue informally using logic, making use of facts about what $\mathbf{Q}$ derives. Suppose x is arbitrary and $\mathrm{Prf}_T(x, \ulcorner \rho_T \urcorner)$. We already know that $\mathbf{T} \nvdash \rho_T$, and so for every k, $\mathbf{Q} \vdash \neg\mathsf{Prf}_T(\overline{k}, \ulcorner \rho_T \urcorner)$. Thus, for every k it follows that $x \neq \overline{k}$. In particular, we have (a) that $x \neq \overline{n}$. We also have $\neg(x = \overline{0} \vee x = \overline{1} \vee \cdots \vee x = \overline{n-1})$ and so by Lemma 11.24, (b) $\neg(x < \overline{n})$. By Lemma 11.25, $\overline{n} < x$. Since $\mathbf{Q} \vdash \mathsf{Ref}_T(\overline{n}, \ulcorner \rho_T \urcorner)$, we have $\overline{n} < x \wedge \mathsf{Ref}_T(\overline{n}, \ulcorner \rho_T \urcorner)$, and from that $\exists z\,(z < x \wedge \mathsf{Ref}_T(z, \ulcorner \rho_T \urcorner))$. Since x was arbitrary we get, as required, that

$$\forall x\,(\mathsf{Prf}_T(x, \ulcorner \rho_T \urcorner) \to \exists z\,(z < x \wedge \mathsf{Ref}_T(z, \ulcorner \rho_T \urcorner))). \qquad \square$$

12.5 Comparison with Gödel's Original Paper

It is worthwhile to spend some time with Gödel's 1931 paper. The introduction sketches the ideas we have just discussed. Even if you just skim through the paper, it is easy to see what is going on at each stage: first Gödel describes the formal system P (syntax, axioms, proof rules); then he defines the primitive recursive functions and relations; then he shows that xBy is primitive recursive, and argues that the primitive recursive functions and relations are represented in **P**. He then goes on to prove the incompleteness theorem, as above. In Section 3, he shows that one can take the unprovable assertion to be a sentence in the language of arithmetic. This is the origin of the β-lemma, which is what we also used to handle sequences in showing that the recursive functions are representable in **Q**. Gödel doesn't go so far to isolate a minimal set of axioms that suffice, but we now know that **Q** will do the trick. Finally, in Section 4, he sketches a proof of the second incompleteness theorem.

12.6 The Derivability Conditions for PA

Peano arithmetic, or **PA**, is the theory extending **Q** with induction axioms for all formulas. In other words, one adds to **Q** axioms of the form

$$(\varphi(0) \wedge \forall x\,(\varphi(x) \to \varphi(x'))) \to \forall x\,\varphi(x)$$

for every formula φ. Notice that this is really a *schema*, which is to say, infinitely many axioms (and it turns out that **PA** is *not* finitely axiomatizable). But since one can effectively determine whether or not a string of symbols is an instance of an induction axiom, the set of axioms for **PA** is computable. **PA** is a much more robust theory than **Q**. For example, one can easily prove that addition and multiplication are commutative, using induction in the usual way. In fact, most finitary number-theoretic and combinatorial arguments can be carried out in **PA**.

Since **PA** is computably axiomatized, the derivability predicate $\mathrm{Prf}_{\mathbf{PA}}(x, y)$ is computable and hence represented in **Q** (and so, in **PA**). As before, we will take $\mathsf{Prf}_{\mathbf{PA}}(x, y)$ to denote the formula representing the relation. Let $\mathsf{Prov}_{\mathbf{PA}}(y)$ be the formula $\exists x\, \mathrm{Prf}_{\mathbf{PA}}(x, y)$, which, intuitively says, "y is derivable from the axioms of **PA**." The reason we need a little bit more than the axioms of **Q** is we need to know that the theory we are using is strong enough to derive a few basic facts about this derivability predicate. In fact, what we need are the following facts:

P1. If $\mathbf{PA} \vdash \varphi$, then $\mathbf{PA} \vdash \mathsf{Prov}_{\mathbf{PA}}(\ulcorner \varphi \urcorner)$.

P2. For all formulas φ and ψ,

$$\mathbf{PA} \vdash \mathsf{Prov}_{\mathbf{PA}}(\ulcorner \varphi \to \psi \urcorner) \to (\mathsf{Prov}_{\mathbf{PA}}(\ulcorner \varphi \urcorner) \to \mathsf{Prov}_{\mathbf{PA}}(\ulcorner \psi \urcorner)).$$

P3. For every formula φ,

$$\mathbf{PA} \vdash \mathsf{Prov}_{\mathbf{PA}}(\ulcorner \varphi \urcorner) \to \mathsf{Prov}_{\mathbf{PA}}(\ulcorner \mathsf{Prov}_{\mathbf{PA}}(\ulcorner \varphi \urcorner) \urcorner).$$

The only way to verify that these three properties hold is to describe the formula $\mathsf{Prov}_{\mathbf{PA}}(y)$ carefully and use the axioms of **PA** to describe the relevant formal derivations. Conditions (1) and (2) are easy; it is really condition (3) that requires work. (Think about what kind of work it entails ...) Carrying out the details would be tedious and uninteresting, so here we will ask you to take it on faith that **PA** has the three properties listed above. A reasonable choice of $\mathsf{Prov}_{\mathbf{PA}}(y)$ will also satisfy

P4. If $\mathbf{PA} \vdash \mathsf{Prov}_{\mathbf{PA}}(\ulcorner \varphi \urcorner)$, then $\mathbf{PA} \vdash \varphi$.

But we will not need this fact.

Incidentally, Gödel was lazy in the same way we are being now. At the end of the 1931 paper, he sketches the proof of the second incompleteness theorem, and promises the details in a later paper. He never got around to it; since everyone who understood the argument believed that it could be carried out (he did not need to fill in the details.)

12.7 The Second Incompleteness Theorem

How can we express the assertion that **PA** doesn't prove its own consistency? Saying **PA** is inconsistent amounts to saying that $\mathbf{PA} \vdash 0 = 1$. So we can take the consistency statement $\mathsf{Con}_{\mathbf{PA}}$ to be the sentence $\neg\mathsf{Prov}_{\mathbf{PA}}(\ulcorner 0 = 1 \urcorner)$, and then the following theorem does the job:

Theorem 12.8. *Assuming* **PA** *is consistent, then* **PA** *does not derive* $\mathsf{Con}_{\mathbf{PA}}$.

It is important to note that the theorem depends on the particular representation of $\mathsf{Con}_{\mathbf{PA}}$ (i.e., the particular representation of $\mathsf{Prov}_{\mathbf{PA}}(y)$). All we will use is that the representation of $\mathsf{Prov}_{\mathbf{PA}}(y)$ satisfies the three derivability conditions, so the theorem generalizes to any theory with a derivability predicate having these properties.

