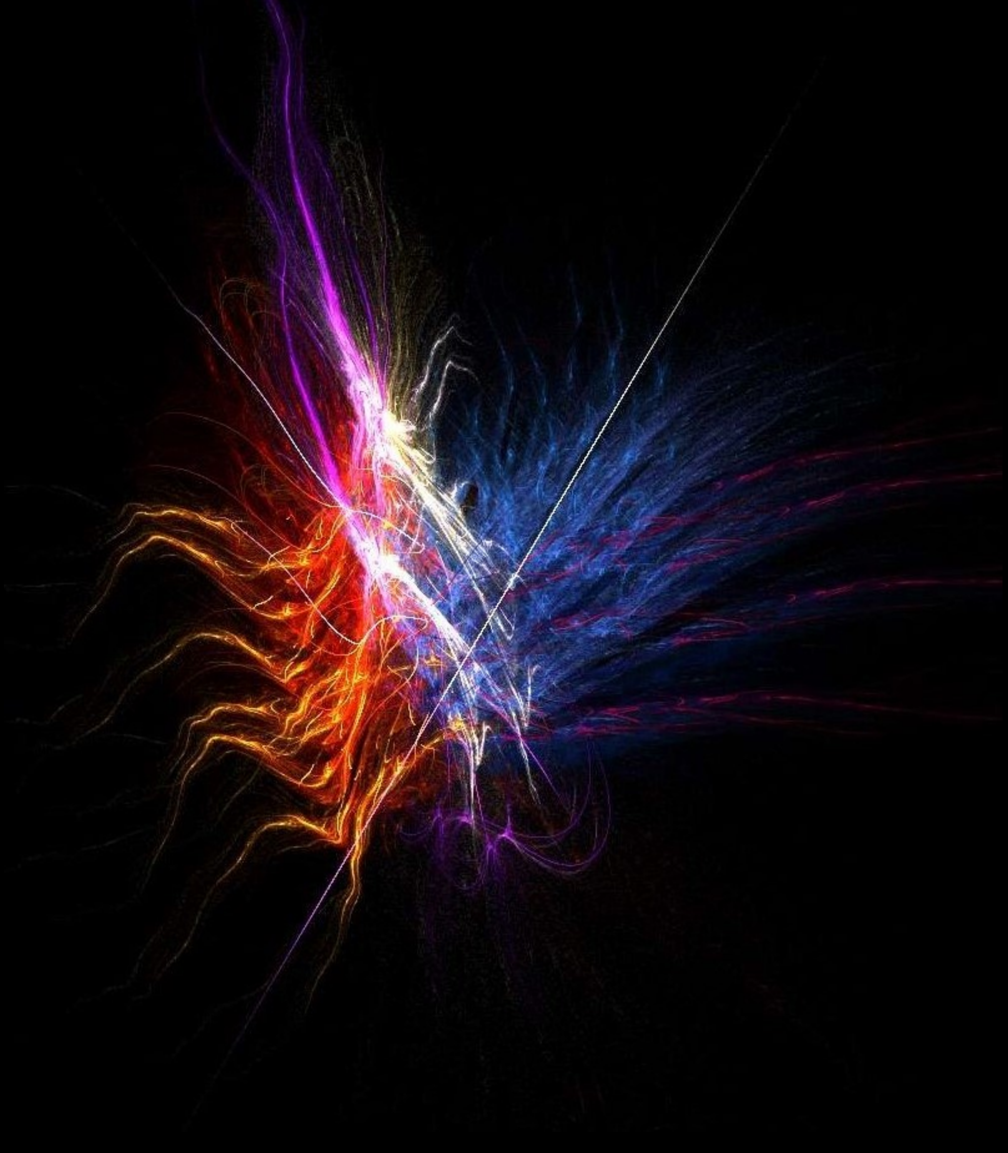
# A Journey Through Formal Mathematics



Pace P. Nielsen

Pace P. Nielsen
Department of Mathematics
Brigham Young University
Provo, UT 84602
pace@math.byu.edu

Version: 1.04
Date: January 17, 2023

# Contents

# Preface

This book is intended as something of a sequel to the book "A Transition to Advanced Mathematics" by Darrin Doud and the author, which will be referred to throughout this book as "Transition". That book acts as an introduction to the language and some of the main ideas of modern mathematics. This book offers a deeper foray into the material covered in "Transition", with additional exercises, topics, and material, suitable for a second course on set theory, logic, and other related advanced topics.

Many sections of this book are intentionally parallel to corresponding sections of "Transition", and they discuss issues that were previously glossed over, or are in need of additional clarification and expansion. However, this book is not meant to be read concurrently with "Transition" as we assume that the reader is familiar with many of the notations, concepts, and definitions from that book, such as functions, sequences, and so forth.

The exercises in this book are key to developing understanding. We will assume that, at the very least, the reader will have read each exercise and understood its content. It is hoped that this book would be suitable for a flipped classroom setting, where students read the sections of the book and come to class prepared with solutions to the exercises.

The material in this book was derived in large part from the following books:

(1) "Set Theory" by Thomas Jech,
(2) "Logic, Induction and Sets" by Thomas Forster,
(3) "Introduction to Mathematical Logic" by Elliot Mendelson, and
(4) "An Invitation to General Algebra and Universal Constructions" by George M. Bergman.

Students interested in an even deeper journey through advanced mathematics may find those texts useful.

There are a few differences between the notation in "Transition" and the notation here. We will explain these differences as they arise. A *countable* set here is always infinite. We will rarely have need to combine countability and finiteness in this text. Uncountable sets are still those sets that are not finite nor countably infinite. Also, in this book the natural numbers will always contain 0.

Pace P. Nielsen

# Chapter I

# Formal Set Theory

*It is a miracle that curiosity survives formal education.* Albert Einstein

One thinks of sets, from an informal standpoint, as unordered collections of objects, where we disregard repetitions. The following two sections describe some limitations and further refinements of this idea, using an axiomatic approach. Particular attention is paid to the paradox of Russell. We will cover the modern methods of avoiding this and other related paradoxes.

From an axiomatic and formal point of view, sets are only what one is allowed to do with them. By changing the axioms one changes the very concept of set, even if the informal idea of "collection" is still a good starting place. Many of the facts about sets that one takes for granted in "Transition" are now treated formally as axioms or as simple consequences of them. The idea of a "class" is also introduced.

In this chapter, whenever the formal version of an axiom is introduced, the reader is encouraged to check that it says what they think it should say. The reader is also encouraged to answer questions posed in the text, as this process can often help with understanding. Moreover, it might also be helpful to try exercises once you have learned the relevant material; exercises are ordered to make this process easier.

# 1 Sets, notations, and axioms, oh my!

## 1.A What is a set?

As a first approximate definition for “set” we might say: “A *set* is a collection of objects”. When pressed to give an example that occurs in day-to-day practice, we might find it difficult to do so. For instance, are music playlists sets? Playlists can contain repetitions of songs, and those repetitions definitely change the nature of the playlist, but repetitions are disregarded in sets. Also, the order of the songs in a playlist matters, whereas the order of the elements in a set does not.

Moreover, to even begin talking about sets, we first need to explain what a “collection” is, what “objects” are, and how objects belong to a collection. This starts getting us into a vicious circle, with no place to get a footing. To handle these issues, the membership relation $\in$ is taken as a *primitive notion* that we do not define. We simply assert its existence, and then add axioms to explain its properties. The hope is that this primitive notion is so simple that it can be used without much explanation. Any formal mathematical system begins this way; e.g., if you are familiar with Euclid’s geometry then you know that the notions of “point” and “line” are (usually) taken as primitive notions (among others).

In formal set theory the membership relation and sets themselves are both primitive notions. We describe their meaning using axioms. So, the assertion that sets really only depend on their elements, ignoring repetition and order, should be taken as an axiom. This axiom is called the *axiom of extensionality*, and it asserts that two sets $S$ and $T$ are equal exactly when their elements are the same. In symbols

$$(S = T) \leftrightarrow \forall x \, (x \in S \leftrightarrow x \in T). \tag{1.1}$$

There are multiple ways to think about this axiom. First, this axiom could be viewed as defining the equality symbol in terms of the membership relation $\in$. This way of thinking appeals to minimalists, who want to reduce all of mathematics to statements in formal set theory subject to only one relation symbol $\in$. The second, more common, way of thinking about this axiom is that equality is already understood (as asserting that $S$ and $T$ are literally the same thing), and the axiom tells us that $S$ and $T$ are literally the same set if and only if they have the same elements.

Thus, when a mathematician is asked to prove that two sets $S$ and $T$ are equal, and they proceed to show that every element $x$ of $S$ belongs to $T$, and vice versa, they are implicitly using this axiom. It is accepted (nearly) universally by mathematicians as a necessary, definitional truth about sets. Sets are uniquely determined as *extensions* of their elements.

## 1.B Do sets exist?

Returning to our previous example of playlists, we might be forced to admit that any playlist itself is not a set, since two playlists with the same songs are not necessarily equal. However, we might assert that there is some abstract set $S$ consisting exactly of the songs from our playlist. What is it that allows us to make this claim? Can we guarantee that a set exists at all?

Without any further axioms, the answer to that last question is no. There need not be any sets at all! To overcome this difficulty we make a very tame assertion, called the *axiom of empty set*. In symbols

$$\exists S\, \forall x\, \neg(x \in S). \tag{1.2}$$

Any set $S$ satisfying this definition is unique by the extensionality axiom. (You might try Exercise 1.3 now). This unique set is named by the special symbol $\emptyset$ and is called the *empty set*. Note that the empty set is a *derived notion*, not a primitive one.

We now have access to one set. Are there any others? Again, without any further axioms the answer is no. Rather than assert the existence of more and more sets, it is more convenient to give general methods to create new sets from previously constructed sets.

## 1.C Pairing

Our next axiom is the *axiom of pairing*, which asserts that if $x$ and $y$ are any two objects we have access to, then their unordered pair exists. In symbols, we write

$$\forall x\, \forall y\, \exists S\, \forall z\, (z \in S \leftrightarrow (z = x \vee z = y)). \tag{1.3}$$

**Proposition 1.4.** *Given $x$ and $y$, the set $S$ whose existence is guaranteed by the axiom of pairing is uniquely determined by $x$ and $y$.*

*Proof.* Let $S_1$ and $S_2$ be any two sets satisfying the axiom of pairing. We wish to show that $S_1 = S_2$. By the axiom of extensionality, it suffices to show that they have the same elements.

Let $z \in S_1$. Since $S_1$ satisfies the axiom of pairing (1.3) we know that $z = x$ or $z = y$. In either case, since $S_2$ also satisfies the axiom of pairing we have $z \in S_2$. By a similar argument, any element of $S_2$ lies in $S_1$. Thus $S_1 = S_2$. □

The unique set whose existence is being asserted by (1.3) is written as $\{x, y\}$.

**Warning 1.5.** The set $\{x, y\}$ does not always have two elements. Consider the case when $x = y$; in that case we write $\{x, x\} = \{x\}$.

**Example 1.6.** If we take $x = y = \emptyset$, then the axiom of pairing gives us the new set $\{\emptyset, \emptyset\} = \{\emptyset\}$, the "set of the empty set". Next, taking $x = y = \{\emptyset\}$, then the axiom of pairing gives us the new set $\{\{\emptyset\}\}$. Repeating this process, we see that we now have access to an infinite number of sets

$$\emptyset,\ \{\emptyset\},\ \{\{\emptyset\}\}, \ldots.$$

Moreover, the axiom of pairing implies that given any set $x$ we can always form the *singleton* $\{x\}$; the unique set whose only element is $x$. △

**Example 1.7.** Starting with just the empty set $\emptyset$, the *only* choice for both $x$ and $y$ in the pairing axiom is $\emptyset$. Initially, the only set that pairing gives us is $\{\emptyset\}$.

Now, with these two sets in hand, there are four choices for $x$ and $y$, namely

$$\begin{array}{ll} x=\emptyset,\ y=\emptyset, & x=\emptyset,\ y=\{\emptyset\}, \\ x=\{\emptyset\},\ y=\emptyset, & x=\{\emptyset\},\ y=\{\emptyset\}. \end{array}$$

We get only two new sets $\{\emptyset,\{\emptyset\}\}$ and $\{\{\emptyset\}\}$.

How many new sets do we get if we repeat this process one more time? △

A natural question is whether we need to assert another axiom to claim the existence of *ordered pairs*, which we will write as $(x, y)$. Surprisingly, the answer is no, due to a clever trick of Kuratowski that allows us to encode ordered pairs using unordered pairs, as follows. Define the ordered pair of two objects $x$ and $y$ as the set

$$(x,y) := \{\{x\},\{x,y\}\}. \tag{1.8}$$

Exercise 1.5 has you show that $(x, y) = (x', y')$ if and only if $x = x'$ and $y = y'$, for any objects $x$, $y$, $x'$, and $y'$, so the encoding is injective. Further, one can tell just from the elements of a set whether or not it is an encoded ordered pair.

**Advice 1.9.** Many mathematicians take "ordered pair" as a primitive notion. Kuratowski's trick allows us to develop set theory without that new primitive notion, at the cost of thinking about ordered pairs in a somewhat unusual way, similar to how computers encode language using binary digits.

Now that we have this trick, one can (as needed) forget about it and treat ordered pairs as primitive after all. In other words, we forget the *syntax* of (1.8) and focus on the *semantics* or *meaning* of the symbols afforded by Exercise 1.5.

**Example 1.10.** If we have access to the numbers 1 and 3, then we can encode the ordered pair $(3, 1)$ as the set $\{\{3\},\{3,1\}\}$. Notice that this is different than the ordered pair $(1, 3)$, which is encoded as the set $\{\{1\},\{1,3\}\}$. △

**Example 1.11.** There are variants of Kuratowski's encoding. The first encoding of ordered pairs as sets was found in 1914 by Wiener, who used the encoding

$$(x,y) := \{\{\{x\},\emptyset\},\{\{y\}\}\}.$$

We use Kuratowski's encoding because it is easier to work with (among other reasons).

Not every scheme gives a valid encoding. For instance, if we take

$$(x,y) := \{x,\{y\}\},$$

this will not correctly capture the meaning of ordered pairs. Can you see why? △

## 1.D Exercises

**Exercise 1.1.** For two sets $x$ and $y$, show that $x \notin y$ is a derived notion, by defining it using a formula in terms of the primitive notion $\in$. (The symbols $x$ and $y$ should appear in the formula; they will act as *parameters*. The symbol $\notin$ should not appear in the formula.)

**Exercise 1.2.** For two sets $x$ and $y$, show that $x \subseteq y$ is a derived notion, by defining it using a formula in terms of the primitive notion $\in$.

**Exercise 1.3.** Assuming the axioms of empty set and extensionality, prove that there is exactly one empty set.

**Exercise 1.4.** Answer the question at the end of Example 1.7, by finding all of the new sets created by the pairing axiom (and the axiom of empty set) on the third iteration. Now do the same for the fourth iteration.

**Exercise 1.5.** For any objects $x$, $y$, $x'$, $y'$, prove that $(x, y) = (x', y')$ if and only if $x = x'$ and $y = y'$.

**Exercise 1.6.** Show that the previous exercise fails if we were to redefine an ordered pair as $(x, y) := \{x, \{y\}\}$.

**Exercise 1.7.** Kuratowski's definition of an ordered pair needs three applications of the pairing axiom. Find an alternative definition of ordered pair that uses only two instances of the pairing axiom and no other axiom. (Your definition should still satisfy Exercise 1.5, but you don't need to provide the proof of that fact.)

**Exercise 1.8.** State an axiom of triplets, in symbols. (We will see in the next section that we will not need such an axiom, but it is valuable practice to translate a concrete idea into a formula.)

The next exercise partially develops the method of encoding the natural numbers as sets, as proposed by Zermelo and von Neumann. Throughout, you may assume the existence of natural numbers as well as basic facts about them, such as induction.

**Exercise 1.9.** If we think of each natural number as encoded by "the set of the previously encoded natural numbers", then we should take

$$\underline{0} := \{\,\} = \emptyset, \qquad \underline{1} := \{\underline{0}\} = \{\emptyset\}, \qquad \underline{2} := \{\underline{0}, \underline{1}\} = \{\emptyset, \{\emptyset\}\}, \qquad \text{and so forth.}$$

The set $\underline{3}$ exists using the axiom of triplets. Write out its elements explicitly (using only braces, commas, and the symbol $\emptyset$). Similarly, write out $\underline{4}$ and $\underline{5}$. (Note how many elements belong to each of these sets.) Construct a set that is not an encoded natural number, and prove that it is not an encoded number (perhaps by using the fact that 0 is smaller than all other natural numbers).

# 2 Axiomatizing the obvious is not so obvious

In the previous section we introduced three axioms: extensionality, empty set, and pairing. Extensionality helps us understand when sets are equal, and the other two axioms aid us in forming very small sets. In this section we add more axioms to our list, which will help us formalize set-builder notation, and thus allow for the formation of many more sets.

## 2.A Sets with more than two elements?

With the axioms of extensionality, empty set, and pairing, we only have access to sets with two elements at most. Rather than assert axioms of triplets, quadruplets, and so forth, we instead assert a general principle that allows us to put together a lot of sets all at once, the *axiom of union*.

Before we state this axiom formally, it is important to point out that the union symbol is used in two somewhat different ways in formal set theory. First, given two sets $A$ and $B$, if we write $A \cup B$, then we mean the collection whose elements are exactly those of $A$ together with those of $B$; i.e., the set $\{x : x \in A \text{ or } x \in B\}$. This notation is the standard one used across multiple disciplines. More generally, if $S$ is a collection of sets, we might write $\bigcup_{A\in S} A$ for the union of all the sets in $S$; i.e., the collection $\{x : \text{there is some } A \in S \text{ such that } x \in A\}$. In other words, this is the set of elements that belong to at least one of the sets in $S$. This notation is also standard in most mathematical disciplines.

The second notation, which is standard in formal set theory, but not as common in other disciplines, and usually not taught to beginning students, is $\bigcup S$. This notation has the same meaning as $\bigcup_{A\in S} A$, which may be slightly confusing because different letters occur to the right of the union symbol. However, of the three ways of writing this set,

$$\bigcup S = \bigcup_{A\in S} A = \{x : \text{there is some } A \in S \text{ such that } x \in A\}, \tag{2.1}$$

the first is the most economical.

**Warning 2.2.** Given a set $S$, when we write $\bigcup S$, using a large union symbol with no subscripts on the union symbol, then this will always refer to the economical notation given in (2.1).

The axiom of union asserts that $T = \bigcup S$ exists as a set, for any set $S$. In symbols

$$\forall S\, \exists T\, \forall x\, (x \in T \leftrightarrow \exists A\, (A \in S \wedge x \in A)). \tag{2.3}$$

By the axiom of extensionality the union, $\bigcup S$, is uniquely determined by $S$.

**Example 2.4.** Given any two sets $A$ and $B$, then the axiom of pairing gives us the set $S = \{A, B\}$. Now $\bigcup S$ in this case is just $A \cup B$. So our axiom of union allows us to define the pairwise union of any two sets. With that given, we can define the union of three sets as $A \cup B \cup C = (A \cup B) \cup C$, and so forth. △

**Example 2.5.** Given any three (not necessarily distinct) objects $a, b, c$, we will show that the triplet set $\{a, b, c\}$ exists. By pairing $\{a, b\}$ exists, and by another application of pairing $\{b, c\}$ exists. By a third application of the pairing axiom, we have the set $S = \{\{a, b\}, \{b, c\}\}$. Now, the axiom of union gives us the set $\bigcup S = \{a, b, c\}$, as desired. Given objects $a, b, c, d$, how many applications of pairing and union are needed to form the quadruplet set $\{a, b, c, d\}$? △

## 2.B Set-builder notation

In the previous subsection we used *set-builder notation* as one way to describe unions. It seems natural that we would, more generally, allow arbitrary sets created using this notation. The *comprehension axioms* assert: Given any property $P(x)$ of objects $x$, the collection $\{x : P(x)\}$ is a set. For instance, the property given by the formula

$$\exists A\,(A \in S \wedge x \in A)$$

is the one we used in (2.1) to describe $\bigcup S$, where the symbol $S$ acts as a parameter. (We will define properties and parameters more formally in Chapter II.)

There is a problem with the comprehension axioms—they are inconsistent!

**Theorem 2.6** (Russell's Paradox)**.** *The comprehension axioms lead to a contradiction.*

*Proof.* Consider the collection $T = \{x : P(x)\}$ where $P(x)$ is the property $x \notin x$. Thus, $T$ is the collection of all the sets $x$ that do not contain themselves as elements. (These are the kinds of sets one actually expects to work with and that have appeared previously in the book.) By comprehension, $T$ would be a set. Thus we can ask whether $T \in T$ is true or not.

**Case 1**: Assume $T \in T$. Then $T$ must satisfy the defining property of being a member of $T$, hence $P(T)$ holds. In other words $T \notin T$, a contradiction.

**Case 2**: Assume $T \notin T$. As $P(T)$ holds, then $T \in T$, a contradiction. □

When set theory was first being developed, the axioms of comprehension were used freely, until Russell discovered this paradoxical fact. Thus, the comprehension axioms are nowadays *not* accepted as valid axioms. On reflection, you might have previously used set-builder notation freely, without realizing the possibility that your notation might be prone to error. On the other hand, set-builder notation is extremely useful, powerful, and flexible. We've already used it to describe unions, for instance. It turns out that much of set-builder notation can be salvaged, in one way or another, using three main ideas.

**Idea 1.** *We should allow comprehension* inside *of an already existing set.*

In other words, we allow set-builder notation of the form

$$T = \{x \in A : P(x)\}. \tag{2.7}$$

These restricted versions of the comprehension axioms are called the *separation axioms*. For each property $P(x)$, which may have $A$ and $S$ act as parameters, but not $T$, the following statement is an instance of a separation axiom:

$$\forall A\, \forall S\, \exists T\, \forall x\, (x \in T \leftrightarrow (x \in A \wedge P(x))). \tag{2.8}$$

You might check that this is a fancy way of asserting the same thing as (2.7).

**Example 2.9.** For a moment, assume that the set $\mathbb{N}$ of natural numbers exists. We can write the collection of squared natural numbers $\{0, 1, 4, 9, \ldots\}$ using set-builder notation as

$$\{x \,:\, x \text{ is a squared natural number}\}. \tag{2.10}$$

Using that notation we cannot, *a priori*, guarantee that this collection is a set, because we do not assume the comprehension axioms. However, the modified set-builder notation

$$\{x \in \mathbb{N} \,:\, x \text{ is a squared natural number}\} \tag{2.11}$$

is automatically a set by a separation axiom. As (2.10) and (2.11) have the same elements, we now know, *a fortiori*, that the collection in (2.10) is a set. △

**Advice 2.12.** When possible, restrict the elements that appear in set-builder notation (before the colon) to be members of a set you already know exists.