It is informative to read Gödel's sketch of an argument, since the theorem follows like a good punch line. It goes like this. Let $\gamma_{\mathbf{PA}}$ be the Gödel sentence that we constructed in the proof of Theorem 12.6. We have shown "If **PA** is consistent, then **PA** does not derive $\gamma_{\mathbf{PA}}$." If we formalize this *in* **PA**, we have a proof of

$$\mathsf{Con}_{\mathbf{PA}} \to \neg\mathsf{Prov}_{\mathbf{PA}}(\ulcorner \gamma_{\mathbf{PA}} \urcorner).$$

Now suppose **PA** derives $\mathsf{Con}_{\mathbf{PA}}$. Then it derives $\neg\mathsf{Prov}_{\mathbf{PA}}(\ulcorner \gamma_{\mathbf{PA}} \urcorner)$. But since $\gamma_{\mathbf{PA}}$ is a Gödel sentence, this is equivalent to $\gamma_{\mathbf{PA}}$. So **PA** derives $\gamma_{\mathbf{PA}}$.

But: we know that if **PA** is consistent, it doesn't derive $\gamma_{\mathbf{PA}}$! So if **PA** is consistent, it can't derive $\mathsf{Con}_{\mathbf{PA}}$.

To make the argument more precise, we will let $\gamma_{\mathbf{PA}}$ be the Gödel sentence for **PA** and use the derivability conditions (P1)–(P3) to show that **PA** derives $\mathsf{Con}_{\mathbf{PA}} \to \gamma_{\mathbf{PA}}$. This will show that **PA** doesn't derive $\mathsf{Con}_{\mathbf{PA}}$. Here is a sketch of the proof, in **PA**. (For simplicity, we drop the **PA** subscripts.)

$$!G \leftrightarrow \neg\mathsf{Prov}(\ulcorner\gamma\urcorner) \tag{12.5}$$

γ is a Gödel sentence

$$!G \to \neg\mathsf{Prov}(\ulcorner\gamma\urcorner) \tag{12.6}$$

from eq. (12.5)

$$!G \to (\mathsf{Prov}(\ulcorner\gamma\urcorner) \to \bot) \tag{12.7}$$

from eq. (12.6) by logic

$$\mathsf{Prov}(\ulcorner\gamma \to (\mathsf{Prov}(\ulcorner\gamma\urcorner) \to \bot)\urcorner) \tag{12.8}$$

by from eq. (12.7) by condition P1

$$\mathsf{Prov}(\ulcorner\gamma\urcorner) \to \mathsf{Prov}(\ulcorner(\mathsf{Prov}(\ulcorner\gamma\urcorner) \to \bot)\urcorner) \tag{12.9}$$

from eq. (12.8) by condition P2

$$\mathsf{Prov}(\ulcorner\gamma\urcorner) \to (\mathsf{Prov}(\ulcorner\mathsf{Prov}(\ulcorner\gamma\urcorner)\urcorner) \to \mathsf{Prov}(\ulcorner\bot\urcorner)) \tag{12.10}$$

from eq. (12.9) by condition P2 and logic

$$\mathsf{Prov}(\ulcorner\gamma\urcorner) \to \mathsf{Prov}(\ulcorner\mathsf{Prov}(\ulcorner\gamma\urcorner)\urcorner) \tag{12.11}$$

by P3

$$\mathsf{Prov}(\ulcorner\gamma\urcorner) \to \mathsf{Prov}(\ulcorner\bot\urcorner) \tag{12.12}$$

from eq. (12.10) and eq. (12.11) by logic

$$\mathsf{Con} \to \neg\mathsf{Prov}(\ulcorner\gamma\urcorner) \tag{12.13}$$

contraposition of eq. (12.12) and $\mathsf{Con} \equiv \neg\mathsf{Prov}(\ulcorner\bot\urcorner)$

$$\mathsf{Con} \to \gamma$$

from eq. (12.5) and eq. (12.13) by logic

The use of logic in the above just elementary facts from propositional logic, e.g., eq. (12.7) uses $\vdash \neg\varphi \leftrightarrow (\varphi \to \bot)$ and eq. (12.12) uses $\varphi \to (\psi \to \chi), \varphi \to \psi \vdash \varphi \to \chi$. The use of condition P2 in eq. (12.9) and eq. (12.10) relies on instances of P2, $\mathsf{Prov}(\ulcorner\varphi \to \psi\urcorner) \to (\mathsf{Prov}(\ulcorner\varphi\urcorner) \to \mathsf{Prov}(\ulcorner\psi\urcorner))$. In the first one, $\varphi \equiv \gamma$ and $\psi \equiv \mathsf{Prov}(\ulcorner\gamma\urcorner) \to \bot$; in the second, $\varphi \equiv \mathsf{Prov}(\ulcorner G\urcorner)$ and $\psi \equiv \bot$.

The more abstract version of the second incompleteness theorem is as follows:

Theorem 12.9. *Let* **T** *be any consistent, axiomatized theory extending* **Q** *and let* $\mathsf{Prov}_T(y)$ *be any formula satisfying derivability conditions P1–P3 for* **T**. *Then* **T** *does not derive* Con_T.

The moral of the story is that no "reasonable" consistent theory for mathematics can derive its own consistency statement. Suppose **T** is a theory of

mathematics that includes **Q** and Hilbert's "finitary" reasoning (whatever that may be). Then, the whole of **T** cannot derive the consistency statement of **T**, and so, a fortiori, the finitary fragment can't derive the consistency statement of **T** either. In that sense, there cannot be a finitary consistency proof for "all of mathematics."

There is some leeway in interpreting the term "finitary," and Gödel, in the 1931 paper, grants the possibility that something we may consider "finitary" may lie outside the kinds of mathematics Hilbert wanted to formalize. But Gödel was being charitable; today, it is hard to see how we might find something that can reasonably be called finitary but is not formalizable in, say, **ZFC**, Zermelo-Fraenkel set theory with the axiom of choice.

12.8 Löb's Theorem

The Gödel sentence for a theory **T** is a fixed point of $\neg \mathsf{Prov}_T(y)$, i.e., a sentence γ such that

$$\mathbf{T} \vdash \neg \mathsf{Prov}_T(\ulcorner \gamma \urcorner) \leftrightarrow \gamma.$$

It is not derivable, because if $\mathbf{T} \vdash \gamma$, (a) by derivability condition (1), $\mathbf{T} \vdash \mathsf{Prov}_T(\ulcorner \gamma \urcorner)$, and (b) $\mathbf{T} \vdash \gamma$ together with $\mathbf{T} \vdash \neg \mathsf{Prov}_T(\ulcorner \gamma \urcorner) \leftrightarrow \gamma$ gives $\mathbf{T} \vdash \neg \mathsf{Prov}_T(\ulcorner \gamma \urcorner)$, and so **T** would be inconsistent. Now it is natural to ask about the status of a fixed point of $\mathsf{Prov}_T(y)$, i.e., a sentence δ such that

$$\mathbf{T} \vdash \mathsf{Prov}_T(\ulcorner \delta \urcorner) \leftrightarrow \delta.$$

If it were derivable, $\mathbf{T} \vdash \mathsf{Prov}_T(\ulcorner \delta \urcorner)$ by condition (1), but the same conclusion follows if we apply modus ponens to the equivalence above. Hence, we don't get that **T** is inconsistent, at least not by the same argument as in the case of the Gödel sentence. This of course does not show that **T** *does* derive δ.

We can make headway on this question if we generalize it a bit. The left-to-right direction of the fixed point equivalence, $\mathsf{Prov}_T(\ulcorner \delta \urcorner) \to \delta$, is an instance of a general schema called a *reflection principle*: $\mathsf{Prov}_T(\ulcorner \varphi \urcorner) \to \varphi$. It is called that because it expresses, in a sense, that **T** can "reflect" about what it can derive; basically it says, "If **T** can derive φ, then φ is true," for any φ. This is true for sound theories only, of course, and this suggests that theories will in general not derive every instance of it. So which instances can a theory (strong enough, and satisfying the derivability conditions) derive? Certainly all those where φ itself is derivable. And that's it, as the next result shows.

Theorem 12.10. *Let* **T** *be an axiomatizable theory extending* **Q**, *and suppose* $\mathsf{Prov}_T(y)$ *is a formula satisfying conditions P1–P3 from section 12.7. If* **T** *derives* $\mathsf{Prov}_T(\ulcorner \varphi \urcorner) \to \varphi$, *then in fact* **T** *derives* φ.

Put differently, if $\mathbf{T} \nvdash \varphi$, then $\mathbf{T} \nvdash \mathsf{Prov}_T(\ulcorner \varphi \urcorner) \to \varphi$. This result is known as Löb's theorem.

The heuristic for the proof of Löb's theorem is a clever proof that Santa Claus exists. (If you don't like that conclusion, you are free to substitute any other conclusion you would like.) Here it is:

1. Let X be the sentence, "If X is true, then Santa Claus exists."

2. Suppose X is true.

3. Then what it says holds; i.e., we have: if X is true, then Santa Claus exists.

4. Since we are assuming X is true, we can conclude that Santa Claus exists, by modus ponens from (2) and (3).

5. We have succeeded in deriving (4), "Santa Claus exists," from the assumption (2), "X is true." By conditional proof, we have shown: "If X is true, then Santa Claus exists."

6. But this is just the sentence X. So we have shown that X is true.

7. But then, by the argument (2)–(4) above, Santa Claus exists.

A formalization of this idea, replacing "is true" with "is derivable," and "Santa Claus exists" with φ, yields the proof of Löb's theorem. The trick is to apply the fixed-point lemma to the formula $\mathsf{Prov}_T(y) \to \varphi$. The fixed point of that corresponds to the sentence X in the preceding sketch.

Proof of Theorem 12.10. Suppose φ is a sentence such that $\mathbf{T}$ derives $\mathsf{Prov}_T(\ulcorner \varphi \urcorner) \to \varphi$. Let $\psi(y)$ be the formula $\mathsf{Prov}_T(y) \to \varphi$, and use the fixed-point lemma to find a sentence θ such that $\mathbf{T}$ derives $\theta \leftrightarrow \psi(\ulcorner \theta \urcorner)$. Then each of the following

is derivable in **T**:

$$!D \leftrightarrow (\mathsf{Prov}_T(\ulcorner\theta\urcorner) \to \varphi) \tag{12.14}$$

θ is a fixed point of $\psi(y)$

$$!D \to (\mathsf{Prov}_T(\ulcorner\theta\urcorner) \to \varphi) \tag{12.15}$$

from eq. (12.14)

$$\mathsf{Prov}_T(\ulcorner\theta \to (\mathsf{Prov}_T(\ulcorner\theta\urcorner) \to \varphi)\urcorner) \tag{12.16}$$

from eq. (12.15) by condition P1

$$\mathsf{Prov}_T(\ulcorner\theta\urcorner) \to \mathsf{Prov}_T(\ulcorner\mathsf{Prov}_T(\ulcorner\theta\urcorner) \to \varphi\urcorner) \tag{12.17}$$

from eq. (12.16) using condition P2

$$\mathsf{Prov}_T(\ulcorner\theta\urcorner) \to (\mathsf{Prov}_T(\ulcorner\mathsf{Prov}_T(\ulcorner\theta\urcorner)\urcorner) \to \mathsf{Prov}_T(\ulcorner\varphi\urcorner)) \tag{12.18}$$

from eq. (12.17) using P2 again

$$\mathsf{Prov}_T(\ulcorner\theta\urcorner) \to \mathsf{Prov}_T(\ulcorner\mathsf{Prov}_T(\ulcorner\theta\urcorner)\urcorner) \tag{12.19}$$

by derivability condition P3

$$\mathsf{Prov}_T(\ulcorner\theta\urcorner) \to \mathsf{Prov}_T(\ulcorner\varphi\urcorner) \tag{12.20}$$

from eq. (12.18) and eq. (12.19)

$$\mathsf{Prov}_T(\ulcorner\varphi\urcorner) \to \varphi \tag{12.21}$$

by assumption of the theorem

$$\mathsf{Prov}_T(\ulcorner\theta\urcorner) \to \varphi \tag{12.22}$$

from eq. (12.20) and eq. (12.21)

$$(\mathsf{Prov}_T(\ulcorner\theta\urcorner) \to \varphi) \to \theta \tag{12.23}$$

from eq. (12.14)

$$!D \tag{12.24}$$

from eq. (12.22) and eq. (12.23)

$$\mathsf{Prov}_T(\ulcorner\theta\urcorner) \tag{12.25}$$

from eq. (12.24) by condition P1

$!A$ from eq. (12.21) and eq. (12.25) □

With Löb's theorem in hand, there is a short proof of the second incompleteness theorem (for theories having a derivability predicate satisfying conditions P1–P3): if $\mathbf{T} \vdash \mathsf{Prov}_T(\ulcorner\bot\urcorner) \to \bot$, then $\mathbf{T} \vdash \bot$. If **T** is consistent, $\mathbf{T} \nvdash \bot$. So, $\mathbf{T} \nvdash \mathsf{Prov}_T(\ulcorner\bot\urcorner) \to \bot$, i.e., $\mathbf{T} \nvdash \mathsf{Con}_{\mathbf{T}}$. We can also apply it to show that δ, the fixed point of $\mathsf{Prov}_T(x)$, is derivable. For since

$$\mathbf{T} \vdash \mathsf{Prov}_T(\ulcorner\delta\urcorner) \leftrightarrow \delta$$

in particular

$$\mathbf{T} \vdash \mathsf{Prov}_T(\ulcorner\delta\urcorner) \to \delta$$

and so by Löb's theorem, $\mathbf{T} \vdash \delta$.

12.9 The Undefinability of Truth

The notion of *definability* depends on having a formal semantics for the language of arithmetic. We have described a set of formulas and sentences in the language of arithmetic. The "intended interpretation" is to read such sentences as making assertions about the natural numbers, and such an assertion can be true or false. Let $\mathfrak{N}$ be the structure with domain $\mathbb{N}$ and the standard interpretation for the symbols in the language of arithmetic. Then $\mathfrak{N} \vDash \varphi$ means "φ is true in the standard interpretation."