**Example 2.13.** Let $A$ and $B$ be sets. Their *intersection*, written $A \cap B$, should be

$$\{x \,:\, x \in A \text{ and } x \in B\}.$$

We do not know (yet) if this describes a set. However, this collection is the same as

$$\{x \in A \,:\, x \in B\},$$

which we now know is a set by separation. (This is one reason we don't need an axiom of intersection.) △

The name "separation" comes from the fact that the separation axioms allow us to separate out the elements $x \in A$ satisfying property $P$, discarding the elements that do not satisfy property $P$. The reason that—unlike for the axiom of extensionality, the axiom of empty set, the axiom of pairing, or the axiom of union—we pluralize the *axioms* of separation is because there is a different axiom for each property $P$. (It is an interesting question, with an even more interesting answer, whether or not a finite number of separation axioms can cover all cases. We will address this issue later in the book.)

**Idea 2.** *Applying functions to already existing sets should output sets.*

**Example 2.14.** The set of squares of natural numbers was written previously in the form (2.11). Alternatively, we could write

$$\{x^2 : x \in \mathbb{N}\}. \tag{2.15}$$

Or, in other words, if $f\colon \mathbb{N} \to \mathbb{N}$ is the function given by the rule $f(x) = x^2$, then the set we are looking at is $\{f(x) : x \in \mathbb{N}\}$. △

This new version of set-builder notation is authorized by the *replacement axioms*, which assert that applying any function to a set *replaces* it with a set. Thus (2.15) will automatically be a set. We will define functions in Section 4, but there are additional issues related to the replacement axioms that need explanation, so we will formally define these axioms in a later chapter.

**Idea 3.** *Unrestricted comprehension gives us collections that are just* too big *to be sets, but they* are *collections of some kind.*

The standard terminology is to call arbitrary collections by the name *classes*, and call those collections that are too big to be sets (on pain of contradiction) *proper classes*. Thus, the collection described in Russell's paradox, namely $\{x : x \notin x\}$, is a proper class.

**Example 2.16.** The collection of all sets, written $V = \{x : x \text{ is a set}\}$, is a proper class. We leave the proof as Exercise 2.8. △

The previous example acts as leverage in proving other classes are proper classes.

**Proposition 2.17.** *The class of all singletons, $W = \{\{x\} : x \text{ is a set}\}$, is a proper class.*

*Proof.* If it were a set, then $V = \bigcup W$ would be a set, by the axiom of union. But $V$ is not a set, by Exercise 2.8. □

**Warning 2.18.** Sets are classes, just not proper classes.

This begs the question: Can we create "the class of all classes that don't contain themselves" and thus run into Russell's paradox? To prevent this problem we make a general rule that we always follow: Proper classes can **never** be elements of classes.

In formal set theory we go even further by asserting that elements of classes are **always** sets. In other words, the only objects available to be put in a collection are sets; no other objects appear on the left of the $\in$-relation. Thus, in formal set theory there is no such thing as a set of songs from a playlist, or even a set of natural numbers (if "number" is taken as a primitive notion). However, just as Kuratowski's trick allows us to encode ordered pairs as sets, we will see that numbers, playlists, and other objects can also be encoded.

## 2.C Exercises

**Exercise 2.1.** The axiom of union asserts that $\bigcup S$ exists for any set $S$. What is the set $\bigcup \emptyset$? What is $\bigcup\{\emptyset\}$? What is $\bigcup\{\emptyset, \{\emptyset\}\}$?

**Exercise 2.2.** Encode the number 0 as the set $\underline{0} := \emptyset$. Also, assuming that $n \in \mathbb{N}$ has been encoded as the set $\underline{n}$, then encode $n + 1$ as the set $\underline{n+1} := \underline{n} \cup \{\underline{n}\}$.

Prove that every natural number has an encoding as a set. (Hint: Use induction on the natural numbers. The proof may seem easy.) This exercise shows that the idea introduced in Exercise 1.9 works for all natural numbers, once we assume the axiom of union. (Is the encoding injective?)

**Exercise 2.3.** Using the encoding of the previous exercise, is $\bigcup \underline{n}$ an encoded natural number? If so, state which number it is (without proof). If not, give a counterexample.

**Exercise 2.4.** Given two sets $A$ and $B$, we defined $A \cup B$ as $\bigcup\{A, B\}$, which requires one application of the pairing axiom followed by a single use of the union axiom. Similarly, given three sets $A, B, C$, we can define $A \cup B \cup C$ as the set $(A \cup B) \cup C$, or in other words

$$\bigcup\left\{\bigcup\{A, B\}, C\right\}.$$

This set is created by using four axioms: pairing, union, pairing, and union, in that order. Thus, notice that the proof that $A \cup B \cup C$ exists is longer than the proof that $A \cup B$ exists.

Given the four sets $A, B, C, D$, define $A \cup B \cup C \cup D$. Also describe which axioms (counting repetitions) are needed to create this set.

If you are given ten sets, how many axioms (counting repetitions) are needed to describe their union?

**Exercise 2.5.** Given three sets $A, B, C$, use the axiom of extensionality to prove the associative property $(A \cup B) \cup C = A \cup (B \cup C)$. (Hint: For the forward inclusion, start by assuming $x \in (A \cup B) \cup C$, and show that $x \in A \cup (B \cup C)$.)

**Exercise 2.6.** For any four objects $a, b, c, d$, show that the quadruplet set $\{a, b, c, d\}$ exists, by using only three applications of the pairing axiom and one application of the union axiom. What is the minimum number of applications of the pairing and union axioms needed to form quintuplet sets? Do we need any applications of pairing if we assume a quadruplet axiom?

**Exercise 2.7.** Show that the *difference* of two sets $A$ and $B$, given by the class $A - B = \{x : x \in A \text{ and } x \notin B\}$, is a set.

**Exercise 2.8.** Prove that the class consisting of all sets, denoted $V$, is a proper class. (Hint: If $V$ is a set, consider $T = \{x \in V : x \notin x\}$.)

**Exercise 2.9.** Prove that if $C$ is a nonempty class, then $\bigcap C = \bigcap_{A \in C} A$ is a set. (Very small hint: Fix some set $A_0 \in C$.) Some mathematicians take $\bigcap \emptyset = V$, which is not a set.

**Exercise 2.10.** Assuming just the axioms of extensionality and union, does the existence of a nonempty set $S$ such that $S = \bigcup S$ lead to a contradiction?

# 3 Infinity

In the previous section we introduced the axioms of extensionality, empty set, pairing, and union, as well as the separation axioms, as the first steps of describing formal set theory. We will continue to accept these axioms in this section and throughout the remainder of the book as appropriate. Historically speaking, none of these axioms has elicited much angst or disbelief.

In this section and the next we will introduce two new axioms—infinity and power set—both of which have a richly historied past and have involved some controversy. These axioms will open up a host of other standard constructions. For instance, we will be able to encode both relations and functions as sets of ordered pairs.

## 3.A Infinity axiomatized

The sets we constructed in the previous section were all finite. It should seem intuitively obvious (even if currently difficult to formally prove) that we cannot obtain any infinite set using only the axioms of the previous sections. We thus need a new axiom to assert the existence of some infinite set $S$. The idea of the axiom is that when $S$ has an element $x$, then it will also have another (supposedly bigger) element $x \cup \{x\}$. The formal statement of the *axiom of infinity* is:

$$\exists S\,(\emptyset \in S \wedge \forall x\,(x \in S \to x \cup \{x\} \in S)). \tag{3.1}$$

Any set $S$ satisfying the condition after the first quantifier in (3.1) is called an *inductive set*, because its structure is parallel to an induction argument (the empty set corresponds to the "base case" and from any set $x$ we can reach the next stage $x \cup \{x\}$ just like in an "inductive step"). The axiom of extensionality does *not* imply that a set satisfying the axiom of infinity is unique, as the axiom does not tell us precisely what elements belong to an inductive set $S$. Indeed, we'll see later there are many distinct inductive sets. However, the following theorem does distinguish one very important inductive set.

**Theorem 3.2.** *There exists a unique smallest inductive set, in the sense that it is a subset of every other inductive set.*

*Proof.* Let $C$ be the class of all inductive sets. It is nonempty by the axiom of infinity. By Exercise 2.9, $\bigcap C$ is a set. Clearly, $\bigcap C$ is a subset of every inductive set, so it suffices to prove that $\bigcap C$ is itself inductive.

First, given any inductive set $S \in C$, we have $\emptyset \in S$. Thus $\emptyset \in \bigcap C$.

Next, assume that $x \in \bigcap C$. Given any $S \in C$ we have $x \in S$. Since $S$ is inductive, we have $x \cup \{x\} \in S$. Since $S \in C$ was arbitrary, we have $x \cup \{x\} \in \bigcap C$. This concludes the proof that $\bigcap C$ is inductive. □

**Definition 3.3.** We call the unique intersection of all inductive sets the *formal set of natural numbers*, written as $\omega$ (for reasons to be explained later).

To fully justify the name "natural numbers" used in this definition, we should prove that the elements of $\omega$ behave like the natural numbers. In particular, we should show that the encoded natural numbers of Exercise 2.2 belong to $\omega$. To that end, we introduce one more definition that is very useful.

**Definition 3.4.** A set (or even a class) $S$ is called *transitive* if every element is a subset. In other words, if $y \in S$ and $x \in y$, then $x \in S$.

**Example 3.5.** The set $\underline{0} = \emptyset$ is transitive, vacuously.

The set $\underline{1} = \{\emptyset\}$ is transitive. Its one element, $\emptyset$, is a subset.

One can check directly that $\underline{2} = \{\emptyset, \{\emptyset\}\}$ is transitive, as is $\underline{3}$.

The set $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$ is not an element of $\omega$, but it is transitive.

The set $\{\{\emptyset\}\}$ is not transitive, since its one element is not a subset. △

The next theorem will help us show that $\omega$ is transitive. It also gives a nice example of how some concepts can be self-improving.

**Theorem 3.6.** *If $S$ is any inductive set (or even a class), then $\{x \in S : x \subseteq S\}$ is also inductive.*

*Proof.* Assume $S$ is inductive, and let $T := \{x \in S : x \subseteq S\}$. First $\emptyset \in T$ since $\emptyset \in S$ and $\emptyset \subseteq S$.

Next, suppose that $x \in T$. Then $x \in S$ and $x \subseteq S$. As $S$ is inductive, $x \cup \{x\} \in S$. Further, we see that $x \cup \{x\} \subseteq S$, since all of the elements of $x$ are members of $S$ and so is $x$ itself. Thus $x \cup \{x\} \in T$. □

The previous theorem says, in effect, that inside any inductive set is another inductive set that is closer to being transitive. This yields:

**Corollary 3.7.** *The set $\omega$ is transitive.*

*Proof.* The set $\{x \in \omega : x \subseteq \omega\}$ is inductive by Theorem 3.6. It is a subset of $\omega$, which is the smallest inductive set. Thus, it equals $\omega$. In particular, each element $x \in \omega$ is a subset of $\omega$, so $\omega$ is transitive. □

In the exercises you will show, moreover, that each element of $\omega$ is a transitive set and equals the set of all "smaller" formal natural numbers. Thus, the idea expressed in Exercises 1.9 and 2.2, of thinking about each natural number as the collection of the previous natural numbers, is now fully realized.

Following the notation of that exercise, given any set $x$ we sometimes write $x + 1$ to mean the set $x \cup \{x\}$.

## 3.B Overused names

There is a technical subtlety that should be pointed out. There are now two different collections of numbers that go by the same name of “set of natural numbers”. The numbers in the first collection are the numbers we personally use to count; these are the numbers being encoded in formal set theory. These cerebral numbers, the collection of which is written as $\mathbb{N}$, are often called the *meta natural numbers*, because they occur outside and prior to every formal structure. The second collection is the (formal) set $\omega$ named in Definition 3.3. Some people refer to the meta natural numbers as the “true” or “real” natural numbers. The elements of $\omega$ are the “formal” natural numbers, which include the “encoded” meta natural numbers.

**Example 3.8.** If someone says they own two cats, they mean the meta number 2, and not the set $\underline{2} = \{\emptyset, \{\emptyset\}\}$. △

In some situations this is a distinction without a difference. However, in other situations this distinction is very important.

The most important difference between these two types of numbers is that we cannot talk about meta natural numbers inside formal set theory. The encoding only goes one way. There is no way to escape from the formal system back into the “real” world, so to speak. This is a cost we must pay when working with formal theories—they are no longer informal. Formal theories are only approximate likenesses of the informal meta concepts we take for granted.

Doesn’t Exercise 2.2 have us at least prove that every meta natural number has an encoding as an element of $\omega$? Yes, but this proof does not take place inside formal set theory. We have only given an informal argument that there is a map (which is injective) from our meta numbers $\mathbb{N}$ to $\omega$. The informal argument presupposes the existence of those meta numbers and basic facts about them. This is what is called a *metatheorem*. It is a theorem outside the formal system. After we develop the language of formal set theory in a later chapter we will see that there is a fundamental roadblock in talking about the concept of meta natural numbers inside the formal system.

Even worse, it is entirely possible that the metatheoretical map we constructed

$$\mathbb{N} \to \omega$$

where

$$0 \mapsto \underline{0} = \emptyset,\ 1 \mapsto \underline{1} = \{\emptyset\},\ 2 \mapsto \underline{2} = \{\emptyset, \{\emptyset\}\},\ \ldots,$$

is not surjective. There may be elements of $\omega$, the formal natural numbers, that are not encoded meta natural numbers. One might wish to assert, as an additional axiom of the formal system, that the meta natural numbers also surject onto $\omega$. However, meta numbers are not formal objects, so turning such an assertion into a formal statement is very difficult, if not impossible, depending on your formalization.

We will see that the meta numbers are needed to develop the formal language of set theory, so we do need to talk about them at times, but only and always *outside* the formal system, in the nonformalized *metatheory*.

**Example 3.9.** Here is another concrete example of the difference between meta natural numbers and elements of $\omega$. In Example 2.4 we defined the union of two sets, $A$ and $B$, using the notation $A \cup B$. This is talking about the meta number two. For $n \geq 2$, we can define repeated unions recursively by the formula

$$A_1 \cup A_2 \cup \cdots \cup A_n \cup A_{n+1} = (A_1 \cup A_2 \cup \cdots \cup A_n) \cup A_{n+1}.$$

The symbol $n$ used here is a meta natural number, not an element of the set $\omega$. It refers to repeating a process in the metatheory some meta number of times. △

**Example 3.10.** The claim in Example 1.6 that we have access to an infinite number of sets $\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots$ is a metatheorem, about the meta process of repeatedly adjoining an additional pair of set braces. On the other hand, if we pick out any one of the sets in this list, such as

$$S = \{\{\{\{\{\{\{\{\{\{\{\emptyset\}\}\}\}\}\}\}\}\}\}\}, \tag{3.11}$$

it is a theorem (not just a metatheorem) that $S$ exists in the formal system. The proof that $S$ exists is formal, not informal; it uses only the axioms we have asserted. In point of fact, the proof of our metatheorem gives us an algorithm to create each such formal proof. For instance, our proof tells us that by applying the pairing axiom eleven times, to the successive pairs of elements

$$x_1 = y_1 = \emptyset,\ x_2 = y_2 = \{\emptyset\}, \ldots, x_{11} = y_{11} = \{\{\{\{\{\{\{\{\{\{\emptyset\}\}\}\}\}\}\}\}\}\},$$

we obtain the set $S$ in (3.11). (The ellipses in the line above are used only to save space, and can easily be filled in.) △

The metatheorem

$$\text{“For each meta natural number } n \text{, there is a set } \underbrace{\{\{\ldots\{\emptyset\}\ldots\}\}}_{n \text{ times}}.\text{”}$$

can be thought of as a (meta) list of formal theorems, one for each meta number $n$, and so we may call it a *theorem schema*. The proof of this metatheorem is an informal description of how to construct a formal proof of each of the corresponding formal theorems. We can explicitly construct the proof of each successive formal theorem, such as we did in the case $n = 11$ in Example 3.10, but these proofs become longer as $n$ increases since they require using the pairing axiom more often.

Note that each axiom of a formal system is also a theorem in that system (with a one-line proof!). Thus, the axioms of separation also give a theorem schema. These axioms are called an *axiom schema* because of this fact. Similarly, the replacement axioms are an axiom schema.

We will take pains, from this point on, when working in formal set theory, to explicitly mention when a symbol $n$ means a meta natural number, and when a theorem is really a theorem schema or a metatheorem. Many authors do not bother with such details, but that is a disservice to their readers.

Just as many mathematical theories can be formalized, so can metatheories. This creates an additional level of ambiguity, because the natural numbers in the formalized metatheory are sometimes still called the "true" natural numbers, even though there is now another level—what one might call the metatheory of the metatheory—where the "truly true" natural numbers live. On the flip side, if you are doing normal mathematics without developing any formalized system, then you never have to worry about meta natural numbers at all.

**Advice 3.12.** When studying any formal theory, pay attention to indices and other repeated processes. The author may be talking about a meta process without pointing it out to the reader.

If an author asserts that the natural numbers in a formal system are the "real" ones, this probably is meant in the context of a larger formal system.

## 3.C Defining infinity

To end this section, we will define what it means to say that some set is infinite. If you ask a random mathematician what it means for a set $S$ to be infinite you are likely (but not guaranteed) to get one of two distinct answers.

(1) There is no bijection between $S$ and $\{1, 2, \dots, n\}$, where $n \in \mathbb{N}$.

(2) The set $S$ has a proper subset $A \subsetneq S$ where there is a bijection from $A$ to $S$.

(If the mathematician instead says "it is not finite" then ask them to formally define "finite" and check that the negation of their definition is one of the options above.)

We will follow the first convention, which is incidentally the most popular, and make the following definitions.

**Definition 3.13.** A set $S$ is *finite* if there is a bijection between $S$ and some (formal) natural number $n \in \omega$.

A set $S$ is *infinite* if it is not finite.

Bijections will be formally defined in the next section. By the way, the sets in option (2) are called the *Dedekind-infinite* sets because of Dedekind's work with them. Without the axiom of choice, which will be handled in a later section, there might exist infinite sets that are not Dedekind-infinite. Thus, the behavior of infinite sets can depend quite a lot on your axioms.

## 3.D Exercises

**Exercise 3.1.** Show that the empty set $\underline{0} = \emptyset$ is not inductive. Similarly, show that $\underline{1} = \{\emptyset\}$ is not inductive.

**Exercise 3.2.** Verify the claim in Example 3.5 that $\underline{2}$ is transitive. Is it inductive?

**Exercise 3.3.** Rewrite the axiom of infinity (3.1) without using the symbol $\emptyset$.

**Exercise 3.4.** Show that the axiom of empty set is a consequence of any axiom asserting the existence of some set (such as the axiom of infinity) together with an axiom of separation.

**Exercise 3.5.** Given $m, n \in \omega$, let us say that $m < n$ holds exactly when $m \in n$. Prove for each $n \in \omega$ that $n = \{m \in \omega : m < n\}$. (Hint: Use Corollary 3.7.)

**Exercise 3.6.** Prove that each $n \in \omega$ is transitive. (Hint: Modify the proofs of Theorem 3.6 and Corollary 3.7.)

**Exercise 3.7.** Prove the principle of mathematical induction in formal set theory: Suppose $S$ is a subset of $\omega$ such that $\underline{0} \in S$, and if $x \in S$ then $x+1 \in S$. Then $S = \omega$.

**Exercise 3.8.** Prove the following basic facts about the relation $<$ on elements of $\omega$.

1. There is no element $n \in \omega$ with $n < n$. In particular, $n \neq n+1$.
2. If $m, n, p \in \omega$, with $m < n$ and $n < p$, then $m < p$. (Hint: Use a previous exercise.)
3. If $m, n \in \omega$, then exactly one of the three conditions $m = n$, $m < n$, or $n < m$ holds.
4. If $X \subseteq \omega$ and $X \neq \emptyset$, then there is some $x \in X$ such that $y \not< x$ for each $y \in X$. (Hint: You might begin by first proving that each $n \in \omega$ has a similar property.)

**Exercise 3.9.** For each meta natural number $n$, give a recursive definition of

$$\{a_1, a_2, \ldots, a_{n+1}\},$$

thus proving the theorem schema that the set $\{a_1, a_2, \ldots, a_n\}$ exists for each such meta number $n$. (Make sure your definition works when $n = 0$.) Now do the same for ordered $n+1$-tuples $(a_1, a_2, \ldots, a_{n+1})$.

**Exercise 3.10.** Give a (very) simple description of $\bigcup \omega$, and prove that claim.

# 4 Power sets

Given a finite set like $S = \{1, 2, 3\}$, it is straightforward to create the collection of all its subsets—its power set:

$$\mathscr{P}(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Most mathematicians similarly assume that the power set of any infinite set exists. As we will see, the power set operation precipitates a plethora of constructions.

## 4.A Power set axiom

For each set $S$, the *axiom of power set* asserts that the class $\mathscr{P}(S) = \{x : x \subseteq S\}$ is a set. In formal symbols,

$$\forall S\, \exists P\, \forall x\, (x \in P \leftrightarrow \forall y\, (y \in x \rightarrow y \in S)) \tag{4.1}$$

By extensionality, the power set of $S$ is unique.

The power set axiom is key in defining many other basic concepts in set theory. For instance, given any two sets $A$ and $B$, we want to define their *product* to be

$$A \times B = \{z : z = (x, y) \text{ for some } x \in A,\ y \in B\}.$$

This shows that $A \times B$ is a class, but we want to prove that $A \times B$ is a set. To do that we will use an axiom of separation. Recall that the ordered pair $z = (x, y)$ is defined as the set $\{\{x\}, \{x, y\}\}$. This set is an element of $\mathscr{P}(\mathscr{P}(\{x, y\}))$. Thus,

$$A \times B = \{z \in \mathscr{P}(\mathscr{P}(A \cup B)) : z = (x, y) \text{ for some } x \in A,\ y \in B\},$$

which is a set by separation. For efficiency, we may now write

$$A \times B = \{(x, y) : x \in A,\ y \in B\}.$$

We write $A^2$ in place of $A \times A$. The 2 here is a meta natural number, counting the number of factors.

## 4.B Relations

In “Transition” a *relation* from a set $A$ to a set $B$ is defined to be a subset of $A \times B$. In formal set theory, it is common to leave $A$ and $B$ unnamed, and make the following simpler definition.

**Definition 4.2.** A (binary) *relation* is a set of ordered pairs.

In other words, if $R$ is a relation and $z \in R$, then $z = (x, y)$ for some sets $x$ and $y$.