Definition 12.11. A relation $R(x_1, \ldots, x_k)$ of natural numbers is *definable* in $\mathfrak{N}$ if and only if there is a formula $\varphi(x_1, \ldots, x_k)$ in the language of arithmetic such that for every $n_1, \ldots, n_k$, $R(n_1, \ldots, n_k)$ if and only if $\mathfrak{N} \vDash \varphi(\overline{n}_1, \ldots, \overline{n}_k)$.

Put differently, a relation is definable in $\mathfrak{N}$ if and only if it is representable in the theory **TA**, where $\mathbf{TA} = \{\varphi : \mathfrak{N} \vDash \varphi\}$ is the set of true sentences of arithmetic. (If this is not immediately clear to you, you should go back and check the definitions and convince yourself that this is the case.)

Lemma 12.12. *Every computable relation is definable in* $\mathfrak{N}$.

Proof. It is easy to check that the formula representing a relation in **Q** defines the same relation in $\mathfrak{N}$. □

Now one can ask, is the converse also true? That is, is every relation definable in $\mathfrak{N}$ computable? The answer is no. For example:

Lemma 12.13. *The halting relation is definable in* $\mathfrak{N}$.

Proof. Let H be the halting relation, i.e.,

$$H = \{\langle e, x\rangle : \exists s\, T(e, x, s)\}.$$

Let θ_T define T in $\mathfrak{N}$. Then

$$H = \{\langle e, x\rangle : \mathfrak{N} \vDash \exists s\, \theta_T(\overline{e}, \overline{x}, s)\},$$

so $\exists s\, \theta_T(z, x, s)$ defines H in $\mathfrak{N}$. □

What about **TA** itself? Is it definable in arithmetic? That is: is the set $\{^{\#}\varphi^{\#} : \mathfrak{N} \vDash \varphi\}$ definable in arithmetic? Tarski's theorem answers this in the negative.

Theorem 12.14. *The set of true sentences of arithmetic is not definable in arithmetic.*

Proof. Suppose $\theta(x)$ defined it, i.e., $\mathfrak{N} \vDash \varphi$ iff $\mathfrak{N} \vDash \theta(\ulcorner\varphi\urcorner)$. By the fixed-point lemma, there is a formula φ such that $\mathbf{Q} \vdash \varphi \leftrightarrow \neg\theta(\ulcorner\varphi\urcorner)$, and hence $\mathfrak{N} \vDash \varphi \leftrightarrow \neg\theta(\ulcorner\varphi\urcorner)$. But then $\mathfrak{N} \vDash \varphi$ if and only if $\mathfrak{N} \vDash \neg\theta(\ulcorner\varphi\urcorner)$, which contradicts the fact that $\theta(y)$ is supposed to define the set of true statements of arithmetic. □

Tarski applied this analysis to a more general philosophical notion of truth. Given any language L, Tarski argued that an adequate notion of truth for L would have to satisfy, for each sentence X,

'X' is true if and only if X.

Tarski's oft-quoted example, for English, is the sentence

'Snow is white' is true if and only if snow is white.

However, for any language strong enough to represent the diagonal function, and any linguistic predicate $T(x)$, we can construct a sentence X satisfying "X if and only if not $T(\text{'}X\text{'})$." Given that we do not want a truth predicate to declare some sentences to be both true and false, Tarski concluded that one cannot specify a truth predicate for all sentences in a language without, somehow, stepping outside the bounds of the language. In other words, a the truth predicate for a language cannot be defined in the language itself.

Problems

Problems for Chapter 1

Problem 1.1. Prove that there is at most one empty set, i.e., show that if A and B are sets without elements, then $A = B$.

Problem 1.2. List all subsets of $\{a, b, c, d\}$.

Problem 1.3. Show that if A has n elements, then $\wp(A)$ has 2^n elements.

Problem 1.4. Prove that if $A \subseteq B$, then $A \cup B = B$.

Problem 1.5. Prove rigorously that if $A \subseteq B$, then $A \cap B = A$.

Problem 1.6. Show that if A is a set and $A \in B$, then $A \subseteq \bigcup B$.

Problem 1.7. Prove that if $A \subsetneq B$, then $B \setminus A \neq \emptyset$.

Problem 1.8. Using Definition 1.23, prove that $\langle a, b\rangle = \langle c, d\rangle$ iff both $a = c$ and $b = d$.

Problem 1.9. List all elements of $\{1, 2, 3\}^3$.

Problem 1.10. Show, by induction on k, that for all $k \geq 1$, if A has n elements, then A^k has n^k elements.

Problems for Chapter 2

Problem 2.1. List the elements of the relation $\subseteq$ on the set $\wp(\{a, b, c\})$.

Problem 2.2. Give examples of relations that are (a) reflexive and symmetric but not transitive, (b) reflexive and anti-symmetric, (c) anti-symmetric, transitive, but not reflexive, and (d) reflexive, symmetric, and transitive. Do not use relations on numbers or sets.

Problem 2.3. Show that $\equiv_n$ is an equivalence relation, for any $n \in \mathbb{N}$, and that $\mathbb{N}/_{\equiv_n}$ has exactly n members.

Problem 2.4. Give a proof of Proposition 2.26.

Problem 2.5. Consider the less-than-or-equal-to relation $\leq$ on the set $\{1, 2, 3, 4\}$ as a graph and draw the corresponding diagram.

Problem 2.6. Show that the transitive closure of R is in fact transitive.

Problems for Chapter 3

Problem 3.1. Show that if $f\colon A \to B$ has a left inverse g, then f is injective.

Problem 3.2. Show that if $f\colon A \to B$ has a right inverse h, then f is surjective.

Problem 3.3. Prove Proposition 3.18. You have to define f^{-1}, show that it is a function, and show that it is an inverse of f, i.e., $f^{-1}(f(x)) = x$ and $f(f^{-1}(y)) = y$ for all $x \in A$ and $y \in B$.

Problem 3.4. Prove Proposition 3.19.

Problem 3.5. Show that if $f\colon A \to B$ and $g\colon B \to C$ are both injective, then $g \circ f\colon A \to C$ is injective.

Problem 3.6. Show that if $f\colon A \to B$ and $g\colon B \to C$ are both surjective, then $g \circ f\colon A \to C$ is surjective.

Problem 3.7. Suppose $f\colon A \to B$ and $g\colon B \to C$. Show that the graph of $g \circ f$ is $R_f \mid R_g$.

Problem 3.8. Given $f\colon A \pfun B$, define the partial function $g\colon B \pfun A$ by: for any $y \in B$, if there is a unique $x \in A$ such that $f(x) = y$, then $g(y) = x$; otherwise $g(y) \uparrow$. Show that if f is injective, then $g(f(x)) = x$ for all $x \in \mathrm{dom}(f)$, and $f(g(y)) = y$ for all $y \in \mathrm{ran}(f)$.

Problems for Chapter 4

Problem 4.1. Define an enumeration of the positive squares $1, 4, 9, 16, \ldots$

Problem 4.2. Show that if A and B are enumerable, so is $A \cup B$. To do this, suppose there are surjective functions $f\colon \mathbb{Z}^+ \to A$ and $g\colon \mathbb{Z}^+ \to B$, and define a surjective function $h\colon \mathbb{Z}^+ \to A \cup B$ and prove that it is surjective. Also consider the cases where A or $B = \emptyset$.