**Example 4.3.** The set $R = \{(\underline{1}, \underline{3}), (\underline{2}, \underline{3})\}$ is a (very simple) relation; all its elements are ordered pairs. We write $\underline{1} R \underline{3}$ to express the fact that $\underline{1}$ relates to $\underline{3}$. On the other hand, we write $\underline{3} \not R \underline{1}$ to express the fact that $\underline{3}$ does not relate to $\underline{1}$. △

**Example 4.4.** Assuming the existence of the set of real numbers $\mathbb{R}$, then the set $\{(x, y) \in \mathbb{R}^2 : x^2 + 9y^2 = 16\}$ is a relation. In this sort of situation, it is common to interpret an ordered pair as a point in the plane, in which case this relation is an ellipse in the plane. △

Each ordered pair has a first coordinate and a second coordinate. Likewise, for each relation there is a set of first coordinates and a set of second coordinates. These sets have special names.

**Definition 4.5.** Let $R$ be a relation.

The *domain* of $R$ is $\operatorname{dom}(R) := \{x : \text{there is some } y \text{ such that } (x, y) \in R\}$.

The *range* of $R$ is $\operatorname{ran}(R) := \{y : \text{there is some } x \text{ such that } (x, y) \in R\}$.

**Warning 4.6.** In "Transition" we only defined "domain" and "range" for functions, but the definitions work just as well for relations.

**Example 4.7.** For the relation $R = \{(\underline{1}, \underline{3}), (\underline{2}, \underline{3})\}$, its domain is $\operatorname{dom}(R) = \{\underline{1}, \underline{2}\}$ and its range is $\operatorname{ran}(R) = \{\underline{3}\}$. △

**Example 4.8.** For the ellipse relation $\{(x, y) \in \mathbb{R}^2 : x^2 + 9y^2 = 16\}$, its domain is the interval $[-4, 4]$ and its range is the interval $[-4/3, 4/3]$.

Consider the parabola $\{(x, y) \in \mathbb{R}^2 : y = x^2\}$. What are its domain and range? Can they be described using interval notation? △

We defined the domain and range of a relation as classes, but we can prove that they are always sets.

**Proposition 4.9.** *The domain and range of a relation are both sets.*

*Proof.* Given a relation $R$, we defined the domain as the class

$$\operatorname{dom}(R) = \{x : \text{there is some } y \text{ such that } (x, y) \in R\}$$

The membership $(x, y) \in R$ means that $\{\{x\}, \{x, y\}\} \in R$. Hence $\{x\} \in \bigcup R$, and so $x \in \bigcup \bigcup R$. This means that we could have defined the domain as

$$\operatorname{dom}(R) = \left\{x \in \bigcup \bigcup R : \text{there is some } y \text{ such that } (x, y) \in R\right\},$$

which is a set by separation. The argument for $\operatorname{ran}(R)$ is similar. □

There are many real-life situations where two different relations are combined in important ways. For instance, let $R$ be the set of ordered pairs $(x, y)$ where $x$ is a person allergic to item $y$. Let $S$ be the set of ordered pairs $(y, z)$ where $y$ is an item found in food $z$. By combining the information from $R$ and $S$ in the correct way, we can tell which people should avoid which foods. This "combining" is captured by the following definition.

**Definition 4.10.** If $R$ and $S$ are relations, then the *composition* of $R$ and $S$ is

$$S \circ R = \{(x, z) \,:\, \text{there is some } y \text{ such that } (x, y) \in R \text{ and } (y, z) \in S\}.$$

**Example 4.11.** Let $R = \{(\underline{1}, \underline{3}), (\underline{2}, \underline{3})\}$ and $S = \{(\underline{3}, \underline{4}), (\underline{1}, \underline{2})\}$. By running through all possible pairs, we find that

$$S \circ R = \{(\underline{1}, \underline{4}), (\underline{2}, \underline{4})\}.$$

Can you find $R \circ S$? △

**Example 4.12.** Again consider the ellipse relation $R = \{(x, y) \in \mathbb{R}^2 \,:\, x^2 + 9y^2 = 16\}$ and the parabola relation $S = \{(y, z) \in \mathbb{R}^2 \,:\, z = y^2\}$. The composition $S \circ R$ is just

$$\{(x, z) \in \mathbb{R}^2 \,:\, \text{there is some } y \in \mathbb{R} \text{ such that } x^2 + 9y^2 = 16 \text{ and } z = y^2\},$$

which simplifies to $\{(x, z) \in \mathbb{R}^2 \,:\, x^2 + 9z = 16 \text{ and } z \geq 0\}$. Try to work out the reversed composition $R \circ S$. △

**Warning 4.13.** In "Transition" we only defined "composition" in very special circumstances. Here, the composition is always defined for two relations, even when the range of the first relation does not equal the domain of the second relation.

## 4.C Functions

Now that relations are defined, we can also talk about functions.

**Definition 4.14.** A relation $f$ is a *function* if $(x, y_1), (x, y_2) \in f$ implies $y_1 = y_2$.

In other words, an element of the domain of a function can relate to only one element of the range.

**Example 4.15.** The ellipse relation $R$ in Example 4.12 is not a function, but the parabola relation $S$ is a function. △

There is some disagreement among mathematicians whether or not a function should come equipped with a codomain. In formal set theory, functions do not have codomains. However, in mathematical practice, most functions do come attached with a codomain. Whenever we want to talk about codomains, we will write $f\colon A \to B$ to express the fact that $f$ is a function whose domain is $A$, and whose range is contained in the codomain $B$.

In this book, two functions are equal when they have the same ordered pairs. (We will not worry about whether or not they have the same intended codomains.)

We end this section by recalling the following three standard definitions.

**Definition 4.16.** Some function $f$ is *injective* (or an *injection*, or *1-1*) when $(x_1, y), (x_2, y) \in f$ implies $x_1 = x_2$.

If $B$ is a codomain for $f$, then $f\colon A \to B$ is *surjective* (or a *surjection*, or *onto*) exactly when $B = \operatorname{ran}(f)$.

Finally, $f\colon A \to B$ is *bijective* (or a *bijection*) when it is injective and surjective.

Thus, injectivity is a property of functions, but surjectivity and bijectivity are only properties of functions with intended codomains. Even though a function $f$ is a relation, rather than writing $(x, y) \in f$ or $xfy$, it is much more common to instead write $f(x) = y$.

**Example 4.17.** The function $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, given by the rule $f(x) = \sqrt{x}$, is a bijection.

First we prove injectivity. Let $x_1, x_2 \in \mathbb{R}_{\geq 0}$, and assume $f(x_1) = f(x_2)$. Then $\sqrt{x_1} = \sqrt{x_2}$. After squaring, we have $x_1 = x_2$, as desired.

Next we prove surjectivity. Letting $y \in \mathbb{R}_{\geq 0}$, our goal is to show that $y \in \operatorname{ran}(f)$. Fix $x = y^2 \in \mathbb{R}_{\geq 0}$. We see that $f(x) = \sqrt{x} = \sqrt{y^2} = y$. Thus, indeed, $y \in \operatorname{ran}(f)$. △

## 4.D Exercises

**Exercise 4.1.** For the parabola relation $\{(x, y) \in \mathbb{R}^2 : y = x^2\}$, find its domain and its range.

**Exercise 4.2.** Prove that the order relation defined on $\omega$ in Exercise 3.5 is a set.

**Exercise 4.3.** Finish the proof of Proposition 4.9, by showing that the range of a relation is a set.

**Exercise 4.4.** Letting $R$ and $S$ be the relations from Example 4.12, describe $R \circ S$. Also prove that neither $R$ nor $S$ is injective.

**Exercise 4.5.** If $R$ and $S$ are relations, prove that $S \circ R$ is a set.

**Exercise 4.6.** For any relation $R$, define $R^{-1} := \{(y, x) : (x, y) \in R\}$, called the *inverse relation*.

(1) Prove that $R$ is a function if and only if $R^{-1}$ is injective (where injectivity for a relation is defined just as for a function).
(2) Find an example showing that even when $R$ is a function, $R^{-1} \circ R$ may not be a function.

**Exercise 4.7.** A *class relation* is a class of ordered pairs. In other words, we just drop the assumption that the relation is a set. Prove that the class relation

$$\in \; = \{(x, y) \,:\, x \in y\}$$

is a proper class.

**Exercise 4.8.** Prove that Dedekind-infinite sets are infinite.

# Chapter II

# First Order Logic

*A man may be born, but in order to be born he must first die, and in order to die he must first awake.* Carl Sandburg

Propositional logic is the formalization of the simple logical arguments about truth that occur in everyday mathematical proof. It is extremely robust in what it does, but it also is quite limited in its expressibility. First order logic goes a step beyond propositional logic, formalizing logical arguments that involve quantified variables. The sentence “if it rains then I get wet” is modeled in propositional logic by a formula of the form $p \to q$. The sentence “all cats are blue” is more complicated, quantifying over a collection of objects (the cats) and proposing that they satisfy a property (being blue). This goes beyond what propositional logic was meant to do, but it fits perfectly into the framework of first order logic. First order logic was developed in the late 1800s and early 1900s, and is now a staple of many formal mathematical systems.

We’ve seen, in Chapter I, some sentences about sets expressed in the language of first order logic. The set theory we have already developed will serve as an important example to motivate our study of this even more fundamental topic. The work in this chapter will allow us to completely formalize the language of set theory, as well as that of many other mathematical theories. A study of first order logic also reveals interesting questions about the nature of language, teaches us how to create “well-formed” formulas, and provides a formal way to “minimize” logical considerations.

# 5 Language and semantics of propositional logic

In this section we describe the syntax—the formal language structure—of propositional logic (PL). We then connect that syntax to semantics—an intended meaning for the symbols.

As PL will be our current (formal) theory of study, the theorems *about* PL will be in our (informal) metatheory. Every instance of a natural number in this section is referring to a meta natural number, as there are no natural numbers in PL. So, with that understanding we will refrain from using the word "meta" in this section (and the next) unless it is deemed necessary or helpful.

## 5.A The building blocks of formulas

There are two parts to the syntax of propositional logic.

First come the *variables*, also called *propositional variables* or *symbols.* These are the $p$'s and $q$'s in formulas like $p \vee q$ and $\neg(p \to q)$. They are placeholders for actual statements. They are meant to take a value of true (T) or false (F).

It is possible to develop PL with just a few variables, but it is much more convenient to assume, from the start, that we have a (meta) countable number of variables. This gives authors freedom to use new variables at will. In this section we will fix, once and for all, our list of propositional variables to be

$$p, q, r, s, t, u, p_0, p_1, p_2, p_3, \dots.$$

We will use the letter $v$ to stand for an arbitrary variable; thus, $v$ ranges over the list of variables above, and the variables range over the values $\{\mathrm{T}, \mathrm{F}\}$.

Second come the *logical connectives*, also called logical operators. The most commonly used connectives are

| | |
|---|---|
| $\neg$ | negation |
| $\vee$ | disjunction |
| $\wedge$ | conjunction |
| $\to$ | implication |
| $\leftrightarrow$ | biconditional |

but there are others, such as XOR ("exclusive or"), which is commonly used in computer science. We let $\Omega$ be the collection of whichever logical connectives we decide to use.

**Remark 5.1.** Sometimes the word "material" appears next to "implication" to distinguish it from "*causal* implication". As we won't discuss temporal logic in this book, we will find no need to use that qualifier. ▲

Every connective has an *arity*, which is the number of operands that the connective operates on. Have you ever noticed that negation is different than the other connectives? When we write a negation, like $\neg p$, it acts on only one input. On the other hand, conjunction acts on two inputs, like $p \wedge q$.

Negation thus has arity 1, or is said to be a 1-*ary* or *unary* operator. All of the other connectives we have described above act on two variables, and so they have arity 2, which are called 2-*ary* or *binary* connectives. All connectives are written to the left of their operands, except that binary connectives are commonly written between the two operands. Thus, we write $p \vee q$ instead of $\vee(p, q)$.

**Example 5.2.** There are, of course, higher arity operators. For instance, consider the 3-ary "majority rules" connective $M$. Here, $M(p, q, r)$ represents the function "Take the most common of the three truth values that $p$, $q$, and $r$ represent". △

**Example 5.3.** A 0-ary connective has zero operands. Thus, any zeroary connective needs no variables to act on. There are two natural 0-ary operators. They are $\top$, called *tautology*, and $\bot$, called *contradiction*. (We will explain their meaning shortly.) △

## 5.B Formulas in PL

From the variables and the logical connectives in $\Omega$ we can construct propositional formulas. We have seen a few examples of such formulas above, including

$$p \vee q \text{ and } \neg(p \to q).$$

Even more complicated formulas like $q \to (\neg p \to \neg\neg q)$ or

$$(\neg(p \to (p \vee q))) \wedge ((r \vee p) \to q)$$

are also possible.

Looking at the examples above, we see that propositional formulas are built up out of simpler pieces. Each (well-formed) *formula* is constructed recursively by the following two rules.

(1) Any variable is a formula.
(2) If $\varphi_1, \varphi_2, \ldots, \varphi_n$ are each formulas and $C$ is an $n$-ary connective in $\Omega$, then $C(\varphi_1, \varphi_2, \ldots, \varphi_n)$ is a formula.

**Example 5.4.** The variables $p$ and $q$ are formulas, by the first rule. If $\neg, \vee \in \Omega$, then $\neg p$ is a formula by the second rule, and hence $(\neg p) \vee q$ is a formula by another application of the second rule. △

**Example 5.5.** We do not allow expressions like $p \vee (p \vee (p \vee \cdots))$ to count as formulas, because such expressions take an infinite number of steps to construct. We only allow those expressions that are built up, using only a (meta) finite numbers of steps, using these rules.

We also do not allow expressions like $p\vee$ to count as formulas. Rule (2) requires that the number of operands matches the arity of any connective. However, expressions like $p \vee p$ are formulas; the operands can be repeated.

Finally, if $\wedge \notin \Omega$, then an expression like $p \wedge q$ will not technically be a formula in our given language. However, we will see later that it is sometimes possible to express it as a (different) formula in our language. △

For any formula we recursively define its *complexity* as follows: Any variable has complexity 0. If $\varphi_1, \varphi_2, \ldots, \varphi_n$ are formulas with complexities $m_1, m_2, \ldots, m_n$, and $C$ is an $n$-ary connective, then we define the complexity of $C(\varphi_1, \varphi_2, \ldots, \varphi_n)$ to be $\max\{m_1, m_2, \ldots, m_n\} + 1$. (Note that the 0-ary connectives have complexity 1.) For example, the complexity of $(\neg p) \vee q$ is 2. What is the complexity of $p \to (p \to \neg p)$?

We can prove statements about formulas by induction on their complexity.

**Metatheorem 5.6.** *Any well-formed formula involves only a finite number of variables and connectives.*

*Proof.* We induct on the complexity of the formula. If the formula has complexity 0, then it is a variable and the statement is obvious. So assume the complexity is greater than 0, but also that the metatheorem is true for all formulas of smaller complexity. Now, our formula is of the form $C(\varphi_1, \varphi_2, \ldots, \varphi_n)$ for some $n$-ary connective $C$ and some formulas $\varphi_1, \varphi_2, \ldots, \varphi_n$ of smaller complexity. By our inductive assumption, each $\varphi_i$ involves only finitely many variables and connectives. Those, together with $C$, still make a finite total. □

Remember that we let $\Omega$ be the collection of whatever set of logical connectives that we decide to use. The set of possible formulas changes according to our choice of $\Omega$. The following are two different typical choices for $\Omega$, but there are many others.

$$\begin{aligned} \text{Standard connectives} &= \{\top, \bot, \neg, \vee, \wedge, \to, \leftrightarrow\} \\ \text{Boolean connectives} &= \{\top, \bot, \neg, \vee, \wedge\} \end{aligned}$$

If one takes $\Omega = \emptyset$, then the only formulas are the variables, and it seems intuitively obvious that these formulas are not expressive enough to allow PL to do what it was designed to do. To settle on what makes one choice of $\Omega$ better than another we will first need to discuss another property that the connectives themselves possess.

## 5.C Semantics of connectives and formulas

Each logical connective comes not only with an arity but also with an intended interpretation, as a function on truth values. The values of that function are captured by a truth table. For instance, conjunction has a four line truth table

| $p$ | $q$ | $p \wedge q$ |
|---|---|---|
| T | T | T |
| T | F | F |
| F | T | F |
| F | F | F |

and so we think of conjunction as representing the function $\{\mathrm{T}, \mathrm{F}\}^2 \to \{\mathrm{T}, \mathrm{F}\}$ given by the rule

$$\{((\mathrm{T}, \mathrm{T}), \mathrm{T}), ((\mathrm{T}, \mathrm{F}), \mathrm{F}), ((\mathrm{F}, \mathrm{T}), \mathrm{F}), ((\mathrm{F}, \mathrm{F}), \mathrm{F})\}.$$

Similarly, tautology has the one line truth table

| $\top$ |
|---|
| T |

and represents the function $\{((), \mathrm{T})\}$ where the empty tuple maps to T. (Contradiction has the opposite truth table.) An $n$-ary connective has a $2^n$ line truth table to describe its intended interpretation. As a quick exercise in understanding, create the eight line truth table for $M$, the majority rules connective, and also write its intended function.

Generally, an *interpretation* of a mathematical theory is a way to assign semantics, or meaning, to the formulas of that theory. In the case of propositional logic, an interpretation is simply a map from the set of all well-formed formulas to the set $\{\mathrm{T}, \mathrm{F}\}$ that also evaluates each logical connective according to the function it represents. An interpretation forces each variable to take some value in $\{\mathrm{T}, \mathrm{F}\}$; this choice picks out which line of a truth table we are considering. After that, we send formulas of complexity 1 to their intended values, then formulas of complexity 2 to their values, and so forth.

We will work out one example; the details of the general case are left as an exercise. Suppose we map $p \mapsto \mathrm{T}$ and $q \mapsto \mathrm{F}$, and all the other variables are also sent to F. We see that $\neg p \mapsto \mathrm{F}$. This in turn forces $\neg p \vee q \mapsto \mathrm{F}$. The truth values of other formulas in this interpretation are found similarly.

It can be beneficial to think of interpretations as *possible worlds*. When we want to know the truth value of a formula, like $\neg p \vee q$, in a possible world, we only need to know the truth values of the variables in that world. The variables can change their truth values from world to world, but the functional way that logical connectives handle those truth values does not change—they have a fixed intended meaning that does not vary from world to world.

Now that formulas have interpretations we can make an important definition.

**Definition 5.7.** Two formulas $\varphi, \psi$ are *logically equivalent*, written $\varphi \equiv \psi$, if their truth values are the same under every interpretation.

**Example 5.8.** The formulas $\neg p \vee q$ and $p \to q$ are logically equivalent, as can easily be checked by constructing a four line truth table. One consequence of this fact is that we don't need $\to$ in our set $\Omega$ of logical connectives if $\Omega$ already contains $\neg$ and $\vee$, since implication can now be defined in terms of the other two connectives.

One benefit of removing $\to$ from $\Omega$ is that when proving (meta)theorems about formulas, there are less logical connectives that need to be considered. One drawback is that formulas become much harder (for humans) to read and write. $\triangle$

If our set $\Omega$ of logical connectives is too small, then it is possible that our formulas are too few to express arbitrary formulas.

**Example 5.9.** Consider, for a moment, what happens if we take $\Omega = \{\neg\}$. It is clear (by induction) that the only formulas are of the form $v, \neg v, \neg\neg v, \ldots$, for some variable $v$. Further, $v$ is logically equivalent to $\neg\neg v$ and all other formulas with an even number of negations (in total). Similarly, $\neg v$ is logically equivalent to those formulas with an odd number of negations.

We wish to show that there is no formula $\varphi$, built using $\Omega = \{\neg\}$, that is logically equivalent to $p \wedge q$. There are three cases to consider: $v = p$, $v = q$, or $v$ is some other variable. In all three cases, a direct check using truth tables shows us that neither $v$ nor $\neg v$ has the required truth value in some possible world. △

Let $\Omega^*$ be the set of all possible logical connectives. A subset $\Omega \subseteq \Omega^*$ of logical connectives is *functionally complete* if every formula built using $\Omega^*$ is logically equivalent to a formula using only the connectives in $\Omega$. For instance, $\Omega^*$ itself is, obviously, functionally complete. An important, and somewhat magical, fact is that some *finite* sets of logical connectives are functionally complete. The most natural example is the set of Boolean connectives, but even that set isn't minimal (see the exercises). We want $\Omega$ to be functionally complete so that our formulas are as expressive as possible. For the rest of the book, unless we explicitly say otherwise (such as in the exercises below), our set $\Omega$ of connectives will be the standard ones, and it is thus functionally complete.

## 5.D Exercises

**Exercise 5.1.** For each of the following expressions, if it is a formula (formed from the standard connectives) find its complexity, and if it is not a formula explain why.

(1) $((p \leftrightarrow p) \leftrightarrow p) \leftrightarrow p$.
(2) $q \vee \ \vee r$.
(3) $\neg$.
(4) $\neg((p \wedge \neg q) \vee (r \vee s))$.
(5) $\mathrm{XOR}(p, p)$.

**Exercise 5.2.** Let $M$ be the majority rules connective. Write the truth table for $M(p, q, r)$, and also write the function $\{\mathrm{T}, \mathrm{F}\}^3 \to \{\mathrm{T}, \mathrm{F}\}$ that $M$ represents (as a set of ordered pairs, where the first coordinate of the ordered pair is a triple).

**Exercise 5.3.** Prove that $(p \leftrightarrow q) \leftrightarrow r$ is not logically equivalent to $(p \leftrightarrow q) \wedge (q \leftrightarrow r)$.

In linear algebra, students are told that the following are equivalent conditions on a square matrix: (1) The matrix is invertible. (2) The rows are linearly independent. (3) The columns are linearly independent.

You might see this equivalence written as: A matrix is invertible if and only if its rows are linearly independent, if and only if its columns are linearly independent.

Which of the two formulas above correctly describes this situation?

**Exercise 5.4.** It is well-known that disjunction is associative, meaning that

$$(p \vee q) \vee r \equiv p \vee (q \vee r).$$

Thus, it is common to simply drop parentheses and write $p \vee q \vee r$ to represent either of those formulas.

Prove that the biconditional is also associative, or in other words $(p \leftrightarrow q) \leftrightarrow r \equiv p \leftrightarrow (q \leftrightarrow r)$.

**Exercise 5.5.** It is possible to write all propositional formulas without parentheses or commas, by simply placing all connectives to the left of their operands (even the binary connectives). This is called Polish notation. For instance, the formula $p \wedge \neg q$ becomes $\wedge p \neg q$, and the more complicated formula $(p \wedge \neg q) \to r$ becomes $\to \wedge p \neg q r$.

Write $p \wedge ((q \vee \neg r) \to s)$ in Polish notation.

**Exercise 5.6.** Using the standard connectives we can often describe some of the other logical connectives. For instance, we can take NOR$(p, q)$ as shorthand for the (primitive) formula $\neg(p \vee q)$.

What is the arity of NOR? Write the truth table for NOR.

**Exercise 5.7.** Find a formula $\varphi$ in the Boolean connectives that is logically equivalent to XOR$(p, q)$. Similarly, find a formula $\psi$ in the Boolean connectives that is logically equivalent to $p \leftrightarrow q$.

**Exercise 5.8.** For any natural number $n$, how many $n$-ary connectives are there? Give a proof of that claimed metatheorem.

**Exercise 5.9** (Disjunctive normal form)**.** In this exercise we describe how to give, for most logical connectives of positive arity, a logically equivalent formula in terms of $\neg$, $\wedge$, and $\vee$, in a very specific form.

(1) Given variables $p_1, p_2, \ldots, p_n$, the truth table of all possible truth assignments has $2^n$ rows. Show that any $n$-fold conjunction $x = x_1 \wedge x_2 \wedge \cdots \wedge x_n$, where each $x_i$ is either $p_i$ or $\neg p_i$, will be true on exactly one row.

(2) Show that by choosing each $x_i$ appropriately, as either $p_i$ or $\neg p_i$, we can make $x$ as in part (1) have its true value occur on any given row. (Hint: A counting argument might make this step simpler.)

(3) Given a nonempty subset of rows, show that a disjunction of formulas of the previous form (with the number of terms in the disjunction matching the size of the subset) is true exactly on that subset. This is called the *disjunctive normal form* for the formula.

(4) Write the majority rules formula $M(p, q, r)$ in disjunctive normal form.

**Exercise 5.10.** Show that the set $\{\neg, \vee\}$ is functionally complete, by defining $\wedge$ using these connectives and finding formulas logically equivalent to $\top$ and $\bot$.

**Exercise 5.11.** Show that the set {NOR} is functionally complete.

**Exercise 5.12.** If $\Omega$ is the set of all 0-ary and 1-ary logical connectives, prove that it is not functionally complete. Thus, functionally complete sets need to contain at least one $n$-ary connective with $n \geq 2$.

**Exercise 5.13.** Prove that $\{\wedge\}$ is not functionally complete.

**Exercise 5.14.** Prove that if $\Omega$ consists of exactly one binary logical connective and also $\Omega$ is functionally complete, then $\Omega$ is either {NOR} or {NAND}, where NAND$(p, q) \equiv \neg(p \wedge q)$. (Hint: Assume $\{C\}$ is functionally complete, and refine the possible truth tables for $C(p, q)$.)

**Exercise 5.15.** Let $S$ be the set of all well-formed formulas and let Var be the set of variables. Let $f_0\colon \mathrm{Var} \to \{\mathrm{T}, \mathrm{F}\}$ be an arbitrary function. Prove (the metatheorem) that $f_0$ extends to a unique map $f\colon S \to \{\mathrm{T}, \mathrm{F}\}$ such that $f$ respects the semantic meaning of the logical connectives. (Hint: Extend $f_0$ by induction.)

**Exercise 5.16.** A concept (only very slightly) stronger than functional completeness is *connective completeness*, which means that each logical connective $C$ can be defined—as a derived connective—using only the connectives in $\Omega$, and where the defining formula uses the same number of variables as the arity of $C$.

For example, even though $p \to q$ is logically equivalent to $((\neg p) \vee q) \wedge (r \vee \neg r)$, the second formula involves three variables, and so it defines a 3-ary operation, and thus it does not define implication. However, the simpler formula $(\neg p) \vee q$ does define implication as a derived connective.

(1) Show that no 0-ary operation can be defined by a formula using only connectives of positive arity. (Note that the formula $p \vee \neg p$ is logically equivalent to $\top$, but it defines a unary operation, not a zeroary operation.)

(2) Show that if $\Omega$ is any functionally complete set of connectives, then every connective of positive arity can be defined as a derived connective.

# 6 Deductive systems for propositional logic

In the previous section we defined the language of PL, and we described how to interpret formulas in possible worlds. This allows us to study the important set of formulas that happen to be true in every possible world, which we do in this section.

We will also introduce an alternative means of studying this collection of formulas, by introducing a deductive system in the style of Hilbert and Frege. This system gives us a syntactic method for proving formulas.

## 6.A Validity

It is tempting to say that a formula like $p \vee \neg p$ is true. Technically speaking, it only obtains a truth value when we interpret it in some possible world. However, $p \vee \neg p$ is special because it evaluates as true in *every* possible world. The formulas with this property are given a special name.

**Definition 6.1.** A formula $\varphi$ is *valid* if it is logically equivalent to $\top$, i.e., it is true in every interpretation. In this case we write $\models \varphi$. We also write $\models_{\mathrm{PL}} \varphi$ when we want to emphasize that we are working in propositional logic.

Valid formulas are also (by an overuse of the term) often called tautologies, but we won't follow that convention in this book. For a valid formula $\varphi$, we read "$\models \varphi$" as "$\varphi$ is *semantically entailed* in propositional logic" or "propositional logic *models* the formula $\varphi$" or just "$\varphi$ is *valid*".

**Example 6.2.** We have $\models p \vee \neg p$, as a quick truth table computation verifies. Other valid formulas include $p \to p$ and $p \leftrightarrow p$. △

**Example 6.3.** For the variable $p$ (or any other variable), we know that $p$ is not valid, since there is a possible world where $p$ is false. Thus, we may write $\not\models p$. Similarly, we also have $\not\models \neg\neg p$. △

**Example 6.4.** We know that the formula $p \vee (q \wedge s)$ is not valid, since it evaluates as false in the possible world where every variable evaluates to F. When proving validity an entire truth table must be constructed, but to disprove validity it suffices to find a single possible world where the formula evaluates as false. △

The symbol $\models$ is a binary relation symbol. This may seem quite surprising, since there does not appear to be anything to the left of symbol. That is because there is a hidden object appearing to the left; it is $\emptyset$. However, rather than writing $\emptyset \models \varphi$ we drop the empty set symbol for simplicity.

More generally, for any two sets of formulas $X$ and $Y$, we write $X \models Y$ to mean that every formula of $Y$ only has truth value T in the possible worlds where every formula in $X$ also evaluates to T. In this case, we say that "$X$ models $Y$". We might say that in possible worlds where the formulas in $X$ are true, then so are the formulas in $Y$. When $X$ or $Y$ is a singleton set we usually drop the set braces.

**Example 6.5.** Is it the case that $q \models p \to q$? Yes, because if we limit ourselves to worlds where $q$ must be true, then $p \to q$ must also be true in those worlds. △

**Example 6.6.** For what formulas $\varphi$ is it the case that $\{p, \neg p\} \models \varphi$? Since there are no possible worlds where both $p$ and $\neg p$ hold true, every formula $\varphi$ is semantically entailed in such worlds, vacuously. △

**Example 6.7.** Consider the three formulas $p$, $q$, and $\neg q \to \neg p$. For which of these formulas $\varphi$ is it the case that $p \to q \models \varphi$?

Make the truth table with all four possibilities for truth values of the variables $p$ and $q$. We can cross out the second line, where $p \to q$ is false. Looking at these three formulas, we see that only $\neg q \to \neg p$ is true in all three remaining possible worlds.

| $p$ | $q$ | $p \to q$ | $\neg q \to \neg p$ |
|---|---|---|---|
| T | T | T | T |
| ~~T~~ | ~~F~~ | ~~F~~ | ~~F~~ |
| F | T | T | T |
| F | F | T | T |

△

## 6.B Syntactic deduction

Rather than relying alone on the semantic meaning of symbols, one can alternatively study a mathematical system by formalizing aspects of its structure via axioms and rules of inference. This creates another way to bring mathematical tools to bear in analyzing mathematics itself. Moreover, the axiomatic method often reveals how a theorem is produced from the underlying assumptions, or perhaps showing it needs fewer assumptions than were previously thought necessary.

The set of axioms and inference rules shouldn't be too large, or there is a lot of repetition and complication. On the other hand, there shouldn't be too few of them, or the system becomes difficult to work with, or simply doesn't prove much. The axiom system we will use is due to Mendelson, and it takes $\Omega = \{\neg, \to\}$ (which is functionally complete). Mendelson's system is by no means the only axiomatic system available, nor the first, nor the simplest to state. We use it just for convenience of exposition. This, and other examples, are sometimes called "Hilbert deductive systems". Mendelson's system has three axioms, which are:

(1) $A \to (B \to A)$.
(2) $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$.
(3) $(\neg A \to \neg B) \to ((\neg A \to B) \to A)$.

Technically, these are not three axioms, but rather three axioms schemas, where each of the symbols $A$, $B$, and $C$ can be replaced by an arbitrary formula. It is this property—axioms are true under substitutions—that gives propositional logic the name "logic".

As is the usual practice for deductive systems in the style of Hilbert, we take only *one* rule of inference: *modus ponens* (MP). It states that if we have deduced both $A$ and $A \to B$, then we may deduce $B$.

**Example 6.8.** Using the axioms above, we will show that $A \to A$. Here is the formal proof:

| | | |
|---|---|---|
| 1. | $A \to ((A \to A) \to A)$ | Axiom 1. |
| 2. | $(A \to ((A \to A) \to A)) \to ((A \to (A \to A)) \to (A \to A))$ | Axiom 2. |
| 3. | $(A \to (A \to A)) \to (A \to A)$ | MP lines 1, 2. |
| 4. | $A \to (A \to A)$ | Axiom 1. |
| 5. | $A \to A$ | MP lines 4, 3. |

Technically this is not a theorem of propositional logic; it only becomes a theorem when we replace $A$ with a formula. Instead, this is a theorem schema, giving us a template how to prove $\varphi \to \varphi$ for any formula $\varphi$. With this theorem schema now proven, we could now add $A \to A$ to our list of axioms schemas, to simplify future computations. △

**Example 6.9.** We will prove that $(\neg A \to A) \to A$.

| | | |
|---|---|---|
| 1. | $\neg A \to \neg A$ | Example 6.8, with $\neg A$ replacing $A$. |
| 2. | $(\neg A \to \neg A) \to ((\neg A \to A) \to A)$ | Axiom 3. |
| 3. | $((\neg A \to A) \to A)$ | MP lines 1, 2. |

Technically speaking, this proof is using Example 6.8 as a sort of lemma. If we want to only use the axioms, then the first line of this proof must be replaced by five new lines (created from Example 6.8, replacing each $A$ with $\neg A$). △

**Example 6.10.** We will show that assuming $\neg B \to \neg A$, we can prove $A \to B$.

| | | |
|---|---|---|
| 1. | $\neg B \to \neg A$ | Assumption. |
| 2. | $(\neg B \to \neg A) \to ((\neg B \to A) \to B)$ | Axiom 3. |
| 3. | $(\neg B \to A) \to B$ | MP lines 1, 2. |
| 4. | $A \to (\neg B \to A)$ | Axiom 1. |
| 5. | $A \to B$ | Exercise 6.4(b), lines 4, 3. |

Exercise 6.4 saves many lines in this proof. △

As was noticed immediately by those developing these formal theories, it can be very hard to prove obvious facts. It takes only a two line truth table to verify that $p \to p$ is a valid formula, but it takes a five line (and not necessarily easy to find) deduction using this axiomatic system. However, with a little work and patience, enough lemmas can be built up to make the system usable. Also, note that we can add other logical connectives to our system as shorthand (or *derived*) notions for other longer formulas that only use *primitive* symbols from $\Omega$. For instance $p \vee q$ is shorthand for $(\neg p) \to q$.

There is a special symbol that expresses when a formula is provable.

**Definition 6.11.** If $\varphi$ is a formula that is entailed from our axiom system, using the given rules of inference, we write $\vdash \varphi$, read as "$\varphi$ is provable". The symbol $\vdash$ is called the *syntactic entailment* or *provability* relation. More generally we write $X \vdash Y$ if every formula in the set $Y$ is provable from the axioms, the formulas in $X$, and the inference rules.

The most obvious property that one wants for an axiomatic system is that it only proves those formulas that are true in every possible world—the valid formulas. For PL that means we want our system to only prove formulas that are logically equivalent to tautology. This property is called *soundness*. Of course, if we add the wrong axioms or an unreasonable inference rule to our system, we wouldn't expect it to be sound. However, it is extremely easy to show that Mendelson's system is sound.

**Metatheorem 6.12** (Propositional logic is sound)**.** *Suppose $\varphi$ is a well-formed formula. If $\vdash \varphi$, then $\models \varphi$.*

*Proof.* Truth tables show that the three axioms are valid. Similarly, if $A$ and $A \to B$ both evaluate as true in a possible world, then $B$ must evaluate as true. Therefore, modus ponens respects truth. Thus, any line in a formal proof, including the last one, must be true in every interpretation. □

A natural question is whether the converse of Metatheorem 6.12 holds. In other words we can ask: If a statement is valid, can we prove it using just our three axioms and one inference rule? If so, our axiomatic system is called *complete* or, more precisely, *semantically complete* (to avoid confusion with other forms of completeness, such as functional completeness). The answer is yes. We won't prove here the "Completeness theorem for propositional logic" but it is a straightforward proof by cases (handling each logical connective in $\Omega$ separately) that reduces valid statements from complicated formulas to simpler cases.

There are many interesting examples of axiomatic systems for PL besides the one we have presented here. We will give four more examples. Each of the types of systems below is sound, functionally complete, and semantically complete.

(1) There are axiomatic systems for PL with no rules of inference but lots of axiom schemas. One especially obvious example of this situation is to take the set of axioms to equal the set of valid formulas.

(2) There are axiomatic systems for PL with one rule of inference, one binary logical connective, and one axiom schema. For instance, take the logical connective to be NAND (the negation of conjunction), written $p \mid q$. The inference rule is Nicod's modus ponens, given by $\{A, A \mid (B \mid C)\} \vdash C$. As proven by Łukasiewicz, the single axiom schema

$$(A \mid (B \mid C)) \mid [(D \mid (D \mid D)) \mid ((D \mid B) \mid [(A \mid D) \mid (A \mid D)])]$$

suffices. There are other examples of this type.

(3) There are axiomatic systems for PL with no axioms and finitely many rules of inference. These systems are called "natural deduction systems", because they model closely how mathematicians reason in practice. Usually they involve all of the standard connectives.

(4) Axiom schemas can be replaced by individual axioms if you also add a single new rule of inference called "substitution" that lets you replace variables with more complicated formulas.

## 6.C More on completeness

Earlier we defined functional completeness and semantic completeness. There is another property that goes by the name (with no qualifiers) "completeness", but more properly should be "syntactical completeness". A deductive system is *syntactically complete* if for any formula $\varphi$, either $\vdash \varphi$ or $\vdash \neg\varphi$. Clearly Mendelson's system is not syntactically complete, nor would we want it to be, for then we would only be working inside a single interpretation, and there is no reason in PL to pick one possible world over another. Thus, Mendelson's system is our first example of a formal system where some formulas can neither be proved nor disproved.

In fact there are situations where mathematicians study deduction systems that are not even semantically complete. For instance, intuitionists reject the law of the excluded middle and other nonconstructive methods; they focus on that portion of mathematics that can be explicitly constructed. Thus, they reject the axiom system of Mendelson given above (specifically, replacing axiom schema (3) with other, weaker axioms), and thus they use a weaker system that cannot prove all valid formulas, only the "constructively valid" ones.

## 6.D Exercises

**Exercise 6.1.** For each of the formulas listed below, decide (with proof) whether or not it is valid.

(1) $(p \wedge q) \vee (\neg p \wedge \neg q)$.
(2) $\neg(p \vee q) \to \neg q$.

**Exercise 6.2.** For which of the following formulas $\varphi$ is it the case that $p \leftrightarrow q \models \varphi$? Justify your answers.

(1) $\neg p \leftrightarrow \neg q$.
(2) $p \to q$.
(3) $p \vee q$.
(4) $(p \wedge q) \vee (\neg p \wedge \neg q)$.

**Exercise 6.3.** Let $\varphi$, $\psi$, and $\theta$ be formulas. Prove the following:

(1) $\varphi \models \varphi$.
(2) If $\varphi \models \psi$ and $\psi \models \theta$, then $\varphi \models \theta$.

Thus, $\models$ is a reflexive and transitive relation. Give an example to show that $\models$ is not symmetric.

**Exercise 6.4.** For formulas $\varphi, \psi, \theta$, do the following:

(1) Show $\{\varphi, \varphi \to \psi, \psi \to \theta\} \vdash \theta$.
(2) Show $\{\varphi \to \psi, \psi \to \theta\} \vdash \varphi \to \theta$.

(One of these is much easier to do than the other.)

**Exercise 6.5.** Let $\varphi$, $\psi$, and $\theta$ be formulas. Do the following:

(1) Prove that if $\psi$ is an instance of one of Mendelson's three axioms, then we have $\vdash \varphi \to \psi$. (Hint: Use axiom (1).)
(2) Prove that if $\vdash \varphi \to \psi$ and $\vdash \varphi \to (\psi \to \theta)$, then $\vdash \varphi \to \theta$.

**Exercise 6.6.** Let $\varphi$ and $\psi$ be formulas. Do the following:

(1) Prove that if $\vdash \varphi \to \psi$, then $\varphi \vdash \psi$. (Hint: Assume $\varphi \to \psi$ is provable. Add two lines to the end of a proof, under the extra assumption of $\varphi$.)

(2) Prove conversely that if $\varphi \vdash \psi$, then $\vdash \varphi \to \psi$. (Hint: Assume $\varphi \vdash \psi$, so there is some proof of $\psi$ using only the axioms, the formula $\varphi$, and modus ponens. Let $\theta_1, \theta_2, \ldots, \theta_n$ be the lines of the proof. Show that $\vdash \varphi \to \theta_i$, for each $i \geq 1$. Exercise 6.5 might help.)

**Exercise 6.7** (Deduction theorem)**.** Let $S$ be a set of formulas, and let $\varphi$ and $\psi$ be formulas. Prove the Deduction (meta)theorem, which says: If $S \cup \{\varphi\} \vdash \psi$, then $S \vdash \varphi \to \psi$. (Hint: Use the same idea as in the previous exercise, except that now some of the lines of the proof can be elements of $S$.)

The Deduction theorem shortens many proofs. For instance, it shows that the two parts of Exercise 6.4 are equivalent.

**Exercise 6.8.** Prove that $\{A \to (B \to C), B\} \vdash A \to C$. (Can you do it without the Deduction theorem? If so, how much longer is the proof?)

**Exercise 6.9.** (This is a difficult exercise if you do not have the deduction theorem.) Show $\vdash \neg\neg A \to A$. (Comparatively, to show $\models \neg\neg A \to A$ takes just a two line truth table.)

**Exercise 6.10.** Let $\varphi$ be a formula, and let $S$ be any set of formulas. Prove that $S \vdash \varphi$ implies $S \models \varphi$. This tells us that PL is *strongly sound*. (Hint: Proceed as in Metatheorem 6.12, recursing over the lines of a proof of $\varphi$ that may use $S$.)

**Exercise 6.11.** Describe a deductive system that is not sound.

**Exercise 6.12.** Show that if we replace Mendelson's three axioms with just the single axiom $A \to A$, then the system is not semantically complete.

# 7 The language of first order logic

Propositional logic has been studied for centuries, going at least as far back as to the ancient Greek logicians. Thus, it is somewhat surprising that there are still advances in our knowledge about the subject, such as shorter axiomatic presentations, results about useful sublogics, and so forth.

First order logic (FOL) extends PL by allowing quantification. The mathematical foundations of FOL are much more recent, starting near the turn of the 20th century. There continue to be major philosophical, logical, and mathematical questions concerning FOL, but it has become *the* preeminent formal language of mathematicians.

## 7.A Logical and nonlogical symbols

First order logic is more expressive than propositional logic, allowing us to formalize sentences like “All the dogs in my neighborhood have homes” and “Every set has a power set”. We build formulas using a language $\mathscr{L}$ of FOL. An example of a formula in formal set theory is

$$\forall S\,\exists T\,(S \in T \to S = T).$$

To form this formula, it appears that not only do we need variables (like $S$ and $T$) but relations like $\in$ and $=$, as well as logical connectives and quantifiers. One should think of FOL as a template to which can be added any number of relations, axioms, and other needed structures, as necessary. We make this precise as follows.

There are two natural components to a language $\mathscr{L}$: the logical symbols and the nonlogical symbols. The logical symbols consist of the following four ingredients.

**(L1) Variables.** The *variables* are the most basic building blocks of formulas in FOL. Rather than fixing a countable list of them, as we did in the previous two sections, it is usual practice to simply introduce them as needed. For instance, in Section 1 the set variables we employed included

$$S, T, A, B, C, x, y, z, x', y', a, b, c, d.$$

Unlike the variables of PL, the variables in FOL do not (usually) range over $\{\mathrm{T}, \mathrm{F}\}$. Instead they range over whatever objects you wish to study. If you are studying sets, the variables range over sets. If you are studying numbers, then the variables range over numbers. In particular, variables of FOL do not normally have truth values.

**(L2) Logical connectives.** In FOL we use the standard *logical connectives* with their usual meanings. We may sometimes limit ourselves to a functionally complete subset of the connectives when proving metatheorems, but we use all the standard connectives when actually working in FOL. Note that because variables have no truth values, it makes no sense to form sentences like $S \to T$. (What would it mean for a set to imply another set?) Instead, the logical connectives will only operate on more complicated expressions, such as $S \in T \to S = T$. We will elucidate this shortly.

**(L3) Quantifiers.** There are just two primitive quantifiers in FOL, the *universal quantifier* $\forall$ and the *existential quantifier* $\exists$. Any parentheses, brackets, commas, or other punctuation are just conveniences to make formulas readable. Just as logical connectives have a specific semantic meaning attached to them, the two quantifiers also have specific meanings, which you have likely learned previously and which we will further explain shortly.

**(L4) Equality.** The *equality relation* is usually (but not always) thought of as a part of the logic. In this book we will always consider it a part of the logic. Equality has the fixed meaning of asserting that two objects are identical, as usual.

There are three types of nonlogical symbols.

**(NL1) Constant symbols.** The *constants* are symbols used to describe specific objects being studied. They do not represent varying quantities as variables do. Some constants are primitive, while others are not. We illustrate this idea with two examples.

**Example 7.1.** In the formal set theory described in Chapter I, we used symbols like $\emptyset$ and $\omega$ to represent specific sets. These are derived constants, not primitive constants, because they were defined in terms of more basic concepts. △

**Example 7.2.** Suppose we want to formalize the theory of natural numbers. We might take 0 as a primitive constant symbol, to represent the number zero.

Let $S$ represent the "successor" function (i.e., it takes a natural number to the next natural number). Thus $S(0)$ would represent the number one, and $S(S(0))$ would represent two, and so forth. To shorten notation, we may write 1 for $S(0)$, in which case 1 is a derived constant symbol. It is not primitive, as it is defined in terms of the more primitive concepts 0 and $S$. △

**(NL2) Relation symbols.** The second type of nonlogical symbols are the *relation symbols* (also sometimes called predicates). Just like with the logical connectives, these come equipped with an arity, but we will see that unlike the connectives they do not come equipped with a fixed meaning.