Problem 4.3. Show that if $B \subseteq A$ and A is enumerable, so is B. To do this, suppose there is a surjective function $f\colon \mathbb{Z}^+ \to A$. Define a surjective function $g\colon \mathbb{Z}^+ \to B$ and prove that it is surjective. What happens if $B = \emptyset$?

Problem 4.4. Show by induction on n that if A_1, A_2, ..., A_n are all enumerable, so is $A_1 \cup \cdots \cup A_n$. You may assume the fact that if two sets A and B are enumerable, so is $A \cup B$.

Problem 4.5. According to Definition 4.4, a set A is enumerable iff $A = \emptyset$ or there is a surjective $f\colon \mathbb{Z}^+ \to A$. It is also possible to define "enumerable set" precisely by: a set is enumerable iff there is an injective function $g\colon A \to \mathbb{Z}^+$. Show that the definitions are equivalent, i.e., show that there is an injective function $g\colon A \to \mathbb{Z}^+$ iff either $A = \emptyset$ or there is a surjective $f\colon \mathbb{Z}^+ \to A$.

Problem 4.6. Show that $(\mathbb{Z}^+)^n$ is enumerable, for every $n \in \mathbb{N}$.

Problem 4.7. Show that $(\mathbb{Z}^+)^*$ is enumerable. You may assume problem 4.6.

Problem 4.8. Give an enumeration of the set of all non-negative rational numbers.

Problem 4.9. Show that $\mathbb{Q}$ is enumerable. Recall that any rational number can be written as a fraction z/m with $z \in \mathbb{Z}$, $m \in \mathbb{N}^+$.

Problem 4.10. Define an enumeration of $\mathbb{B}^*$.

Problem 4.11. Recall from your introductory logic course that each possible truth table expresses a truth function. In other words, the truth functions are all functions from $\mathbb{B}^k \to \mathbb{B}$ for some k. Prove that the set of all truth functions is enumerable.

Problem 4.12. Show that the set of all finite subsets of an arbitrary infinite enumerable set is enumerable.

Problem 4.13. A subset of $\mathbb{N}$ is said to be *cofinite* iff it is the complement of a finite set $\mathbb{N}$; that is, $A \subseteq \mathbb{N}$ is cofinite iff $\mathbb{N} \setminus A$ is finite. Let I be the set whose elements are exactly the finite and cofinite subsets of $\mathbb{N}$. Show that I is enumerable.

Problem 4.14. Show that the enumerable union of enumerable sets is enumerable. That is, whenever A_1, A_2, ... are sets, and each A_i is enumerable, then the union $\bigcup_{i=1}^{\infty} A_i$ of all of them is also enumerable. [NB: this is hard!]

Problem 4.15. Let $f\colon A \times B \to \mathbb{N}$ be an arbitrary pairing function. Show that the inverse of f is an enumeration of $A \times B$.

Problem 4.16. Specify a function that encodes $\mathbb{N}^3$.

Problem 4.17. Show that $\wp(\mathbb{N})$ is non-enumerable by a diagonal argument.

Problem 4.18. Show that the set of functions $f : \mathbb{Z}^+ \to \mathbb{Z}^+$ is non-enumerable by an explicit diagonal argument. That is, show that if f_1, f_2, ..., is a list of functions and each $f_i : \mathbb{Z}^+ \to \mathbb{Z}^+$, then there is some $\overline{f} : \mathbb{Z}^+ \to \mathbb{Z}^+$ not on this list.

Problem 4.19. Show that if there is an injective function $g : B \to A$, and B is non-enumerable, then so is A. Do this by showing how you can use g to turn an enumeration of A into one of B.

Problem 4.20. Show that the set of all *sets of* pairs of positive integers is non-enumerable by a reduction argument.

Problem 4.21. Show that the set X of all functions $f : \mathbb{N} \to \mathbb{N}$ is non-enumerable by a reduction argument (Hint: give a surjective function from X to $\mathbb{B}^\omega$.)

Problem 4.22. Show that $\mathbb{N}^\omega$, the set of infinite sequences of natural numbers, is non-enumerable by a reduction argument.

Problem 4.23. Let P be the set of functions from the set of positive integers to the set $\{0\}$, and let Q be the set of *partial* functions from the set of positive integers to the set $\{0\}$. Show that P is enumerable and Q is not. (Hint: reduce the problem of enumerating $\mathbb{B}^\omega$ to enumerating Q).

Problem 4.24. Let S be the set of all surjective functions from the set of positive integers to the set $\{0,1\}$, i.e., S consists of all surjective $f : \mathbb{Z}^+ \to \mathbb{B}$. Show that S is non-enumerable.

Problem 4.25. Show that the set $\mathbb{R}$ of all real numbers is non-enumerable.

Problem 4.26. Show that if $A \approx C$ and $B \approx D$, and $A \cap B = C \cap D = \emptyset$, then $A \cup B \approx C \cup D$.

Problem 4.27. Show that if A is infinite and enumerable, then $A \approx \mathbb{N}$.

Problem 4.28. Show that there cannot be an injection $g : \wp(A) \to A$, for any set A. Hint: Suppose $g : \wp(A) \to A$ is injective. Consider $D = \{g(B) : B \subseteq A \text{ and } g(B) \notin B\}$. Let $x = g(D)$. Use the fact that g is injective to derive a contradiction.

Problems for Chapter 5

Problem 5.1. Prove Lemma 5.8.

Problem 5.2. Prove that for any term t, $l(t) = r(t)$.

Problem 5.3. Prove Lemma 5.12.

Problem 5.4. Prove Proposition 5.13 (Hint: Formulate and prove a version of Lemma 5.12 for terms.)

Problem 5.5. Prove Proposition 5.19.

Problem 5.6. Prove Proposition 5.20.

Problem 5.7. Prove Lemma 5.28.

Problem 5.8. Prove Proposition 5.30. Hint: use a similar strategy to that used in the proof of Theorem 5.29.

Problem 5.9. Prove Proposition 5.32.

Problem 5.10. Give an inductive definition of the bound variable occurrences along the lines of Definition 5.33.

Problem 5.11. Is $\mathfrak{N}$, the standard model of arithmetic, covered? Explain.

Problem 5.12. Let $\mathcal{L} = \{c, f, A\}$ with one constant symbol, one one-place function symbol and one two-place predicate symbol, and let the structure $\mathfrak{M}$ be given by

1. $|\mathfrak{M}| = \{1, 2, 3\}$
2. $c^{\mathfrak{M}} = 3$
3. $f^{\mathfrak{M}}(1) = 2, f^{\mathfrak{M}}(2) = 3, f^{\mathfrak{M}}(3) = 2$
4. $A^{\mathfrak{M}} = \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 3, 3\rangle\}$

(a) Let $s(v) = 1$ for all variables v. Find out whether

$$\mathfrak{M}, s \vDash \exists x\,(A(f(z), c) \to \forall y\,(A(y, x) \vee A(f(y), x)))$$

Explain why or why not.