In formal set theory there is exactly one primitive binary relation symbol, $\in$. The binary relation symbol $\subseteq$ is not a primitive symbol, because it is defined in terms of more primitive notions. When studying real numbers, one might take $<$ to be a primitive binary relation symbol.

**(NL3) Function symbols.** The last of the nonlogical symbols are the *function symbols*. These also have arities.

There are no primitive function symbols in formal set theory, but there are many derived functions. For instance, $\cup$ is a derived 2-ary function symbol and $\mathscr{P}$ is a derived 1-ary function symbol. In the theory of the natural numbers described in Example 7.2, the successor function $S$ is a primitive 1-ary function symbol.

**Remark 7.3.** Functions are technically relations with extra properties. One could try to do away with function symbols altogether and just use relation symbols. However, that makes FOL harder to explain, as functions play a special role. So, few people choose to do away with them.

Constants are technically 0-ary functions. There is no harm in thinking of them as separate from the functions symbols, nor any harm in thinking of them as part of the same family.

Finally, some people propose moving variables from the list of logical symbols to the list of nonlogical symbols by thinking of them as 0-ary relation symbols. On one hand this change makes some sense since variables don't have a fixed interpretation, unlike the other logical symbols. However, this change affects the language of FOL in other drastic ways, so we won't follow this proposal. (Also, 0-ary relations are hard to think about!) ▲

The list of the *primitive* nonlogical symbols, according to their type and arity, is called the *signature* of the language $\mathscr{L}$. For instance, the signature of the language formalized in Example 7.2 is one constant symbol and one unary function symbol. The signature of formal set theory is one binary relation symbol. We might write these two signatures as $(0, S)$ and $(\in)$, respectively, although the actual symbols used in the signature do not matter; only their arities and types matter.

**Example 7.4.** The set of real numbers, $\mathbb{R}$, has an ordering relation $<$, as well as the two functions addition $+$ and multiplication $\cdot$. It also has two important constants 0 and 1. Thus, one could try to formalize the structure of $\mathbb{R}$ using the signature of two constant symbols, one binary relation symbol, and two binary function symbols. We might use the same symbols and write $(0, 1, <, +, \cdot)$ for the signature.

It should be unclear, at this point, if this signature, or any other, can fully capture the nature of $\mathbb{R}$. △

## 7.B Formula formation

Recall the formula

$$\forall S\, \exists T\, (S \in T \to S = T).$$

If we break it down into its constituent parts, some subparts, such as $S \in T \to S = T$, or even $S = T$, represent sentences that can be true or false (depending on the values of the variables). If we break it down even further, say to the variables $S$ and $T$ themselves, they no longer represent sentences, but refer to the objects of study.

Conversely, building formulas is a two-step process. We first build what are called the *terms*—these are strings of symbols that refer to the objects under study. Using the terms we build the formulas, which represent open sentences and statements.

We build the terms recursively using the following steps.

(1) Variables are terms.

(2) Primitive constant symbols are terms.

(3) If $t_1, t_2, \ldots, t_n$ are previously constructed terms, and $f$ is a primitive $n$-ary function symbol, then $f(t_1, t_2, \ldots, t_n)$ is a term.

(Notice that rule (2) is a special case of (3) if we think of constants as 0-ary functions.) Terms are exactly those strings of symbols which can be built using only a finite number of steps according to these rules.

The reader should check that these three rules guarantee that the terms will always be referring to the objects under study.

**Example 7.5.** Using the signature $(0, S)$ given in Example 7.2, we see that each of $0, S(0), S(S(0)), \ldots$ is a term. If $v$ is a variable, then each of $v, S(v), S(S(v)), \ldots$ is a term. These together are the only possible types of terms in this signature. △

**Example 7.6.** If we take $\emptyset$ and $\omega$ as primitive constant symbols, and $\cup$ as a primitive binary function symbol, then $\emptyset \cup \omega$ is a term, as is $(\emptyset \cup \omega) \cup \omega$, and so is $(v_1 \cup \emptyset) \cup (\emptyset \cup v_2)$ for any two variables $v_1$ and $v_2$. However,

$$\emptyset \cup (\emptyset \cup (\emptyset \cup \cdots))$$

is not a term.

What are the terms in the signature $(\in)$, which has no constant symbols nor function symbols? △

**Example 7.7.** Some of the terms in the signature $(0, 1, <, +, \cdot)$ include

$$0,\ 1,\ 1 + (0 \cdot (1 + 1)),\ v_1 + v_2, \ldots.$$

To test your understanding, figure out which of the three steps for creating terms were used to construct these terms.

Any expression using the symbol $<$ will not be a term, since $<$ is a relation symbol, not a function symbol. △

**Warning 7.8.** Terms are not formulas. In particular, variables are not formulas in FOL, but they are in PL.

With terms now defined we can construct (well-formed) *formulas*, recursively, using the following steps.

(1) If $t_1, t_2, \ldots, t_n$ are terms and $R$ is a primitive $n$-ary relation symbol, then $R(t_1, t_2, \ldots, t_n)$ is a formula.
(2) If $t_1, t_2$ are terms, then $t_1 = t_2$ is a formula.
(3) If $\varphi_1, \varphi_2, \ldots, \varphi_n$ are formulas and if $C$ is a primitive $n$-ary logical connective, then $C(\varphi_1, \varphi_2, \ldots, \varphi_n)$ is a formula.
(4) If $v$ is a variable and $\varphi$ is a formula, then $\forall v\, \varphi$ and $\exists v\, \varphi$ are formulas.

The formulas created from rules (1) and (2) are the *atomic formulas*, so named for the fact that they are not built up from simpler formulas—just from the terms.

We define the *complexity* of FOL formulas in nearly the same way as we did for formulas of PL; we leave this as Exercise 7.5. (There is also a notion of "term complexity" if one needs it.)

**Example 7.9.** For this example, we use the signature $(0, S)$ of Example 7.2. Each of the following three strings of symbols are formulas (taking $x$ and $y$ to be variables):

$$x = 0,\ S(0) = S(S(0)),\quad \text{and}\quad \exists x\,(y = S(S(S(x)))).$$

However, none of the following three strings of symbols are formulas:

$$0,\ (x \wedge y) \to (S(0) = y),\quad \text{and}\quad \forall x\, x.$$

Can you explain why? Which ones are terms? △

**Example 7.10.** In this example we work with set theory using the single primitive binary relation symbol $\in$. In this signature,

$$(S = T) \wedge \forall S\,(x \in S)$$

is a formula. However,

$$\forall S\,(S \to S = \emptyset)$$

is not a formula, as the premise of the implication is not a formula. If we instead consider

$$\forall S\,(S \neq \omega \to S = \emptyset)$$

then this is a formula, or more correctly a shorthand *derived formula*, since it involves the nonprimitive symbols $\emptyset$, $\omega$, and $\neq$.

When you see $S = \emptyset$, this is shorthand for a "primitive" formula like

$$\forall x\, \neg(x \in S).$$

By saying "the formula $S = \emptyset$" one means "a formula represented by $S = \emptyset$". △

**Warning 7.11.** It isn't always obvious how to translate a shorthand derived formula into a primitive formula. Moreover, there could be different ways to do so. It helps to practice.

There are (at least) two strange behaviors that occur for formulas in FOL. First, some well-formed formulas do not look well-formed. Consider

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

which *is* a formula.

Second, in propositional logic a valid statement remains valid after replacing one variable with another one. In FOL there are some subtleties. The variable $x$ in the formula $\forall x\,(x = x)$ can be replaced by any other variable. However, now consider

$$\forall x\, \exists y\, \neg(x = y), \tag{7.12}$$

which intuitively expresses the idea "it is not the case that there is exactly one object". If we replace the variable $x$ here with $y$, we get the new formula

$$\forall y\, \exists y\, \neg(y = y),$$

which doesn't represent the same idea as (7.12).

In the following subsection we discuss "free" and "bound" variables, which will help us fix these types of issues.

## 7.C Free and bound variables

Informally, one might say that a quantifier $\exists x$ or $\forall x$ binds the variable $x$. However, this is overly simple for two reasons. First, in a formula such as $\forall x\, \varphi$, the variable $x$ might not occur anywhere in $\varphi$ (take $\varphi$ to be $y = y$, for instance). Second, even if $x$ appears in $\varphi$ it might already be bound by a different quantifier; this is the case for

$$\forall x\, \exists x\, (x \in y). \tag{7.13}$$

So we formally define *free variables* and *bound variables* recursively as follows. First, any variable that occurs in an atomic formula is free, and none are bound. Second, if $\varphi_1, \varphi_2, \ldots, \varphi_n$ are formulas and $C$ is an $n$-ary logical connective, then $C(\varphi_1, \varphi_2, \ldots, \varphi_n)$ has the same free and bound variables as occur among the $\varphi_i$'s. Third, if $\varphi$ is a formula, then the free variables in $\exists x\, \varphi$ and $\forall x\, \varphi$ are the free variables of $\varphi$ different than $x$, while the bound variables are the bound variables of $\varphi$ along with $x$ if it occurs anywhere in $\varphi$.

Thus in (7.13), the variable $x$ is bound (and it was first bound by the quantifier $\exists x$) whereas the variable $y$ is free. In formulas like

$$(x = y) \wedge \exists x\, (x \in S) \tag{7.14}$$

the variable $x$ occurs as *both* a free and bound variable.

A variable must actually occur, other than beside a quantifier, to be considered free or bound in a formula.

Hereafter, (except on rare occasions) we will avoid writing formulas $\exists x\, \varphi$ and $\forall x\, \varphi$ if $x$ is bound in $\varphi$ or doesn't appear at all. We will also avoid writing formulas with variables that are both bound and free. The fact that every formula is (semantically) equivalent to a formula avoiding these issues is left to the exercises in Section 9, after introducing the semantics of FOL in the next section. For instance, (7.13) will be equivalent to

$$\exists x\, (x \in y)$$

(with the first quantifier simply dropped) while (7.14) will be equivalent to

$$(x = y) \wedge \exists z\, (z \in S),$$

with the bound copies of $x$ replaced by a new variable $z$.

**Remark 7.15.** We promised in Section 1 to define *properties* and *parameters*. We are now able to do so. Properties are simply formulas. The formula itself describes the property we want to work with. (Thus, in FOL there are no *ineffable* properties; all properties have a description via a formula.) Parameters are simply free variables in the formula. There will be more to say about this in future sections. ▲

There is one further convention we will hereafter follow. If $\varphi$ is a formula, and the free variables of $\varphi$ are among $x_1, x_2, \ldots, x_n$, then we will write $\varphi(x_1, x_2, \ldots, x_n)$ in place of $\varphi$. (Why will there only be finitely many free variables?) If a formula has no free variables it is called a *statement*. (Are there any statements without quantifiers in FOL?)

## 7.D Exercises

**Exercise 7.1.** Describe the collection of terms when using the signature $(\in)$, and then describe the collection of atomic formulas.

**Exercise 7.2.** For each of the following strings of symbols, decide whether or not it is a formula in some signature, giving reasons for your answers. If it is a formula, determine if it is a statement, and also determine if it is shorthand.

(1) $\exists y\ (\neg(y = y) \wedge B(y))$, with $y$ a variable.
(2) $\emptyset \neq \omega$.
(3) $\forall a\, \exists b\, ((a < b) \wedge (b < c))$, with $a$, $b$, and $c$ variables.

**Exercise 7.3.** Translate the following sentences into formulas in some signature.

(1) Some dogs bark. (The objects of study could be dogs, and whether or not a dog is barking could be represented by a 1-ary relation symbol $B$.)
(2) There is nobody in my logic class older than the teacher. (Treat the teacher as a primitive constant.)
(3) If all the prime numbers greater than 2 are odd, then 2 is very odd.
(4) Every bully hates being bullied. (Treat bullying as a binary relation.)

**Exercise 7.4.** Define term complexity for terms of FOL, where the variables are exactly the terms with term complexity 0.

**Exercise 7.5.** Define complexity for formulas of FOL, with the formulas of complexity 0 being exactly the atomic formulas.

**Exercise 7.6.** Using the signature $(\in)$, rewrite the following shorthand formulas as primitive formulas.

(1) $\forall x \in T\, (x \subseteq y)$ (which means that every element of $T$ is a subset of $y$).
(2) $\exists x\, ((\omega \cap x = \emptyset) \wedge (x \neq \emptyset))$ (which means that there is a nonempty set containing no natural numbers).

The moral of this exercise is: To make formulas readable, use derived concepts.

**Exercise 7.7.** The language of FOL doesn't contain the "exists unique" quantifier, which is often written as $\exists!$. Find a formula in FOL that represents the shorthand formula $\exists! x\, \varphi$. (Hint: Write $\varphi(x)$ in place of $\varphi$ as needed.)

**Exercise 7.8.** For each of the following formulas, list the bound variables and also list the free variables.

(1) $\forall x\, (x < S(y))$.
(2) $(x < y) \wedge (\forall x\, (x < z))$.
(3) $\forall z\, \exists y\, (x = x)$.

# 8 Semantics and models of first order logic

To interpret a formula in propositional logic we use a row on a truth table. The different rows give different truth assignments to the variables. After the variables are assigned truth values then the formula also obtains a truth value.

Things are more complicated for first order logic. The quantifiers do a lot of the work of assigning meaning to the variables. Further, in FOL we allow open sentences, with variables not given any assignment at all.

In this section we make all of these ideas precise by introducing the standard semantics assigned for terms and formulas of first order logic. We also define models, which will be a fundamental concept throughout this book.

## 8.A Models

We don't think of the variables as representing arbitrary values from the set $\{\mathrm{T}, \mathrm{F}\}$, nor from any other *fixed* set. Rather, the domain of the variables is allowed to vary from one possible world to another. Sometimes variables represent numbers, and other times sets, and sometimes more exotic objects. The symbols in the signature also have varying meanings. A *model* (or an *interpretation*) is the way we give values to the nonlogical symbols. A model is also called an $\mathscr{L}$*-structure* when emphasizing its dependence on the signature of the language $\mathscr{L}$.

A model $\mathcal{M}$ consists of the following data. First, there is a nonempty meta set $M$, called the *domain* (or, more fully, the *domain of definition*). This is the set of objects from which the variables may take their values. Second, if $p$ is any primitive $n$-ary function (or $n$-ary relation, or constant) symbol in the signature, then we have a fixed $n$-ary function (respectively, $n$-ary relation, constant) $p_{\mathcal{M}}$ on the set $M$.

**Example 8.1.** Suppose we are working with the signature $(0, S)$, consisting of one constant symbol and one unary function symbol. We will construct a model $\mathcal{M}$ of this signature.

First we choose a nonempty domain, say

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

There are, of course, many other choices available for the domain. Any nonempty (meta) set would do.

Second, we must assign to the constant symbol 0 some value in $M$. One choice is

$$0_{\mathcal{M}} = a.$$

Any element of $M$ would do.

Finally, we need a unary function $S_{\mathcal{M}}$ on $M$; one option would be

$$S_{\mathcal{M}} = \{(a, a), (b, a), (c, b)\}.$$

Any (unary) function $M^1 \to M$ would do.

The model $\mathcal{M}$ consists of the triple of information $(M, 0_{\mathcal{M}}, S_{\mathcal{M}})$. △

**Example 8.2.** There is a *standard model* for the signature $(0, S)$. It is given by taking $M = \mathbb{N} = \{0, 1, 2, \ldots\}$, with $0_{\mathcal{M}} = 0 \in M$ and taking $S_{\mathcal{M}}$ to be the usual successor function on these (meta) natural numbers. △

**Example 8.3.** Define a model $\mathcal{M}$ for the signature $(0, S)$ by taking $M = \{0, 1\}$, with $0_{\mathcal{M}} = 1 \in M$ and with $S_{\mathcal{M}} = \mathrm{id}_M = \{(0, 0), (1, 1)\}$.

Notice that there are three different concepts with the symbol 0 in their names: the constant symbol 0 of the signature, the interpretation of the constant symbol $0_{\mathcal{M}} = 1 \in M$, and the meta number $0 \in M$. To avoid overload, we will often instead write $\mathbf{0}$ for the constant symbol in the signature. △

It is tempting to think that the only model of the signature $(\mathbf{0}, S)$ is the standard model given in Example 8.2. That is because the signature $(\mathbf{0}, S)$ was initially introduced in an attempt to talk more formally about the successor operation on the (meta) natural numbers. However, there are many other nonempty sets with a distinguished constant and distinguished unary function, which is all that is required.

**Example 8.4.** Consider the signature $(\in)$. To construct a model we must fix a (meta) set $M$, such as

$$M = \{1, 2, 3, 4\},$$

as well as a binary relation symbol $\in_{\mathcal{M}}$ on $M$, such as

$$\in_{\mathcal{M}} = \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}.$$ △

**Example 8.5.** Some models are very simple, even when the signature is not. Consider the signature $(0, 1, <, +, \cdot)$. Let $\mathcal{M}$ be a model whose domain is $M = \{a\}$, which has just one element. We are forced to assign the values of the constant symbols as $0_{\mathcal{M}} = a$ and $1_{\mathcal{M}} = a$. Similarly, the binary function symbols are assigned to the function

$$+_{\mathcal{M}} = \cdot_{\mathcal{M}} = \{((a, a), a)\}.$$

There are two possibilities for $<_{\mathcal{M}}$. Can we see what they are? △

## 8.B Variable assignments

Once we have a domain of definition it also makes sense to talk about assigning a fixed meaning to the variables. A *variable assignment* is a function $\nu$ that assigns to each variable $x$ a value $\nu(x) \in M$ (which is one reason why we need $M \neq \emptyset$).

**Advice 8.6.** Variable assignments are not considered part of the data that comprises a model. You can think of a variable assignment as an extra amount of information that may be tacked onto a model.

The possible worlds in PL are given by truth assignments to variables. However, in FOL you should think of the possible worlds as the models. Any variable assignment is additional information.

**Example 8.7.** Given the model of Example 8.1, one possible variable assignment $\nu$ would be to take $x \mapsto a$ for each variable $x$. There are other options. Can you think of a variable assignment that sends only some of the variables to $a$? △

## 8.C Interpreting terms

Let $\mathcal{M}$ be a model and let $\nu$ be a variable assignment. If $t$ is a term, we can now describe the interpretation $t_{(\mathcal{M},\nu)} \in M$ of the term in the given domain. Terms are always interpreted as elements of the domain of definition. The interpretation is built up in three stages, just as the terms are built up in three stages:

(1) The variables are interpreted according to the given variable assignment.
(2) The constant symbols are interpreted according to their chosen assignment.
(3) If $t_1, t_2, \ldots, t_n$ are previously constructed terms with corresponding interpretations $t_{1,(\mathcal{M},\nu)}, t_{2,(\mathcal{M},\nu)}, \ldots, t_{n,(\mathcal{M},\nu)} \in M$, and $f$ is an $n$-ary function symbol, then $f(t_1, t_2, \ldots, t_n)_{(\mathcal{M},\nu)}$ is simply $f_{\mathcal{M}}(t_{1,(\mathcal{M},\nu)}, t_{2,(\mathcal{M},\nu)}, \ldots, t_{n,(\mathcal{M},\nu)}) \in M$. In short, interpret $f$ according to its chosen assignment $f_{\mathcal{M}}$.

To make this clearer, we work out a couple examples.

**Example 8.8.** For this example, again work with the signature $(\mathbf{0}, S)$. Use the model described in Example 8.1 where we have

$$M = \{a, b, c\}, \quad \mathbf{0}_{\mathcal{M}} = a, \quad S_{\mathcal{M}} = \{(a, a), (b, a), (c, b)\}.$$

Let our variable assignment function $\nu$ be the constant function that takes each variable $x \mapsto c$.

Consider the term $S(S(\mathbf{0}))$. Its interpretation is built up in stages. First, we see $\mathbf{0}_{(\mathcal{M},\nu)} = \mathbf{0}_{\mathcal{M}} = a \in M$. Next,

$$S(\mathbf{0})_{(\mathcal{M},\nu)} = S_{\mathcal{M}}(\mathbf{0}_{(\mathcal{M},\nu)}) = S_{\mathcal{M}}(a) = a \in M.$$

Similarly, $S(S(\mathbf{0}))_{(\mathcal{M},\nu)} = a \in M$. In the computations we did not use $\nu$ since these terms involve no variables.

Next, if $x$ is a variable, we see that $x_{(\mathcal{M},\nu)} = \nu(x) = c$. Hence $S(x)_{(\mathcal{M},\nu)} = b$ and then $S(S(x))_{(\mathcal{M},\nu)} = a$.

If we use a new model $\mathcal{N}$, with the same domain and constant symbol assignment as $\mathcal{M}$, but with $S_{\mathcal{N}} = \{(a, b), (b, c), (c, a)\}$, then $\mathbf{0}_{\mathcal{N}} = a$, $S(\mathbf{0})_{\mathcal{N}} = b$, and finally $S(S(\mathbf{0}))_{\mathcal{N}} = c$. What is $S(S(S(\mathbf{0})))_{\mathcal{N}}$? (We have dropped the variable assignment function from the notation, as it is not needed if there are no variables.) △

Without a variable assignment, variables are allowed to range over the domain of definition, as in the next example.

**Example 8.9.** Consider the signature $(0, 1, <, +, \cdot)$ consisting of two constant symbols, one binary relation symbol, and two binary function symbols. Let $\mathcal{M}$ be the (standard) model where the domain of definition is $\mathbb{R}$ and these symbols are each given their usual meaning.

If $x$ and $y$ are variables, then $x + y$ is a term. Without a variable assignment, one should think of $(x+y)_{\mathcal{M}}$ as an element of $M$ that varies according to how $x$ and $y$ range over the domain. Once we assign values $\nu(x), \nu(y) \in \mathbb{R}$, then $(x+y)_{(\mathcal{M},\nu)} = \nu(x)+\nu(y)$ also obtains a fixed value. △

## 8.D Interpreting formulas

**Example 8.10.** Consider the formula $\forall x\,(x^2 \geq 0)$. In the signature $(0, 1, <, +, \cdot)$ this is really shorthand for

$$\forall x\,((0 < x \cdot x) \vee (0 = x \cdot x)).$$

This formula is interpreted as true in the standard model with domain $\mathbb{R}$.

The way we test its truth is to break it down into its component parts. The universal quantifier tells us to look at the subformula

$$(0 < x \cdot x) \vee (0 = x \cdot x)$$

and check that it is interpreted as true for every possible value assignment $\nu(x) \in \mathbb{R}$. The disjunction tells us to check that at least one of the two disjuncts is interpreted as true (still under every possible real value assignment for $x$), and so forth. We will make this process precise, below. △

With terms interpreted, we can now interpret formulas. Formulas represent assertions about the elements of the domain of definition. Thus, every formula is interpreted, after fixing the variables, as an element of $\{\mathrm{T}, \mathrm{F}\}$. We interpret formulas using four stages, just as the formulas themselves are defined in four stages.

We begin with the atomic formulas. If $t_1, t_2, \ldots, t_n$ are terms and $R$ is an $n$-ary relation symbol, then $R(t_1, t_2, \ldots, t_n)_{(\mathcal{M},\nu)}$ is true exactly when the statement $R_{\mathcal{M}}(t_{1,(\mathcal{M},\nu)}, t_{2,(\mathcal{M},\nu)}, \ldots, t_{n,(\mathcal{M},\nu)})$ is true. In other words, we think of this situation as follows: Replace the symbol $R$ with the actual relation $R_{\mathcal{M}}$ on the set $M$. Since $t_{1,(\mathcal{M},\nu)}, t_{2,(\mathcal{M},\nu)}, \ldots, t_{n,(\mathcal{M},\nu)}$ are elements of $M$, we may ask if they satisfy the relation $R_{\mathcal{M}}$ in their given order. If so, then we interpret the formula $R(t_1, t_2, \ldots, t_n)$ as true, otherwise it is false.

The other type of atomic formula is $t_1 = t_2$, for terms $t_1$ and $t_2$. Interpreting these formulas is even easier. We say $(t_1 = t_2)_{(\mathcal{M},\nu)}$ is true when $t_{1,(\mathcal{M},\nu)}$ is the same element as $t_{2,(\mathcal{M},\nu)}$ in the domain $M$, otherwise $(t_1 = t_2)_{(\mathcal{M},\nu)}$ is false.

Interpreting formulas built using logical connectives is no different than in propositional logic; we replace the connective symbol $C$ by the function $C_{\mathcal{M}}$ on $\{\mathrm{T}, \mathrm{F}\}$ that it represents. In other words $C(\varphi_1, \varphi_2, \ldots, \varphi_n)_{(\mathcal{M},\nu)}$ is true or false according to the value of $C_{\mathcal{M}}(\varphi_{1,(\mathcal{M},\nu)}, \varphi_{2,(\mathcal{M},\nu)}, \ldots, \varphi_{n,(\mathcal{M},\nu)})$, which can be determined by a truth table. Often the subscript on $C$ is dropped, since $C_{\mathcal{M}}$ is the same function in any model; but remember that the logical connective symbol and the function it represents are technically different things.

Finally, we need to interpret the quantified formulas $\forall x\,\varphi$ and $\exists x\,\varphi$, where $\varphi$ is a formula and $x$ is a variable. Working by induction on the complexity of formulas, we may assume that $\varphi$ has an interpretation in every model with a variable assignment. Given $m \in M$ let $\nu[x \mapsto m]$ be the same assignment as $\nu$ except that the assigned value of $x$ is now $m$. We say $(\exists x\,\varphi)_{(\mathcal{M},\nu)}$ is true if $\varphi_{(\mathcal{M},\nu[x\mapsto m])}$ is true for some $m \in M$, and false otherwise. Similarly, $(\forall x\,\varphi)_{(\mathcal{M},\nu)}$ is true if $\varphi_{(\mathcal{M},\nu[x\mapsto m])}$ is true for each $m \in M$, and false otherwise. (Thus, the truth value of quantified formulas depends on the truth values of simpler formulas under slightly modified variable assignments.)

If this four step definition is confusing, the reader is encouraged to look back at Example 8.10 and also to look forward to the next three examples.

**Example 8.11.** The formula $\forall x\, \neg(\mathbf{0} = S(x))$ is interpreted as true in the standard model of the signature $(\mathbf{0}, S)$, because 0 is not the successor of any natural number. In other words, the subformula $\neg(\mathbf{0} = S(x))$ is true under any variable assignment. However, without a variable assignment we think of that subformula as an open sentence without a truth value. △

**Example 8.12.** Given any formula $\varphi$ and any variable $x$, we will show that $\forall x\, \varphi$ and $\neg(\exists x\, \neg\varphi)$ always evaluate to the same truth value, i.e., they are *logically equivalent*. There are two cases to consider.

**Case 1**: First, suppose that $\varphi_{(\mathcal{M},\nu[x\mapsto m])}$ is true for each $m$ in the domain of definition. Then $(\forall x\, \varphi)_{(\mathcal{M},\nu)}$ is true, by definition. On the other hand

$$(\neg\varphi)_{(\mathcal{M},\nu[x\mapsto m])} = \neg_{\mathcal{M}}(\varphi_{(\mathcal{M},\nu[x\mapsto m])}) = \neg_{\mathcal{M}}(\mathrm{T}) = \mathrm{F}$$

for each $m$, and so $(\exists x\, \neg\varphi)_{(\mathcal{M},\nu)}$ is false. Hence its negation is true, as desired. (The subscripts on logical connectives are messy, so hereafter we will usually drop them.)

**Case 2**: Suppose $\varphi_{(\mathcal{M},\nu[x\mapsto m_0])}$ is false for some fixed $m_0$ in the domain of definition. Then $(\forall x\, \varphi)_{(\mathcal{M},\nu)}$ is false. Further,

$$(\neg\varphi)_{(\mathcal{M},\nu[x\mapsto m_0])} = \neg(\varphi_{(\mathcal{M},\nu[x\mapsto m_0])}) = \neg(\mathrm{F}) = \mathrm{T}$$

so $(\exists x\, \neg\varphi)_{(\mathcal{M},\nu)}$ is true. Its negation is then false, as desired. △

**Example 8.13.** Consider the signature $(\mathbf{0}, S)$ and the model $\mathcal{M}$ and variable assignment $\nu$ as in Example 8.8. As a quick exercise, to test your understanding, check if $\exists x\, \neg(S(\mathbf{0}) = x)$ is true or false when interpreted in $(\mathcal{M}, \nu)$. △

## 8.E Why do we care about models?

Models do much more than tell us how we can interpret formulas as true or false. They also place limitations on the consequences of formulas. A very old question, arising from the work of Euclid, was whether his important "parallel postulate" was a consequence of his previous axioms. This problem remained unsolved for hundreds of years, until a model was created for a type of geometry that satisfies the other axioms but not the parallel postulate.

Similarly, we can use models to show that some of the axioms of set theory are independent of one another. An easy case is worked out in Exercise 8.7.

Perhaps even more importantly, models aid intuition. Each model provides a new example (or counterexample) against which to test one's understanding of formal statements. The process of trying to create a model often reveals some subtleties in how different formal statements interact with one another. It can also reveal that a proof is missing ingredients or completely impossible (as in the case of trying to show that the parallel postulate follows from certain other axioms). For another example, see Exercise 8.8.

## 8.F Exercises

**Exercise 8.1.** Consider the signature $(\in)$, consisting of one binary relation symbol. Let $\mathcal{M}$ be the model with domain $\mathbb{N}$ where we interpret $\in$ as the relation $<$.

Determine the truth value of the (shorthand) formula

$$\forall S\,(\forall x \in S\,(x \neq S)),$$

as interpreted in this model.

**Exercise 8.2.** Recall that a statement is a formula with no free variables. Give examples for each of the following three phenomena:

(1) a statement that is not true in any model,
(2) a statement that is true in a model, but false in a different model, and
(3) a statement that is true in all models.

Try to make your examples as simple as possible.

**Exercise 8.3.** Show that for any formula $\varphi$ and any variable $x$, the formulas $\exists x\,\varphi$ and $\neg(\forall x\,\neg\varphi)$ always evaluate to the same truth value. (Note that Example 8.12 and this exercise show that, *a fortiori*, only one of the two quantifiers is needed when defining the language of first order logic.)

**Exercise 8.4.** Let $\varphi$ and $\psi$ be formulas. Prove that $(\forall x\,\varphi) \wedge (\forall x\,\psi)$ is logically equivalent to $\forall x\,(\varphi \wedge \psi)$.

**Exercise 8.5.** Consider the signature $(p, D)$ where $p$ is a constant symbol and $D$ is a binary relation symbol.

If we think of $D(x, y)$ as expressing "$x$ chooses to dance with $y$" and we think of the constant symbol $p$ as representing the "premier ballerina", then we can give informal meanings to formulas. Match each of the following formulas to the corresponding description.

(i) $\forall x\,\forall y\,D(x, y)$.
(ii) $\forall y\,D(p, y)$.
(iii) $\forall x\,D(x, p)$.
(iv) $\forall x\,D(x, x)$.
(v) $D(p, p)$.
(vi) $\exists x\,\forall y\,D(x, y)$.
(vii) $\exists y\,\forall x\,D(x, y)$.
(viii) $\forall y\,\exists x\,D(x, y)$.
(ix) $\forall x\,\exists y\,D(x, y)$.
(x) $\exists x\,\exists y\,D(x, y)$.
(xi) $\exists x\,D(x, p)$.
(xii) $\exists y\,D(p, y)$.

(a) Everyone has an admirer.
(b) Two to tango.
(c) Ballerina's solo routine.
(d) He weaves across the dance floor.
(e) Juliet found her Romeo.
(f) The danseur showed up.
(g) Stamina contest.
(h) Adored ballerina.
(i) A ballerina that loves all types of dancing.
(j) Mosh pit.
(k) Someone is quite popular.
(l) Dance instructor intervenes.

**Exercise 8.6.** Using the signature from the previous exercise, find a model $\mathcal{M}$ of the language where the formula

$$(\forall x\,\exists y\,\neg D(x, y)) \wedge (\forall x\,D(x, p))$$

evaluates as true, and where the cardinality of the domain of definition is as small as possible. (You might say that everyone has an enemy, but it isn't the ballerina.)

Next, find another model $\mathcal{N}$ where the previous formula still evaluates as true, but

$$\forall x \, \exists y \, (D(x,y) \wedge y \neq p)$$

also evaluates as true. (How might you informally describe what this new formula is asserting?)

**Exercise 8.7.** There are exactly two models of the signature $(\in)$ for the domain of definition $M = \{0\}$. For both models, test whether or not the axioms of extensionality, empty set, pairing, union, and power set hold. (In particular, these two models will demonstrate that some combinations of the axioms of set theory do not imply some of the others.)

**Exercise 8.8.** A basic fact used throughout mathematics is that the square of a nonzero real number is positive:

(0) $\forall x \, (x \neq 0 \rightarrow 0 < x \cdot x)$.

A standard proof of this fact uses the following three facts, which are even more basic. First, multiplying an inequality by a positive number preserves the inequality:

(1) $\forall a \, \forall b \, \forall c \, \big(((a < b) \wedge (0 < c)) \rightarrow a \cdot c < b \cdot c\big)$.

Second, multiplying an inequality by a negative number flips the inequality:

(2) $\forall a \, \forall b \, \forall c \, \big(((a < b) \wedge (c < 0)) \rightarrow b \cdot c < a \cdot c\big)$.

Third, multiplying any number by zero yields zero:

(3) $\forall y \, (y \cdot 0 = 0)$.

The proof of property (0) proceeds as follows. Let $x$ be a nonzero real number. First, if $0 < x$, then multiplying by $x$ preserves the inequality so $0x < x^2$, and hence $0 < x^2$, as desired. In the other case, when $x < 0$, then multiplying by $x$ flips the inequality, so we have $0x < x^2$, and hence $0 < x^2$ in this case as well.

Consider the signature $(0, <, \cdot)$, which consists of a constant symbol, a binary relation symbol, and a binary function symbol. Find a model of this signature where (1), (2), and (3) all evaluate as true, but (0) evaluates as false. It turns out that the proof of property (0) that was given above is using a (secret) fourth property; write that fourth property as a formula in this signature. (Hint: You might work backwards and first find that secret fourth property, as it may help you eliminate potential models. The fourth property cannot evaluate as true in the model you seek.)

**Exercise 8.9.** Do the following:

(1) Prove that if a term $t$ does not involve the variable $x$, then $t_{(\mathcal{M},\nu)}$ is independent of the value of $\nu(x)$. (Hint: Induct on the term complexity.)

(2) Prove that if a formula $\varphi$ does not involve the variable $x$, then $\forall x \, \varphi$ is logically equivalent to $\varphi$. (Hint: Prove that the evaluation $\varphi_{(\mathcal{M},\nu)}$ is independent of the value $\nu(x)$, by induction on the formula complexity of $\varphi$.)

# 9 Validity in first order logic

In the previous section we used an awkward phraseology to speak about "a formula $\varphi$ that evaluates as true in a model $\mathcal{M}$ under some variable assignment $\nu$". This concept can more succinctly be expressed using the idea of "satisfaction".

The symbol $\models$ that we used for the *satisfaction relation* from propositional logic also has a home in first order logic. We will use this symbol in multiple related ways. This makes it much easier to express many simple concepts without the awkward phraseology we used above.

We will also demonstrate how to use induction to prove "obvious" meta facts about FOL. This will help you, in the exercises, manipulate formulas.

## 9.A The satisfaction relation and validity

The relevant definitions in this section are the following, which build off of each other.

**Definition 9.1.** Let $\mathscr{L}$ be a language, let $\mathcal{M}$ be a model for that language, let $\nu$ be a variable assignment with values in the domain of the model, and let $\varphi$ be a FOL well-formed formula in the language $\mathscr{L}$.

(1) When $\varphi_{(\mathcal{M},\nu)}$ is true, then we write $(\mathcal{M},\nu) \models \varphi$. In this case we say "$\varphi$ is *satisfied* in $(\mathcal{M},\nu)$" or alternatively "$(\mathcal{M},\nu)$ *satisfies* (or *models*) $\varphi$".

(2) We write $\mathcal{M} \models \varphi$ when $(\mathcal{M},\nu) \models \varphi$ holds for all possible $\nu$. In this case, we say "$\varphi$ is *satisfied* in $\mathcal{M}$" or "$\mathcal{M}$ *satisfies* (or *models*) $\varphi$".

(3) If $\varphi$ is satisfied in some model of $\mathscr{L}$, we say $\varphi$ is *satisfiable*.

(4) If $\varphi$ is satisfied in every model of $\mathscr{L}$, we say that $\varphi$ is logically *valid* and in that case we write $\models \varphi$, or more expressively $\models_{\text{FOL}} \varphi$.

(5) If $X$ and $Y$ are sets of formulas, we write $X \models Y$ (or sometimes $\models_X Y$) to mean that the formulas in $Y$ are satisfied in any model where the formulas in $X$ are also satisfied (under every variable assignment).

**Example 9.2.** Consider a model $\mathcal{M}$ of the signature $(\in)$ where $M = \{0,1,2\}$ and $\in_{\mathcal{M}} = \{(0,1),(0,2)\}$. This model will not satisfy the axiom of extensionality (1.1), because this model tells us that even though 1 and 2 are different elements of the domain, the relation $\in_{\mathcal{M}}$ says that they both have only the single "element" 0.

More formally, let $\nu$ be a variable assignment where $S \mapsto 1$, $T \mapsto 2$, and $x \mapsto 0$. We see that $(x \in S)_{(\mathcal{M},\nu)}$ is true, since $0 \in_{\mathcal{M}} 1$. Similarly, $(x \in T)_{(\mathcal{M},\nu)}$ is true. Thus, $(x \in S \leftrightarrow x \in T)_{(\mathcal{M},\nu)}$ is true.

When we replace $\nu$ by the new variable assignment $\nu[x \mapsto 1]$, then it happens that $(x \in S)_{(\mathcal{M},\nu[x\mapsto 1])}$ and $(x \in T)_{(\mathcal{M},\nu[x\mapsto 1])}$ are both false, because $1 \notin_{\mathcal{M}} 1$ and $1 \notin_{\mathcal{M}} 2$. Hence, the biconditional $(x \in S \leftrightarrow x \in T)$ is satisfied in $(\mathcal{M},\nu[x \mapsto 1])$. Similarly, that biconditional is satisfied in $(\mathcal{M},\nu[x \mapsto 2])$.

Since that biconditional in satisfied no matter the value of $x$, this proves that $(\mathcal{M},\nu) \models \forall x\,(x \in S \leftrightarrow x \in T)$. On the other hand $(\mathcal{M},\nu) \not\models (S = T)$, since $1 \neq 2$. Thus, (1.1) is not satisfied in $(\mathcal{M},\nu)$, and hence not satisfied in $\mathcal{M}$. △

**Example 9.3.** Consider the formula $x = y$. When we think of it in terms of some model $\mathcal{M}$, it becomes an open sentence that has no truth value until we describe a variable assignment. If the domain of definition has two or more elements, there are some variable assignments that make the formula true and other assignments that make it false, so in that case $\mathcal{M} \not\models (x = y)$. However, when the domain of definition has exactly one element, then $\mathcal{M} \models (x = y)$. Thus, the formula $x = y$ is satisfiable, but not valid. △

**Example 9.4.** The formula $\forall x\,(x = x)$ is valid. For a formula to be valid, as this one is, it must be satisfied in *every* model. Valid formulas are rare and (dare we say) tautological. △

**Advice 9.5.** Avoid thinking of formulas in formal systems as true or false, even when they are statements. A formula only becomes true upon interpreting it in a model, according to a variable assignment (if needed).

**Warning 9.6.** Satisfaction may vary from model to model, but validity does not vary.

It is quite easy to determine whether a string of symbols in propositional logic is a well-formed formula and valid. To test validity, form a truth table with $2^n$ rows, where $n$ is the number of variables that appear in the formula, and then test every possibility. The process is completely finite.

Determining whether or not a string of symbols in first order logic is a well-formed formula is still a finite process (assuming $\mathscr{L}$ is easy to describe), but determining validity is difficult. There are infinitely many models, so we sometimes need another way to test for validity. One of the most successful methods is syntactic deduction, which we will introduce in Section 10.

## 9.B Proving obvious meta facts

Let $t$ be a term. Given a model $\mathcal{M}$ and two variable assignments $\nu_1$ and $\nu_2$, we might wonder when $t_{(\mathcal{M},\nu_1)}$ is the same element of the domain as $t_{(\mathcal{M},\nu_2)}$. It seems clear that this should be the case if $\nu_1$ and $\nu_2$ agree on the variables in $t$. It should not matter if $\nu_1$ and $\nu_2$ might disagree on variables that do not appear in $t$.

This fact may seem so obvious that no proof is required. However, it is valuable to look at this skeptically for a moment. Our formal method of interpreting terms inside models is complicated, and perhaps we unintentionally introduced a rule that disproves our intuition. Thus, this obvious fact deserves a proof. Seemingly, the only method we have available is to work through the recursive definition of interpreting terms. The following proof demonstrates how this is to be done.

**Metalemma 9.7.** *Let $t$ be a term in some language $\mathscr{L}$. Let $\mathcal{M}$ be a model for $\mathscr{L}$. If $\nu_1$ and $\nu_2$ are variable assignments that agree on the variables in $t$, then*

$$t_{(\mathcal{M},\nu_1)} = t_{(\mathcal{M},\nu_2)}.$$

*Proof.* We work by induction on the term complexity of $t$. For the base case, suppose $t$ is a variable. In this case $t_{(\mathcal{M},\nu_1)} = \nu_1(t)$ and $t_{(\mathcal{M},\nu_2)} = \nu_2(t)$. Since $\nu_1$ and $\nu_2$ agree on the variables in $t$ (which is just $t$ itself), we have the needed equality.

Next, suppose $t = f(t_1, \dots, t_n)$ for some $n$-ary function symbol $f$ and terms $t_1, \dots, t_n$ of smaller complexity. (This includes the case when $t$ is a constant symbol, if necessary, by taking $n = 0$.) Inductively, we may assume that $t_{i,(\mathcal{M},\nu_1)} = t_{i,(\mathcal{M},\nu_2)}$ for each $i$. We find that

$$t_{(\mathcal{M},\nu_1)} = f_{\mathcal{M}}(t_{1,(\mathcal{M},\nu_1)}, \dots, t_{n,(\mathcal{M},\nu_1)}) = f_{\mathcal{M}}(t_{1,(\mathcal{M},\nu_2)}, \dots, t_{n,(\mathcal{M},\nu_2)}) = t_{(\mathcal{M},\nu_2)}$$

as desired. □

This proof technique is extremely useful. A similar technique applies to formulas as well. For instance, we have:

**Metatheorem 9.8.** *Let $\varphi$ be a formula in some language $\mathscr{L}$. Let $\mathcal{M}$ be a model for $\mathscr{L}$. If $\nu_1$ and $\nu_2$ are variable assignments that agree on the free variables in $\varphi$, then*

$$\varphi_{(\mathcal{M},\nu_1)} = \varphi_{(\mathcal{M},\nu_2)}.$$

*Proof.* We induct on the complexity of $\varphi$.

**Case 1**: $\varphi$ is an atomic formula of the form $t_1 = t_2$ for some terms $t_1$ and $t_2$. All the variables in $\varphi$ are free, and so $\nu_1$ and $\nu_2$ agree on the variables in $t_1$ and $t_2$. Thus Metalemma 9.7 tells us that $t_{1,(\mathcal{M},\nu_1)}$ is the same element in the domain as $t_{1,(\mathcal{M},\nu_2)}$, and similarly $t_{2,(\mathcal{M},\nu_1)}$ is the same as $t_{2,(\mathcal{M},\nu_2)}$.

Now $\varphi_{(\mathcal{M},\nu_1)}$ is true if and only if (by the definition of how formulas are interpreted) the equality $t_{1,(\mathcal{M},\nu_1)} = t_{2,(\mathcal{M},\nu_1)}$ is true. This is equivalent, by what we said in the previous paragraph, to $t_{1,(\mathcal{M},\nu_2)} = t_{2,(\mathcal{M},\nu_2)}$ being true, which in turn is equivalent to $\varphi_{(\mathcal{M},\nu_2)}$ being true.

**Case 2**: $\varphi$ is an atomic formula of the form $R(t_1, \dots, t_n)$ for some $n$-ary relation symbol $R$, and terms $t_1, \dots, t_n$. By definition of interpreting formulas, we have

$$\varphi_{(\mathcal{M},\nu_1)} = R_{\mathcal{M}}(t_{1,(\mathcal{M},\nu_1)}, \dots, t_{n,(\mathcal{M},\nu_1)})$$

and similarly

$$\varphi_{(\mathcal{M},\nu_2)} = R_{\mathcal{M}}(t_{1,(\mathcal{M},\nu_2)}, \dots, t_{n,(\mathcal{M},\nu_2)})$$

Applying Metalemma 9.7, and again noting that $\nu_1$ and $\nu_2$ agree on all of the variables, we have $t_{i,(\mathcal{M},\nu_1)} = t_{i,(\mathcal{M},\nu_2)}$ for each $i$, and so we have the needed equality.

**Case 3**: $\varphi$ is of the form $\neg\psi$, for some formula $\psi$ of smaller complexity. We have

$$\varphi_{(\mathcal{M},\nu_1)} = \neg(\psi_{(\mathcal{M},\nu_1)}) = \neg(\psi_{(\mathcal{M},\nu_2)}) = \varphi_{(\mathcal{M},\nu_2)}$$

where the middle equality comes from applying our inductive hypothesis to $\psi$.