(b) Give a different structure and variable assignment in which the formula is not satisfied.

Problem 5.13. Complete the proof of Proposition 5.55.

Problem 5.14. Prove Proposition 5.58

Problem 5.15. Prove Proposition 5.59.

Problem 5.16. Suppose $\mathcal{L}$ is a language without function symbols. Given a structure $\mathfrak{M}$, c a constant symbol and $a \in |\mathfrak{M}|$, define $\mathfrak{M}[a/c]$ to be the structure that is just like $\mathfrak{M}$, except that $c^{\mathfrak{M}[a/c]} = a$. Define $\mathfrak{M} \mid\!\models \varphi$ for sentences φ by:

1. $\varphi \equiv \bot$: not $\mathfrak{M} \mid\!\models \varphi$.
2. $\varphi \equiv R(d_1, \ldots, d_n)$: $\mathfrak{M} \mid\!\models \varphi$ iff $\langle d_1^{\mathfrak{M}}, \ldots, d_n^{\mathfrak{M}} \rangle \in R^{\mathfrak{M}}$.
3. $\varphi \equiv d_1 = d_2$: $\mathfrak{M} \mid\!\models \varphi$ iff $d_1^{\mathfrak{M}} = d_2^{\mathfrak{M}}$.
4. $\varphi \equiv \neg\psi$: $\mathfrak{M} \mid\!\models \varphi$ iff not $\mathfrak{M} \mid\!\models \psi$.
5. $\varphi \equiv (\psi \wedge \chi)$: $\mathfrak{M} \mid\!\models \varphi$ iff $\mathfrak{M} \mid\!\models \psi$ and $\mathfrak{M} \mid\!\models \chi$.
6. $\varphi \equiv (\psi \vee \chi)$: $\mathfrak{M} \mid\!\models \varphi$ iff $\mathfrak{M} \mid\!\models \psi$ or $\mathfrak{M} \mid\!\models \chi$ (or both).
7. $\varphi \equiv (\psi \to \chi)$: $\mathfrak{M} \mid\!\models \varphi$ iff not $\mathfrak{M} \mid\!\models \psi$ or $\mathfrak{M} \mid\!\models \chi$ (or both).
8. $\varphi \equiv \forall x\, \psi$: $\mathfrak{M} \mid\!\models \varphi$ iff for all $a \in |\mathfrak{M}|$, $\mathfrak{M}[a/c] \mid\!\models \psi[c/x]$, if c does not occur in ψ.
9. $\varphi \equiv \exists x\, \psi$: $\mathfrak{M} \mid\!\models \varphi$ iff there is an $a \in |\mathfrak{M}|$ such that $\mathfrak{M}[a/c] \mid\!\models \psi[c/x]$, if c does not occur in ψ.

Let $x_1, \ldots, x_n$ be all free variables in φ, $c_1, \ldots, c_n$ constant symbols not in φ, $a_1, \ldots, a_n \in |\mathfrak{M}|$, and $s(x_i) = a_i$.

Show that $\mathfrak{M}, s \vDash \varphi$ iff $\mathfrak{M}[a_1/c_1, \ldots, a_n/c_n] \mid\!\models \varphi[c_1/x_1] \ldots [c_n/x_n]$.

(This problem shows that it is possible to give a semantics for first-order logic that makes do without variable assignments.)

Problem 5.17. Suppose that f is a function symbol not in $\varphi(x, y)$. Show that there is a structure $\mathfrak{M}$ such that $\mathfrak{M} \vDash \forall x\, \exists y\, \varphi(x, y)$ iff there is an $\mathfrak{M}'$ such that $\mathfrak{M}' \vDash \forall x\, \varphi(x, f(x))$.

(This problem is a special case of what's known as Skolem's Theorem; $\forall x\, \varphi(x, f(x))$ is called a *Skolem normal form* of $\forall x\, \exists y\, \varphi(x, y)$.)

Problem 5.18. Carry out the proof of Proposition 5.60 in detail.

Problem 5.19. Prove Proposition 5.63

Problem 5.20.

1. Show that $\Gamma \vDash \bot$ iff Γ is unsatisfiable.
2. Show that $\Gamma \cup \{\varphi\} \vDash \bot$ iff $\Gamma \vDash \neg\varphi$.
3. Suppose c does not occur in φ or Γ. Show that $\Gamma \vDash \forall x\, \varphi$ iff $\Gamma \vDash \varphi[c/x]$.

Problem 5.21. Complete the proof of Proposition 5.71.

Problems for Chapter 6

Problem 6.1. Find formulas in $\mathcal{L}_A$ which define the following relations:

1. n is between i and j;
2. n evenly divides m (i.e., m is a multiple of n);
3. n is a prime number (i.e., no number other than 1 and n evenly divides n).

Problem 6.2. Suppose the formula $\varphi(v_1, v_2)$ expresses the relation $R \subseteq |\mathfrak{M}|^2$ in a structure $\mathfrak{M}$. Find formulas that express the following relations:

1. the inverse R^{-1} of R;
2. the relative product $R \mid R$;

Can you find a way to express R^+, the transitive closure of R?

Problem 6.3. Let $\mathcal{L}$ be the language containing a 2-place predicate symbol $<$ only (no other constant symbols, function symbols or predicate symbols—except of course $=$). Let $\mathfrak{N}$ be the structure such that $|\mathfrak{N}| = \mathbb{N}$, and $<^{\mathfrak{N}} = \{\langle n, m\rangle : n < m\}$. Prove the following:

1. $\{0\}$ is definable in $\mathfrak{N}$;
2. $\{1\}$ is definable in $\mathfrak{N}$;
3. $\{2\}$ is definable in $\mathfrak{N}$;
4. for each $n \in \mathbb{N}$, the set $\{n\}$ is definable in $\mathfrak{N}$;
5. every finite subset of $|\mathfrak{N}|$ is definable in $\mathfrak{N}$;
6. every co-finite subset of $|\mathfrak{N}|$ is definable in $\mathfrak{N}$ (where $X \subseteq \mathbb{N}$ is co-finite iff $\mathbb{N} \setminus X$ is finite).

Problem 6.4. Show that the comprehension principle is inconsistent by giving a derivation that shows

$$\exists y \, \forall x \, (x \in y \leftrightarrow x \notin x) \vdash \bot.$$

It may help to first show $(A \to \neg A) \wedge (\neg A \to A) \vdash \bot$.

Problems for Chapter 7

Problem 7.1. Give derivations of the following sequents:

1. $\varphi \wedge (\psi \wedge \chi) \Rightarrow (\varphi \wedge \psi) \wedge \chi$.
2. $\varphi \vee (\psi \vee \chi) \Rightarrow (\varphi \vee \psi) \vee \chi$.
3. $\varphi \to (\psi \to \chi) \Rightarrow \psi \to (\varphi \to \chi)$.
4. $\varphi \Rightarrow \neg\neg\varphi$.