**Case 4**: $\varphi$ is of the form $\psi \wedge \theta$, for some formulas $\psi$ and $\theta$ of smaller complexity. A similar argument as in Case 3 works here. Moreover, now that we've handled $\neg$ and $\wedge$, all other logical connectives work too.

**Case 5**: $\varphi$ is of the form $\forall x\, \psi$, for some formula $\psi$ of smaller complexity. By our inductive assumption we have $\psi_{(\mathcal{M},\nu_1[x\mapsto m])} = \psi_{(\mathcal{M},\nu_2[x\mapsto m])}$ for any $m$ in our domain. This is true whether or not $x$ is free in $\psi$, since $\nu_1[x \mapsto m]$ and $\nu_2[x \mapsto m]$ agree on $x$ as well as all the free variables of $\varphi$. Now

$$\begin{aligned} \varphi_{(\mathcal{M},\nu_1)} \text{ is true} \quad &\text{iff} \quad \psi_{(\mathcal{M},\nu_1[x\mapsto m])} \text{ is true for each } m \text{ in the domain,} \\ &\text{iff} \quad \psi_{(\mathcal{M},\nu_2[x\mapsto m])} \text{ is true for each } m \text{ in the domain,} \\ &\text{iff} \quad \varphi_{(\mathcal{M},\nu_2)} \text{ is true.} \end{aligned}$$

Finally, by Exercise 8.3 we do not need to consider existential quantifiers, so we are done. □

**Corollary 9.9.** *Let $\varphi$ be a statement (i.e., a formula without any free variables) in some language $\mathscr{L}$. Let $\mathcal{M}$ be a model for $\mathscr{L}$. Then $\mathcal{M} \models \varphi$ if and only if $(\mathcal{M}, \nu_0) \models \varphi$ for some variable assignment $\nu_0$.*

*Proof.* The forward direction is true because $\mathcal{M} \models \varphi$ means that $(\mathcal{M}, \nu) \models \varphi$ holds for *every* variable assignment $\nu$. Conversely, notice that once $\varphi$ is satisfied for some variable assignment, it is satisfied for all variable assignments by Metatheorem 9.8. □

## 9.C Exercises

**Exercise 9.1.** Is $x = x$ satisfiable in some model? Is it valid? Justify your answers.

**Exercise 9.2.** Is the formula $\forall x\, \exists y\, (x \neq y)$ satisfiable in some model? Is it valid? Justify your answers.

**Exercise 9.3.** Is the axiom of empty set satisfiable in some model? Is it valid? Justify your answers.

**Exercise 9.4.** Let $\mathcal{M}$ be the model of the signature $(\in)$ where $M = \mathbb{N}$ and $\in_{\mathcal{M}}$ is the relation $<$ on $\mathbb{N}$. Which of the axioms of extensionality, empty set, pairing, union, infinity, and power set are satisfied in this model? (You may use an informal argument as was done in the first paragraph of Example 9.2. To gain some footing, you might start by figuring out the $\in_{\mathcal{M}}$-elements of 5.)

**Exercise 9.5.** Let $\mathcal{M}$ be a model, and let $\nu$ be a variable assignment. Let $s$ and $t$ be terms, and let $m = t_{(\mathcal{M},\nu)}$. Let $s' = s[x \mapsto t]$, which is the new term that results from replacing each occurrence of the variable $x$ in $s$ (if any) with $t$. Prove that $s_{(\mathcal{M},\nu[x\mapsto m])} = s'_{(\mathcal{M},\nu)}$. (Hint: Induction on the complexity of $s$.)

**Exercise 9.6.** Use the same notation as in the previous exercise. Suppose that $\varphi$ is a formula satisfying the following technical condition (that will come up again in the next section): If there is a free occurrence of $x$ in $\varphi$ that falls under the scope of a quantifier $\forall y$ or $\exists y$, then $y$ does not occur as a variable in $t$. Let $\varphi'$ be the formula that results from replacing each free occurrence of $x$ in $\varphi$ (if any) with $t$. Prove that $\varphi_{(\mathcal{M},\nu[x\mapsto m])} = \varphi'_{(\mathcal{M},\nu)}$. (Hint: Induction.)

In the next few exercises we prove that not only can we avoid quantifying over a bound variable (or a variable that doesn't appear) and avoid having a variable be both free and bound, we can also move all quantifiers to the very front of formulas.

**Exercise 9.7.** Let $\varphi(x)$ be a formula where $x$ is not bound and $y$ does not occur anywhere in $\varphi(x)$. Prove that $\forall x\, \varphi(x)$ is logically equivalent to $\forall y\, \varphi(y)$. (Hint: Use Exercises 9.5 and 9.6.)

**Exercise 9.8.** Let $\varphi$ be a formula where $x$ does not occur. Prove that $\forall x\, \varphi$ is logically equivalent to $\varphi$.

**Exercise 9.9** (Prenex normal form)**.** Prove that any formula $\psi$ is logically equivalent to a formula with the same free variables, of the form $Q_1x_1\, Q_2x_2\, \ldots\, Q_nx_n\, \varphi$, where

(1) the formula $\varphi$ has no quantifiers,
(2) $x_1, x_2, \ldots, x_n$ are distinct variables that appear in $\varphi$, and are not free in $\psi$, and
(3) $Q_1, Q_2, \ldots, Q_n \in \{\forall, \exists\}$.

(Hint: Treat the previous two exercises and Exercise 8.4 as lemmas. They tell you how to start moving quantifiers to the front of a formula. Perhaps you need a few more lemmas.)

**Exercise 9.10.** Find a prenex normal form for each the following formulas.

(1) $\exists y\, ((\forall x\, (x = y)) \to (y = z))$.
(2) $(\forall x\, (x \in y)) \vee (\forall x\, \neg(x \in y))$.
(3) $\forall x\, \exists x\, (x = x)$.

# 10 Deduction in first order logic

In the previous two sections we described how to interpret first order formulas inside models. In this section we will give an axiomatic system that allows us to deduce, syntactically, the semantically valid formulas.

## 10.A Syntactic deduction

Below, $\varphi$, $\psi$, and $\theta$ will represent arbitrary formulas in some FOL language $\mathscr{L}$. Also, $x$, $y$, and $z$ will be arbitrary variables and $t$ an arbitrary term. To simplify notation, when we write $\varphi[x \mapsto t]$ we mean that every free occurrence of $x$ in $\varphi$ is replaced by the term $t$, and after these replacements **no additional occurrences of variables become bound**. In other words, if any free occurrence of $x$ in $\varphi$ is under the scope of a quantifier $\forall v$, then the variable $v$ does not occur in $t$.

Our axiomatic system for FOL begins with assuming Mendelson's three axiom schemes for formulas. Thus, we assume:

(1) $\varphi \to (\psi \to \varphi)$,
(2) $(\varphi \to (\psi \to \theta)) \to ((\varphi \to \psi) \to (\varphi \to \theta))$, and
(3) $(\neg\varphi \to \neg\psi) \to ((\neg\varphi \to \psi) \to \varphi)$.

Next, to handle equality we add two more axiom schemes:

(4) $x = x$, and
(5) $(x = y) \to (\varphi[z \mapsto x] \to \varphi[z \mapsto y])$.

Finally, we need to handle the quantifiers so we assume two more schemes:

(6) $(\forall x\, \varphi) \to \varphi[x \mapsto t]$, and
(7) $(\forall x\, (\varphi \to \psi)) \to (\varphi \to (\forall x\, \psi))$ if $x$ is not free in $\varphi$.

There are two rules of inference. The first rule is modus ponens. The other rule is called *universal generalization*, which says that from $\varphi$ we may infer $\forall x\, \varphi$.

**Advice 10.1.** Think of universal generalization as expressing the fact that if $\varphi$ is satisfied in a model, then the value of $x$ is irrelevant for $\varphi$ in that model.

This axiomatic system assumes that we have rewritten all nonvariable logical symbols in terms of $=, \neg, \to, \forall$. Due to this fact, it is quite difficult (until one has built up a large number of metalemmas) to prove simple statements. However, this choice simplifies the proofs of certain metatheorems, as we will see shortly.

**Definition 10.2.** If $\varphi$ is a formula that can be deduced from our deduction system, then we write $\vdash \varphi$, read as "$\varphi$ is provable". The symbol $\vdash$ is called the *syntactic entailment* relation.

We leave it to the reader to differentiate between the two different entailment relations $\vdash_{\mathrm{PL}}$ and $\vdash_{\mathrm{FOL}}$, using subscripts as necessary.

A natural question is whether or not our deduction system for FOL is sound. In other words: Are the axioms valid? Do the rules of inference preserve validity? The answer to both of these questions is "yes".

**Metatheorem 10.3** (First order logic is sound)**.** *Suppose $\varphi$ is a well-formed formula. If $\vdash \varphi$, then $\models \varphi$.*

*Proof.* Let $(\mathcal{M}, \nu)$ be an arbitrary, fixed model with variable assignment function. Our job is to show that any formula that can be proved in our formal system will also be modeled in this arbitrary world $(\mathcal{M}, \nu)$. We start by showing that the axiom schemes are valid. We compute directly that

$$(\varphi \to (\psi \to \varphi))_{(\mathcal{M},\nu)} = \varphi_{(\mathcal{M},\nu)} \to (\psi_{(\mathcal{M},\nu)} \to \varphi_{(\mathcal{M},\nu)}),$$

and the truth value of the expression on the right is T, regardless of the truth values of $\varphi_{(\mathcal{M},\nu)}$ and $\psi_{(\mathcal{M},\nu)}$, by checking a truth table. (Remember that the symbols "→" on the two sides of the equality are different things. On the right side, change both "→" symbols to "$\to_{\mathcal{M}}$" to emphasize this difference, if necessary. Some authors use "⇒" in place of "$\to_{\mathcal{M}}$".)

Mendelson's other two axioms are similarly valid.

Next, as $x_{(\mathcal{M},\nu)}$ is always the same element as itself (in the domain of definition), axiom scheme (4) is valid.

Next we will verify axiom scheme (5). If $\nu(x) \neq \nu(y)$, then axiom (5) evaluates as true in $(\mathcal{M}, \nu)$, vacuously. So we may assume $\nu(x) = \nu(y)$. By Exercise 9.6, we have

$$\varphi[z \mapsto x]_{(\mathcal{M},\nu)} = \varphi_{(\mathcal{M},\nu[z \mapsto m])}$$

where $m = x_{(\mathcal{M},\nu)} = \nu(x)$. Since $\nu(x) = \nu(y)$, a symmetric computation shows that $\varphi[z \mapsto x]_{(\mathcal{M},\nu)} = \varphi[z \mapsto y]_{(\mathcal{M},\nu)}$, and so the final implication at the end of axiom scheme (5) evaluates as true since the premise and conclusion have the same truth values.

We leave the proof that axiom schemes (6) and (7) are valid as an exercise.

The lines of every formal proof are of three kinds: axioms, the result of modus ponens applied to two previous lines, or universal generalization applied to a previous line. Thus, to show every line in a proof is a valid statement, it now suffices to show that modus ponens and universal generalization take valid statements to valid statements. The proof for universal generalization is left as an exercise, while the proof for modus ponens follows from the same argument as in Metatheorem 6.12 using truth tables. □

The strange, technical condition used when defining the restricted substitutions $\varphi[x \mapsto t]$ is an unfortunate, but necessary, complication. Some beginning math students want to rename variables willy-nilly, but they learn quite quickly that if a variable name has already been used, then this behavior can sometimes lead to errors.

For example, suppose in a proof you have let $z$ be a variable, and you make the assertion "There is some value for the variable $y$ that does not equal $z$". At this point, it would be an error to replace the variable $z$ with the *bound* variable $y$. More formally, consider the case when $\varphi(z)$ is the formula

$$\exists y\, \neg(z = y) \tag{10.4}$$

and let $(\mathcal{M}, \nu)$ be the following structure: the domain of definition is $M = \{0, 1\}$ and every variable evaluates to 0. (Let any nonlogical symbols be evaluated however you want.)

The premise, $x = y$, of axiom scheme (5) evaluates as true in $(\mathcal{M}, \nu)$. The formula $\varphi(x)$ is $\exists y\, \neg(x = y)$, and it also evaluates as true. However, $\varphi(y)$ is $\exists y\, \neg(y = y)$, which evaluates as false. Thus axiom scheme (5) would be unsound if we allowed arbitrary substitutions.

**Advice 10.5.** Never plug bound variables into free variables.

## 10.B Additional notations

To generalize our notation, let $T$ be any collection of statements (i.e., formulas with no free variables). When we write $T \vdash \varphi$ or $\vdash_T \varphi$, we will mean that $\varphi$ can be proved syntactically, using our deduction system, under the additional assumption that we can use any of the formulas of $T$ in our proof.

Similarly, as we defined previously, $T \models \varphi$ or $\models_T \varphi$ means that $\varphi$ is satisfied in any interpretation in which every formula of $T$ is also satisfied. Thus, if we write $\mathcal{M} \models_T \varphi$, we will mean that $\mathcal{M}$ satisfies all the formulas in $T$ as well as $\varphi$.

We leave it as an easy exercise to extend Metatheorem 10.3 to show that FOL is *strongly sound*: $\vdash_T \varphi$ implies $\models_T \varphi$.

We will often think of $T$ as the collection of axioms for a *theory* we want to study. For instance, $T$ might be the collection of axioms for set theory that we introduced in Chapter I. Many of the remaining sections of this book will consist of specific first order theories that we wish to study.

**Example 10.6.** Let $\varphi$ be the formula $\forall x\, \exists S\, (x \in S)$ in the language $\mathscr{L}$ whose signature is $(\in)$. (If we interpret existential quantification as a shorthand, then $\varphi$ is really $\forall x\, \neg(\forall S\, \neg(x \in S))$.) This formula is not valid; there are models for the language $\mathscr{L}$ where the formula evaluates to false. Thus we write $\not\models \varphi$. Further, by the contrapositive of soundness, $\nvdash \varphi$. There is no way to prove $\varphi$ in our deduction system!

On the other hand, if $T$ is a set of statements containing the axiom of pairing (1.3), then $\varphi$ is valid in that theory, so $\models_T \varphi$. We also have $\vdash_T \varphi$; this is difficult to show without improving our deduction system as outlined in the next subsection. △

## 10.C Adding rules of inference

In this section we give examples of how to metatheoretically add rules of inference and axiom-like schemes to our deduction system.

Let's begin with axioms. By Example 6.8, we know $\vdash \varphi \to \varphi$ holds for any formula $\varphi$. Rather than write out the formal derivation of this fact each time we use it, it is much more convenient to treat Example 6.8 as a metalemma. For instance, suppose we wish to prove $(\neg\varphi \to \varphi) \to \varphi$. Our derivation might follow Example 6.9 and thus look like

| | | |
|---|---|---|
| 1. | $\neg\varphi \to \neg\varphi$ | Example 6.8. |
| 2. | $(\neg\varphi \to \neg\varphi) \to ((\neg\varphi \to \varphi) \to \varphi)$ | Axiom 3. |
| 3. | $(\neg\varphi \to \varphi) \to \varphi$ | MP lines 1, 2. |

This derivation is really a seven line derivation, with the first line referencing the five lines from Example 6.8 (with $\neg\varphi$ in place of $A$).

Another metalemma that the reader might find useful is

$$(A \to B) \to ((B \to C) \to (A \to C));$$

see Exercise 10.3.

In practice, mathematicians do not *merely* build from one result to the next, creating an ever increasing list of lemmas. They also have a cache of useful arguments or methods that can be applied in different settings. Thus, it can be quite powerful to add extra inference rules to our proof system. One of the most useful is the following:

**Metatheorem 10.7** (Deduction theorem)**.** *Let $T$ be a set of formulas and let $\varphi$ and $\psi$ be formulas. If $T \cup \{\varphi\} \vdash \psi$, and no application of universal generalization was used to quantify over any of the free variables of $\varphi$, then $T \vdash \varphi \to \psi$.*

*Proof.* Fix a derivation of $\psi$ where each line is either an axiom, a formula from $T$, the formula $\varphi$, a consequence of modus ponens from two of the earlier lines, or a consequence of universal generalization from an earlier line (subject to the assumed restriction). For each integer $1 \leq i \leq n$, let $\theta_i$ be the $i$th line of this proof. Note that the last line of the proof, $\theta_n$, is just $\psi$.

We will prove that $T \vdash \varphi \to \theta_i$ for each integer $1 \leq i \leq n$, which will give us more than what we need. We may assume recursively that the claim is true for all smaller indexes.

First consider the case when $\theta_i$ is an axiom or an element of $T$. Note that $\theta_i \to (\varphi \to \theta_i)$ is an instance of axiom (1). So modus ponens yields $T \vdash \varphi \to \theta_i$.

Second, $\varphi \to \varphi$ is proven in Example 6.8 so if $\theta_i$ is $\varphi$, then we have $T \vdash \varphi \to \theta_i$.

Third, suppose that $\theta_i$ is the result of applying modus ponens to two previous lines. Thus $\theta_j$ and $\theta_j \to \theta_i$ are two previous lines, and so we have $T \vdash \varphi \to \theta_j$ and $T \vdash \varphi \to (\theta_j \to \theta_i)$. By axiom (2) we have

$$T \vdash (\varphi \to (\theta_j \to \theta_i)) \to ((\varphi \to \theta_j) \to (\varphi \to \theta_i)).$$

Applying modus ponens to the previous two proven formulas, we get

$$T \vdash (\varphi \to \theta_j) \to (\varphi \to \theta_i).$$

Applying modus ponens once again, we conclude $T \vdash \varphi \to \theta_i$, as desired.

Finally, suppose $\theta_i$ is $\forall v\, \theta_j$, where $v$ is not a free variable in $\varphi$ and $j < i$. (In other words, we obtain $\theta_i$ by applying universal generalization to the $j$th line of the proof, under the assumed restriction on $v$.) We know $T \vdash \varphi \to \theta_j$. Applying universal generalization we get

$$T \vdash \forall v\, (\varphi \to \theta_j).$$

Since $v$ is not free in $\varphi$, axiom (7) yields

$$T \vdash (\forall v\, (\varphi \to \theta_j)) \to (\varphi \to (\forall v\, \theta_j)).$$

By modus ponens, $T \vdash \varphi \to \forall v\, \theta_j$. □

The deduction theorem is used by working mathematicians all the time. The following is a classic example.

**Example 10.8.** Recall that $A \subseteq B$ is shorthand for the formula $\forall x\, (x \in A \to x \in B)$.

Consider the standard structure of a proof that $A \subseteq B$. The first sentence is usually "Assume $x \in A$". From that assumption we deduce $x \in B$. In other words, we show $x \in A \vdash x \in B$. If we do this without quantifying over $x$ (indeed, you shouldn't quantify over $x$!), then by the deduction theorem we have deduced the more complicated formula $x \in A \to x \in B$. The last sentence of the proof is usually something like "Since $x$ was arbitrary, we see that $A \subseteq B$", which is just using universal generalization to conclude $\vdash \forall x\, (x \in A \to x \in B)$. △

Here is an example of making the deduction theorem do a lot of the work in deducing formulas.

**Example 10.9.** Let $\varphi$, $\psi$, and $\theta$ be formulas. We will show that

$$\varphi \to (\psi \to \theta) \vdash \psi \to (\varphi \to \theta).$$

By the deduction theorem, it suffices to show—without universal generalization—that

$$\{\psi, \varphi \to (\psi \to \theta)\} \vdash \varphi \to \theta.$$

By another application of the deduction theorem, it suffices to show that

$$\{\varphi, \psi, \varphi \to (\psi \to \theta)\} \vdash \theta$$

without using universal generalization. This we can do with the following five line proof

| | | |
|---|---|---|
| 1. | $\varphi$ | Assumption. |
| 2. | $\varphi \to (\psi \to \theta)$ | Assumption. |
| 3. | $\psi \to \theta$ | MP lines 1, 2. |
| 4. | $\psi$ | Assumption. |
| 5. | $\theta$ | MP lines 4, 3. |

How many lines does it take to prove the claimed result without using the deduction theorem? △

The restriction in the deduction theorem about not quantifying over free variables of $\varphi$ is something that we usually take for granted—we should avoid *simultaneously* treating a variable as both free and bound. Consider taking $\varphi$ to be $x = y$ and $\psi$ to be $\forall x\,(x = y)$. Notice that we are treating $x$ as free in $\varphi$, but bound in $\psi$. We see that $\varphi \vdash \varphi$. Thus, by universal generalization $\varphi \vdash \psi$. However, $\vdash \varphi \to \psi$ is false; to see this, from soundness it suffices to find a model $\mathcal{M}$ where $\mathcal{M} \not\models \varphi \to \psi$. The same model that appeared directly after (10.4), with two distinct elements, works here too.

There is one more thing we should point out about this example. Notice that the metastatement

$$\vdash \varphi \text{ implies } \vdash \psi \tag{10.10}$$

is true, by an application of universal generalization, but the metastatement

$$\vdash \varphi \to \psi \tag{10.11}$$

is false by the soundness theorem, since $\varphi \to \psi$ is not valid.

**Warning 10.12.** Just because you can (informally) justify the formal provability of $\psi$ under the assumption that $\varphi$ is formally provable, this does not mean that you can formally prove that $\psi$ follows from $\varphi$. (Hint: $\varphi$ might not be formally provable in the first place.)

Replacing $\vdash$ by $\models$ in (10.10), it becomes "if $\varphi_{(\mathcal{M},\nu)}$ is true for every model with variable assignment $(\mathcal{M},\nu)$, then so is $\psi_{(\mathcal{M},\nu)}$". Replacing $\vdash$ by $\models$ in (10.11), it becomes "in every model with variable assignment $(\mathcal{M},\nu)$, if $\varphi_{(\mathcal{M},\nu)}$ is true then so is $\psi_{(\mathcal{M},\nu)}$". These are different sentences. (Putting it in other words, the formula

$$((\forall x\, A) \to (\forall x\, B)) \to \forall x\,(A \to B)$$

is not valid.)

## 10.D Semantic completeness

The converse of Metatheorem 10.3 is true. It was first proven by Gödel in 1929, using a different axiomatization of FOL.

**Metatheorem 10.13** (Gödel's (strong semantic) completeness theorem)**.** *Let $T$ be any set of statements and let $\varphi$ be an arbitrary formula, all in a FOL language $\mathscr{L}$. If $\models_T \varphi$, then $\vdash_T \varphi$.*

We will not prove this metatheorem here, but mention it only to further justify our use of deduction. Hopefully this metatheorem comes as something of a shock; there is no *a priori* reason to suppose that our deduction system has enough axioms or rules of inference to allow us to deduce all valid formulas. Indeed, there is no *a priori* reason to suppose that there is even a finite (or countable) list of rules and axiom schemes that would suffice.