Problem 7.2. Give derivations of the following sequents:

1. $(\varphi \vee \psi) \to \chi \Rightarrow \varphi \to \chi$.
2. $(\varphi \to \chi) \wedge (\psi \to \chi) \Rightarrow (\varphi \vee \psi) \to \chi$.
3. $\Rightarrow \neg(\varphi \wedge \neg\varphi)$.
4. $\psi \to \varphi \Rightarrow \neg\varphi \to \neg\psi$.
5. $\Rightarrow (\varphi \to \neg\varphi) \to \neg\varphi$.
6. $\Rightarrow \neg(\varphi \to \psi) \to \neg\psi$.
7. $\varphi \to \chi \Rightarrow \neg(\varphi \wedge \neg\chi)$.
8. $\varphi \wedge \neg\chi \Rightarrow \neg(\varphi \to \chi)$.
9. $\varphi \vee \psi, \neg\psi \Rightarrow \varphi$.
10. $\neg\varphi \vee \neg\psi \Rightarrow \neg(\varphi \wedge \psi)$.
11. $\Rightarrow (\neg\varphi \wedge \neg\psi) \to \neg(\varphi \vee \psi)$.
12. $\Rightarrow \neg(\varphi \vee \psi) \to (\neg\varphi \wedge \neg\psi)$.

Problem 7.3. Give derivations of the following sequents:

1. $\neg(\varphi \to \psi) \Rightarrow \varphi$.
2. $\neg(\varphi \wedge \psi) \Rightarrow \neg\varphi \vee \neg\psi$.
3. $\varphi \to \psi \Rightarrow \neg\varphi \vee \psi$.
4. $\Rightarrow \neg\neg\varphi \to \varphi$.
5. $\varphi \to \psi, \neg\varphi \to \psi \Rightarrow \psi$.
6. $(\varphi \wedge \psi) \to \chi \Rightarrow (\varphi \to \chi) \vee (\psi \to \chi)$.
7. $(\varphi \to \psi) \to \varphi \Rightarrow \varphi$.

8. $\Rightarrow (\varphi \to \psi) \vee (\psi \to \chi)$.

(These all require the CR rule.)

Problem 7.4. Give derivations of the following sequents:

1. $\Rightarrow (\forall x\, \varphi(x) \wedge \forall y\, \psi(y)) \to \forall z\, (\varphi(z) \wedge \psi(z))$.
2. $\Rightarrow (\exists x\, \varphi(x) \vee \exists y\, \psi(y)) \to \exists z\, (\varphi(z) \vee \psi(z))$.
3. $\forall x\, (\varphi(x) \to \psi) \Rightarrow \exists y\, \varphi(y) \to \psi$.
4. $\forall x\, \neg\varphi(x) \Rightarrow \neg\exists x\, \varphi(x)$.
5. $\Rightarrow \neg\exists x\, \varphi(x) \to \forall x\, \neg\varphi(x)$.
6. $\Rightarrow \neg\exists x\, \forall y\, ((\varphi(x,y) \to \neg\varphi(y,y)) \wedge (\neg\varphi(y,y) \to \varphi(x,y)))$.

Problem 7.5. Give derivations of the following sequents:

1. $\Rightarrow \neg\forall x\, \varphi(x) \to \exists x\, \neg\varphi(x)$.
2. $(\forall x\, \varphi(x) \to \psi) \Rightarrow \exists y\, (\varphi(y) \to \psi)$.
3. $\Rightarrow \exists x\, (\varphi(x) \to \forall y\, \varphi(y))$.

(These all require the CR rule.)

Problem 7.6. Prove Proposition 7.16

Problem 7.7. Prove that $\Gamma \vdash \neg\varphi$ iff $\Gamma \cup \{\varphi\}$ is inconsistent.

Problem 7.8. Complete the proof of Theorem 7.28.

Problem 7.9. Give derivations of the following sequents:

1. $\Rightarrow \forall x\, \forall y\, ((x = y \wedge \varphi(x)) \to \varphi(y))$
2. $\exists x\, \varphi(x) \wedge \forall y\, \forall z\, ((\varphi(y) \wedge \varphi(z)) \to y = z) \Rightarrow \exists x\, (\varphi(x) \wedge \forall y\, (\varphi(y) \to y = x))$

Problems for Chapter 8

Problem 8.1. Complete the proof of Proposition 8.2.

Problem 8.2. Complete the proof of Proposition 8.11.

Problem 8.3. Complete the proof of Lemma 8.12.

Problem 8.4. Complete the proof of Proposition 8.14.

Problem 8.5. Complete the proof of Lemma 8.18.

Problem 8.6. Use Corollary 8.21 to prove Theorem 8.20, thus showing that the two formulations of the completeness theorem are equivalent.

Problem 8.7. In order for a derivation system to be complete, its rules must be strong enough to prove every unsatisfiable set inconsistent. Which of the rules of derivation were necessary to prove completeness? Are any of these rules not used anywhere in the proof? In order to answer these questions, make a list or diagram that shows which of the rules of derivation were used in which results that lead up to the proof of Theorem 8.20. Be sure to note any tacit uses of rules in these proofs.

Problem 8.8. Prove (1) of Theorem 8.23.

Problem 8.9. In the standard model of arithmetic $\mathfrak{N}$, there is no element $k \in |\mathfrak{N}|$ which satisfies every formula $\overline{n} < x$ (where $\overline{n}$ is $0'^{\cdots'}$ with n $'$s). Use the compactness theorem to show that the set of sentences in the language of arithmetic which are true in the standard model of arithmetic $\mathfrak{N}$ are also true in a structure $\mathfrak{N}'$ that contains an element which *does* satisfy every formula $\overline{n} < x$.

Problem 8.10. Prove Proposition 8.27. Avoid the use of $\vdash$.

Problem 8.11. Prove Lemma 8.28. (Hint: The crucial step is to show that if Γ_n is finitely satisfiable, so is $\Gamma_n \cup \{\theta_n\}$, without any appeal to derivations or consistency.)

Problem 8.12. Prove Proposition 8.29.

Problem 8.13. Prove Lemma 8.30. (Hint: the crucial step is to show that if Γ_n is finitely satisfiable, then either $\Gamma_n \cup \{\varphi_n\}$ or $\Gamma_n \cup \{\neg\varphi_n\}$ is finitely satisfiable.)

Problem 8.14. Write out the complete proof of the Truth Lemma (Lemma 8.12) in the version required for the proof of Theorem 8.31.

Problem 8.15. Let Γ be the set of all sentences φ in the language of arithmetic such that $\mathfrak{N} \vDash \varphi$, i.e., Γ contains all sentences true in the "standard model." Show that there is a model $\mathfrak{M}$ of Γ which is not covered, i.e., some $a \in |\mathfrak{M}|$ is such that $a \neq \mathrm{Val}^{\mathfrak{M}}(t)$ for all closed terms t.

Problems for Chapter 9

Problem 9.1. Prove Proposition 9.5 by showing that the primitive recursive definition of mult can be put into the form required by Definition 9.1 and showing that the corresponding functions f and g are primitive recursive.

Problem 9.2. Give the complete primitive recursive notation for mult.

Problem 9.3. Prove Proposition 9.13.

Problem 9.4. Show that

$$f(x,y) = 2^{(2^{\cdot^{\cdot^{\cdot^{2^x}}}})} \Big\} y \text{ 2's}$$

is primitive recursive.

Problem 9.5. Show that integer division $d(x,y) = \lfloor x/y \rfloor$ (i.e., division, where you disregard everything after the decimal point) is primitive recursive. When $y = 0$, we stipulate $d(x,y) = 0$. Give an explicit definition of d using primitive recursion and composition.

Problem 9.6. Show that the three place relation $x \equiv y \mod n$ (congruence modulo n) is primitive recursive.

Problem 9.7. Suppose $R(\vec{x}, z)$ is primitive recursive. Define the function $m'_R(\vec{x}, y)$ which returns the least z less than y such that $R(\vec{x}, z)$ holds, if there is one, and 0 otherwise, by primitive recursion from χ_R.

Problem 9.8. Define integer division $d(x,y)$ using bounded minimization.

Problem 9.9. Show that there is a primitive recursive function sconcat(s) with the property that

$$\text{sconcat}(\langle s_0, \ldots, s_k \rangle) = s_0 \frown \ldots \frown s_k.$$

Problem 9.10. Show that there is a primitive recursive function tail(s) with the property that

$$\text{tail}(\Lambda) = 0 \text{ and}$$
$$\text{tail}(\langle s_0, \ldots, s_k \rangle) = \langle s_1, \ldots, s_k \rangle.$$

Problem 9.11. Prove Proposition 9.24.

Problem 9.12. The definition of hSubtreeSeq in the proof of Proposition 9.25 in general includes repetitions. Give an alternative definition which guarantees that the code of a subtree occurs only once in the resulting list.

Problem 9.13. Define the remainder function $r(x,y)$ by course-of-values recursion. (If x, y are natural numbers and $y > 0$, $r(x,y)$ is the number less than y such that $x = z \times y + r(x,y)$ for some z. For definiteness, let's say that if $y = 0$, $r(x,0) = 0$.)

Problems for Chapter 10

Problem 10.1. Show that the function flatten(z), which turns the sequence $\langle {}^{\ulcorner}t_1{}^{\urcorner}, \ldots, {}^{\ulcorner}t_n{}^{\urcorner}\rangle$ into ${}^{\ulcorner}t_1, \ldots, t_n{}^{\urcorner}$, is primitive recursive.

Problem 10.2. Give a detailed proof of Proposition 10.8 along the lines of the first proof of Proposition 10.5.

Problem 10.3. Prove Proposition 10.9. You may make use of the fact that any substring of a formula which is a formula is a sub-formula of it.

Problem 10.4. Prove Proposition 10.12

Problem 10.5. Define the following properties as in Proposition 10.15:

1. FollowsBy$_{\text{Cut}}(p)$,
2. FollowsBy$_{\to\text{L}}(p)$,
3. FollowsBy$_{=}(p)$,
4. FollowsBy$_{\forall\text{R}}(p)$.

For the last one, you will have to also show that you can test primitive recursively if the last inference of the derivation with Gödel number p satisfies the eigenvariable condition, i.e., the eigenvariable a of the $\forall$R does not occur in the end-sequent.

Problems for Chapter 11

Problem 11.1. Show that the relations $x < y$, $x \mid y$, and the function rem(x, y) can be defined without primitive recursion. You may use 0, successor, plus, times, $\chi_{=}$, projections, and bounded minimization and quantification.

Problem 11.2. Prove that $y = \mathrm{o}$, $y = x'$, and $y = x_i$ represent zero, succ, and P^n_i, respectively.

Problem 11.3. Prove Lemma 11.18.

Problem 11.4. Use Lemma 11.18 to prove Proposition 11.17.

Problem 11.5. Using the proofs of Proposition 11.20 and Proposition 11.20 as a guide, carry out the proof of Proposition 11.21 in detail.

Problem 11.6. Show that if R is representable in $\mathbf{Q}$, so is χ_R.

Problems for Chapter 12

Problem 12.1. A formula $\varphi(x)$ is a *truth definition* if $\mathbf{Q} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$ for all sentences ψ. Show that no formula is a truth definition by using the fixed-point lemma.

Problem 12.2. Every ω-consistent theory is consistent. Show that the converse does not hold, i.e., that there are consistent but ω-inconsistent theories. Do this by showing that $\mathbf{Q} \cup \{\neg \gamma_{\mathbf{Q}}\}$ is consistent but ω-inconsistent.

Problem 12.3. Two sets A and B of natural numbers are said to be *computably inseparable* if there is no decidable set X such that $A \subseteq X$ and $B \subseteq \overline{X}$ ($\overline{X}$ is the complement, $\mathbb{N} \setminus X$, of X). Let $\mathbf{T}$ be a consistent axiomatizable extension of $\mathbf{Q}$. Suppose A is the set of Gödel numbers of sentences provable in $\mathbf{T}$ and B the set of Gödel numbers of sentences refutable in $\mathbf{T}$. Prove that A and B are computably inseparable.

Problem 12.4. Show that $\mathbf{PA}$ derives $\gamma_{\mathbf{PA}} \to \mathsf{Con}_{\mathbf{PA}}$.

Problem 12.5. Let $\mathbf{T}$ be a computably axiomatized theory, and let Prov_T be a derivability predicate for $\mathbf{T}$. Consider the following four statements:

1. If $T \vdash \varphi$, then $T \vdash \mathsf{Prov}_T(\ulcorner \varphi \urcorner)$.
2. $T \vdash \varphi \to \mathsf{Prov}_T(\ulcorner \varphi \urcorner)$.
3. If $T \vdash \mathsf{Prov}_T(\ulcorner \varphi \urcorner)$, then $T \vdash \varphi$.
4. $T \vdash \mathsf{Prov}_T(\ulcorner \varphi \urcorner) \to \varphi$

Under what conditions are each of these statements true?

Problem 12.6. Show that $Q(n) \Leftrightarrow n \in \{{}^{\natural}\varphi^{\#} : \mathbf{Q} \vdash \varphi\}$ is definable in arithmetic.

Bibliography

Cantor, Georg. 1892. Über eine elementare Frage der Mannigfaltigkeitslehre. *Jahresbericht der deutschen Mathematiker-Vereinigung* 1: 75–8.

Frege, Gottlob. 1884. *Die Grundlagen der Arithmetik: Eine logisch mathematische Untersuchung über den Begriff der Zahl*. Breslau: Wilhelm Koebner. Translation in Frege (1953).

Frege, Gottlob. 1953. *Foundations of Arithmetic*, ed. J. L. Austin. Oxford: Basil Blackwell & Mott, 2nd ed.

Potter, Michael. 2004. *Set Theory and its Philosophy*. Oxford: Oxford University Press.

Smullyan, Raymond M. 1968. *First-Order Logic*. New York, NY: Springer. Corrected reprint, New York, NY: Dover, 1995.

Zuckerman, Martin M. 1973. Formation sequences for propositional formulas. *Notre Dame Journal of Formal Logic* 14(1): 134–138.