## 10.E Second order logic

Second order logic allows quantification not only over the domain of definition, but over other subsets of the domain. It is thus more expressive than FOL, but also more complicated. There are some mathematical concepts that can only be expressed in a second order (or higher) language; we will point them out as we meet them. On the other hand, the metatheory of second order logic is not as clean. For instance, the semantic completeness theorem fails (for the standard semantics of second order logic).

## 10.F Exercises

**Exercise 10.1.** Assume $T_1 \subseteq T_2$ are sets of formulas and assume $\varphi$ is a formula. Show that if $T_1 \vdash \varphi$ then $T_2 \vdash \varphi$.

**Exercise 10.2.** Give an example of some language $\mathscr{L}$ with a predicate $\varphi(x)$ and a term $t$ such that $(\forall x\, \varphi(x)) \to \varphi(t)$ is not valid, thus showing that the substitution restriction in axiom (6) cannot be dropped. (By a *predicate*, here we mean a formula with one free variable.)

**Exercise 10.3.** Show that $\vdash_{\mathrm{FOL}} (\varphi \to \psi) \to ((\psi \to \theta) \to (\varphi \to \theta))$ for all formulas $\varphi, \psi, \theta$, (Hint: Use the deduction theorem.) Next, show that $\neg\psi \to \neg\varphi \vdash_{\mathrm{FOL}} \varphi \to \psi$ for all formulas $\varphi, \psi$. (How hard is it to write down the full formal derivations?)

**Exercise 10.4.** Prove that equality is an equivalence relation on terms. In other words, for any terms $r, s, t$ show that

(1) $\vdash r = r$,
(2) $\vdash (r = s) \to (s = r)$, and
(3) $\vdash (r = s) \to ((s = t) \to (r = t))$.

(Warning: Part (1) is not merely an application of axiom scheme (4), since that scheme is talking about a variable $x$, rather than an arbitrary term. You will need to use axiom (4) in conjugation with axiom (6) to introduce an arbitrary term.)

**Exercise 10.5.** Finish the proof of Metatheorem 10.3, while additionally improving it to give *strong soundness* for first order logic.

**Exercise 10.6.** Prove the following slight generalization of the deduction theorem: If $T \cup \{\varphi\} \vdash \psi$, and no application of universal generalization was used to quantify over any of the free variables of $\varphi$ *when deducing a formula that depends on* $\varphi$, then $T \vdash \varphi \to \psi$.

(To put this improvement into colloquial terms: If you use a lemma that only assumes $T$, but not $\varphi$, then you don't need to double-check whether or not in the proof of the lemma you bound a variable that you are now treating as free.)

**Exercise 10.7.** (This is a very important exercise.) Prove the *compactness theorem*: Let $T$ be a set of statements. Assuming every finite subset of $T$ is satisfiable, then $T$ is satisfiable. (Hint: Begin by assuming, contrapositively, that $T$ is not satisfiable, and get $\models_T \bot$. Then apply the completeness theorem.)

# Chapter III

# Order Properties and Set Theory

*One of the advantages of being disorderly is that one is constantly making exciting discoveries.* A. A. Milne

In this chapter we apply the language of first order logic to study orderings. Ordered structures are ubiquitous. The natural numbers are ordered, and this gives rise to an ordering on the integers, rationals, and reals. The order on the natural numbers also induces a partial order on formulas, which we have called complexity. We will study other generalizations, specializations, and modifications of orderings, in a wide array of contexts.

Most importantly, we introduce the idea of an ordinal as a way to measure well-ordered sets. Ordinals also provide a means of doing induction in more general contexts, and are a key component in the theory of sets.

# 11 Order properties on binary relations

At its heart, an order is a binary relation. Thus, to speak of orders in FOL we first need our language to have a binary relation symbol. We will use the symbol $\prec$.

Without assuming anything more, the relation symbol $\prec$ can have *many* different interpretations. For instance, we could take $\mathcal{M}$ to be the interpretation whose domain is $\mathbb{N}$ and where $\prec_{\mathcal{M}}$ is just the usual order relation $<$ on $\mathbb{N}$. We could instead use the interpretation $\mathcal{M}^{\text{op}}$, whose domain is $\mathbb{N}$ but where $\prec_{\mathcal{M}^{\text{op}}}$ is the inverse order relation $>$ on $\mathbb{N}$. Or we could just take a random domain with a random binary operation that has nothing to do with order. So our next task is to describe properties for $\prec$ that will make it an order relation, like $<$ on $\mathbb{N}$.

## 11.A Irreflexive versus reflexive

There are two very closely related relations on $\mathbb{N}$, namely: the strict ordering $<$ and the nonstrict ordering $\leq$. The difference between these two relations is encapsulated by the following two properties.

**Definition 11.1.** Let $R$ be a binary relation on a set $M$.
We say $R$ is *reflexive* if $\forall x \in M\,(xRx)$ holds.
We say $R$ is *irreflexive* if $\forall x \in M\,\neg(xRx)$ holds.

Notice that any irreflexive relation on a nonempty set $M$ is not reflexive. The converse is false, since a relation may be "partially" reflexive, such as the relation $R = \{(3,3)\}$ on the set $M = \{1,2,3\}$. Thus, it takes more to be irreflexive than to merely be nonreflexive.

To every binary relation $R$, there is a corresponding best reflexive relation, and similarly a best irreflexive relation, as described in the following proposition.

**Proposition 11.2.** *Let $R$ be a relation on a set $M$. There is a smallest reflexive relation $R^{\text{ref}} \supseteq R$ on $M$ called the* reflexive closure *of $R$. Similarly, there is a largest irreflexive relation $R^{\text{irref}} \subseteq R$ called the* irreflexive reduction *of $R$.*

*Proof.* We will prove the first half, leaving the second half as an exercise. Consider the binary operation $R' = R \cup \{(x,x) : x \in M\}$. It contains $R$, and it is clearly reflexive. Thus, it suffices to show that $R'$ is contained in every other reflexive relation on $M$ containing $R$; but that is also clear.

We will now give an alternative proof since the idea it uses can be applied in many different settings. Let $P$ be the set of all reflexive relations on $M$ that contain $R$. Take $R' = \bigcap_{S \in P} S$. Since $M \times M \in P$, we have $R' \subseteq M \times M$, and so $R'$ is a binary relation on $M$. Next, $(x,x) \in S$ for each $x \in M$ and each $S \in P$, and hence $(x,x) \in R'$. Similarly $R \subseteq R'$. So $R'$ is a reflexive relation on $M$ containing $R$, and contained in every other such relation $S \in P$, as desired. □

In the first proof, we could have denoted the equality relation $\{(x, x) : x \in M\}$ on $M$ as $=_M$, and thus $R^{\text{ref}} = R \cup (=_M)$. However, for readability reasons it is usually best to avoid using the equality symbol to denote a set.

**Example 11.3.** Let $S = \{1, 2, 3\}$, and let

$$M = \mathscr{P}(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, S\}.$$

The subset relation $\subseteq$ on $M$ is reflexive. Its irreflexive reduction is the proper subset relation, so $\subseteq^{\text{irref}} = \subsetneq$. Vice versa, the reflexive closure of the proper subset relation $\subsetneq$ is the usual subset relation, so $\subsetneq^{\text{ref}} = \subseteq$. △

There are some straightforward properties of the reflexive closure and irreflexive reduction; we leave the proof of the following proposition as an exercise.

**Proposition 11.4.** *Let $R, S$ be any two relations on $M$.*
(1) $R \subseteq R^{\text{ref}}$ *and* $R^{\text{irref}} \subseteq R$.
(2) *If* $R \subseteq S$, *then* $R^{\text{ref}} \subseteq S^{\text{ref}}$ *and* $R^{\text{irref}} \subseteq S^{\text{irref}}$.
(3) $(R^{\text{ref}})^{\text{ref}} = R^{\text{ref}}$ *and* $(R^{\text{irref}})^{\text{irref}} = R^{\text{irref}}$.
(4) $(R^{\text{ref}})^{\text{irref}} = R^{\text{irref}}$ *and* $(R^{\text{irref}})^{\text{ref}} = R^{\text{ref}}$.

**Question**: Is it better to prefer orders that are irreflexive or ones that are reflexive?
**Answer**: When talking about inequalities, when possible it is better to use $<$ than $\leq$, because it gives us more information (namely, that the inequality is strict). Also, another minor benefit to irreflexivity is that it depends only on the ordered pairs in the relation, and not on the set $M$. On the other hand, we will see in Section 12 a reason to favor reflexivity over irreflexivity in some situations.

Of course, in cases when it is unknown if an inequality is strict or not, then we prefer to use $\leq$ instead of $<$. However, this does not necessitate the existence of two binary relation symbols in our language, since we can *define* the (nonprimitive) reflexive closure symbol $\preceq$ of a binary relation symbol $\prec$ by a first order formula:

$$x \preceq y \text{ is shorthand for } (x \prec y) \vee (x = y).$$

## 11.B Symmetry, asymmetry, and antisymmetry

An essential property of orders is that there is a notion of "smallness". Intuitively, if $a$ is (strictly) smaller than $b$, we cannot have $b$ smaller than $a$. This leads us to study the following three properties of binary relations.

**Definition 11.5.** Let $R$ be a binary relation on a set $M$.
We say $R$ is *symmetric* if $\forall x, y \in M\,(xRy \to yRx)$ holds.
We say $R$ is *asymmetric* if $\forall x, y \in M\,(xRy \to \neg(yRx))$ holds.
We say $R$ is *antisymmetric* if $\forall x, y \in M\,((xRy \wedge yRx) \to (x = y))$ holds.

Thus, for an irreflexive *order* relation we will want it to be asymmetric. If instead our order relation is reflexive, then asymmetry should be replaced by antisymmetry.

**Example 11.6.** Let $S = \{1,2,3\}$, and let

$$M = \mathscr{P}(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, S\},$$

as in Example 11.3. The reflexive relation $\subseteq$ is antisymmetric, but not asymmetric or symmetric. The irreflexive relation $\subsetneq$ is asymmetric and (vacuously) antisymmetric, but it is not symmetric. △

There are some nice relationships between asymmetry and antisymmetry as expressed in the following lemma; we leave the proof as an exercise.

**Lemma 11.7.** *Let $R$ be a relation.*
(1) *$R$ is antisymmetric if and only if $R^{\mathrm{ref}}$ is antisymmetric, if and only if $R^{\mathrm{irref}}$ is antisymmetric.*
(2) *$R$ is asymmetric if and only if $R$ is antisymmetric and irreflexive.*

As with the reflexive property, there is always a *symmetric closure* of a relation $R$; it is just $R^{\mathrm{sym}} = R \cup R^{-1}$. However, there is no best way to asymmetrize an arbitrary relation. Indeed, consider the case when $R = \{(1,2),(2,1)\}$. This relation is not asymmetric, and no superset will be asymmetric either. Although $R$ becomes asymmetric when we drop either ordered pair, there is no reason to prefer dropping one pair over the other. (Is there a "best" way to make $R$ into an antisymmetric relation?) The problematic nature of this example boils down to the fact that $R$ is not even antisymmetric. In the contrary case there are no problems.

**Proposition 11.8.** *If $R$ is an antisymmetric relation, then $R^{\mathrm{irref}}$ is the largest asymmetric relation contained in $R$.*

*Proof.* That $R^{\mathrm{irref}}$ is asymmetric comes from Lemma 11.7. Any asymmetric relation $S$ is irreflexive, and hence if $S \subseteq R$ then $S = S^{\mathrm{irref}} \subseteq R^{\mathrm{irref}}$ by Proposition 11.4(2). □

## 11.C Transitivity

There is another intuitive condition that "smallness" should possess. If $a$ is smaller than $b$, and $b$ is smaller than $c$, then certainly $a$ is (much) smaller than $c$.

**Definition 11.9.** Let $R$ be a binary relation on a set $M$.
We say $R$ is *transitive* if $\forall x,y,z \in M\ ((xRy \wedge yRz) \to xRz)$.

**Example 11.10.** Let $S = \{1,2,3\}$, and let

$$M = \mathscr{P}(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, S\},$$

as in Example 11.3. Both the subset and proper subset relations, $\subseteq$ and $\subsetneq$, are transitive. △

Just as for the reflexive and symmetric properties, there is a closure operation for the transitive property.

**Proposition 11.11.** *Let $R$ be a binary relation. There is a smallest transitive relation $R^{\text{tran}} \supseteq R$, called the* transitive closure.

*Proof sketch.* The intersection argument used in the proof of Proposition 11.2 works here too. Alternatively, define $R^1 = R$ and recursively take $R^{n+1} = R \circ R^n$, and then show that $\bigcup_{n=1}^{\infty} R^n$ satisfies the conditions of the transitive closure. □

**Corollary 11.12.** *If $R$ is a binary relation on a set $M$, there is a smallest equivalence relation containing $R$.*

*Proof.* Consider $((R^{\text{ref}})^{\text{sym}})^{\text{tran}}$, which is transitive and contains $R$. By Exercise 11.11, this relation is also reflexive and symmetric. Any other equivalence relation $S \supseteq R$ is reflexive, and hence $S = S^{\text{ref}} \supseteq R^{\text{ref}}$ by Proposition 11.4(2). Similarly, $S = S^{\text{sym}} \supseteq (R^{\text{ref}})^{\text{sym}}$, and then $S = S^{\text{tran}} \supseteq ((R^{\text{ref}})^{\text{sym}})^{\text{tran}}$, by Exercise 11.10. □

**Example 11.13.** Let $M = \{1,2,3,4\}$, and let $R = \{(1,2),(4,2)\}$. The reflexive closure is

$$R^{\text{ref}} = \{(1,1),(2,2),(3,3),(4,4),(1,2),(4,2)\}.$$

The symmetric closure of this relation is

$$(R^{\text{ref}})^{\text{sym}} = \{(1,1),(2,2),(3,3),(4,4),(1,2),(2,1),(4,2),(2,4)\}.$$

which is still reflexive. Finally, its transitive closure is

$$((R^{\text{ref}})^{\text{sym}})^{\text{tran}} = \{(1,1),(2,2),(3,3),(4,4),(1,2),(2,1),(4,2),(2,4),(1,4),(4,1)\},$$

which is still symmetric and reflexive, and hence this is an equivalence relation. △

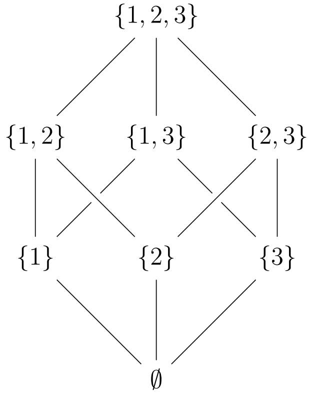
## 11.D Partial Orders

**Definition 11.14.** A binary relation $<$ on a set $M$ that is irreflexive, asymmetric, and transitive is said to be a (strict) *partial ordering* of $M$, or we say that $<$ *partially orders* $M$. We also say that $(M,<)$ (or sometimes just $M$, when $<$ is understood from context) is a *poset*.

The reason for the word “partial” is that our ordering does not necessarily force two elements to be related. In other words, we do not assume that given two distinct elements, one of them has to be bigger than the other. We will cover “total” orders in the next section, but the notion of a partial order is quite natural and much more applicable. For instance, if you are ranking your favorite foods, and you get to “eggplant” and “liver”, then you might not put one above the other (especially if you have never had either one); then again, you might like them both very much and cannot decide which to place above the other.

There are many other examples of posets. The following is perhaps the most important example we will encounter.

**Example 11.15.** Given any set $S$, we can partially order $M = \mathscr{P}(S)$ using the relation $\subsetneq$. This relation is irreflexive, asymmetric, and transitive. Taking $S = \{1, 2, 3\}$ as in previous examples, then we can picture this partially ordered set as follows:



where one set is properly contained in another exactly when you can follow lines upward from one to the other. Here, there is no relationship between $\{1, 2\}$ and $\{3\}$; we do not think of either set as “smaller” than the other; they are incomparable.

From the proper subset relation $\subsetneq$, we can construct the (nonproper) subset relation $\subseteq$, and vice versa. Thus, both of the relations $\subsetneq$ and $\subseteq$ give us this same picture; the only difference is that the proper subset relation $\subsetneq$ excludes a set being related to itself, whereas $\subseteq$ requires it. We can easily pass between these two different relations, treating them as giving us the same picture. △

The theory of posets is a first order theory. Consider any language containing a binary relation symbol $\prec$. Take POSET to be the set of statements asserting that $\prec$ is irreflexive, asymmetric, and transitive. In other words, we take POSET equal to

$$\{\forall x\, \neg(x \prec x), \forall x\, \forall y\, (x \prec y \to \neg(y \prec x)), \forall x\, \forall y\, \forall z\, ((x \prec y \wedge y \prec z) \to x \prec z)\}.$$

These three statements axiomatize what it means to have a poset. Thus we know by Metatheorem 10.13—Gödel’s completeness theorem—that any first order statement about posets that is valid (i.e., satisfied in all models) will have a proof whose only axioms are those of first order logic together with these three axioms of POSET. We write $\vdash_{\text{POSET}} \varphi$ when $\varphi$ has such a proof.

It is interesting to note that transitivity together with irreflexivity implies asymmetry, and asymmetry implies irreflexivity! So, we could have simplified our axiomatization of POSET (by dropping either the first or second axiom) as well as simplified the definition of partial orders.

Given any partial order $<$ on a set $M$, then $<^{\text{ref}}$ will be a reflexive, antisymmetric, and transitive relation, which we call the nonstrict version, $\leq$, of the partial order. Conversely, given any reflexive, antisymmetric, transitive relation $\leq$, then $\leq^{\text{irref}}$ is a strict partial order that we denote as $<$. Moreover, Proposition 11.4(4) tells us that these two constructions are inverses of each other. Thus, the reflexive closure and the irreflexive reduction provide a natural means for passing between $\prec$ and $\preceq$, the strict and nonstrict versions of a partial order relation on a set. (Sometimes we use the notation $\precneq$ for the strict partial order.)

If the words "strict" and "nonstrict" do not appear, then one can sometimes tell from context which of the two versions of the order is meant. However, it is also commonly the case that the difference between these two versions of the order is irrelevant. Thus, we think of $(M, <)$ and $(M, \leq)$ as giving the strict and nonstrict versions of the *same* poset structure on $M$.

## 11.E Exercises

**Exercise 11.1.** Let $M = \{1, 2, 3, 4\}$ and let $R = \{(1, 1), (1, 3), (2, 3)\}$. Write each of $R^{\text{ref}}$, $R^{\text{sym}}$, $R^{\text{tran}}$, $(R^{\text{sym}})^{\text{tran}}$, $(R^{\text{tran}})^{\text{sym}}$, and $((R^{\text{ref}})^{\text{sym}})^{\text{tran}}$ as sets of ordered pairs. (Note that the order in which one takes the symmetric closure and the transitive closure matters. We might say that these two operations do not commute. This is explored further in Exercise 11.11, where we see that the symmetric closure should occur before the transitive closure.)

**Exercise 11.2.** Which of the first six properties on binary relations defined in this section depend only on the ordered pairs in the relation $R$ (and not on $M$)?

**Exercise 11.3.** If $M = \emptyset$ and $R = \emptyset$, which of the first six properties on binary relations defined in this section does $R$ satisfy? (Interpret universally quantified statements as vacuously true on $\emptyset$.)

**Exercise 11.4.** Let $R$ be a binary relation on a set $M \neq \emptyset$. Prove that $R$ cannot be both reflexive and irreflexive. Also prove that $R$ cannot be both symmetric and antisymmetric unless $R \subseteq \{(x, x) : x \in M\}$ (and hence it cannot be both symmetric and asymmetric unless $R = \emptyset$).

**Exercise 11.5.** Finish the proof of Proposition 11.2, showing that every binary relation has an irreflexive reduction.

**Exercise 11.6.** Prove Proposition 11.4. (Hint: One of the parts might help with another part.)

**Exercise 11.7.** Taking $\prec$ as an arbitrary primitive binary relation symbol, give a first order formula that defines $\precneq = \prec^{\mathrm{irref}}$. Similarly, give a first order formula that defines the inverse relation $\succ$ and another that defines the symmetric closure $\prec \cup \succ$.

**Exercise 11.8.** Prove Lemma 11.7.

**Exercise 11.9.** Prove the converse of Proposition 11.8: If $R$ is a binary relation that is not antisymmetric, then there is no largest asymmetric relation $S \subseteq R$. (As a corollary, this implies that there is no canonical way to define an antisymmetrization of a relation that isn't already antisymmetric.)

**Exercise 11.10.** Show that the symmetric closure and the transitive closure satisfy appropriate versions of parts (1), (2), and (3) of Proposition 11.4.

**Exercise 11.11.** For any binary relation $R$ on a set $M$, prove that the reflexive closure commutes with both the symmetric closure and the transitive closure (in other words, $(R^{\mathrm{ref}})^{\mathrm{sym}} = (R^{\mathrm{sym}})^{\mathrm{ref}}$ and $(R^{\mathrm{ref}})^{\mathrm{tran}} = (R^{\mathrm{tran}})^{\mathrm{ref}}$.) Further prove the equality $((R^{\mathrm{sym}})^{\mathrm{tran}})^{\mathrm{sym}} = (R^{\mathrm{sym}})^{\mathrm{tran}}$.

**Exercise 11.12.** (This is a lengthy problem.) Given the first six properties defined on relations in this section, there are a total of sixty-four possible combinations of those properties and their negations holding for some relation $R$ on some *nonempty* set $M$. (We handled the empty set earlier.) Some of those possibilities are disallowed; for instance, any relation that is asymmetric is automatically antisymmetric and irreflexive, hence also not reflexive. Give examples (i.e., models) for the possible cases (with minimal cardinality for $M$, and then for $R$), and prove that the other cases not covered by your examples cannot occur. (Did you discover any new relationships among these properties that were not described in this section?)

**Exercise 11.13.** (Extremely easy exercise.) Show that POSET can be defined using only one axiom. (Hint: It isn't just one of the three defining conditions.)

**Exercise 11.14.** Is there a "posetization" of an arbitrary binary relation?

**Exercise 11.15.** Consider the divisibility relation $\mid$ on $\mathbb{N}$. Show that it is a (nonstrict) partial order, with 0 the biggest element. (When defining "greatest common divisor", the word "greatest" is referring to this partial ordering of $\mathbb{N}$, and not the usual ordering $<$ on $\mathbb{N}$, which is why $\mathrm{GCD}(0,0) = 0$.)

**Exercise 11.16.** Prove that the transitive closure is not, generally, first order definable, even if we have access to two binary relation symbols, as follows.

Let $\triangle$ and $\square$ be the two binary relation symbols. Let $T_1$ be the set of all statements in that signature that are satisfied in every model $\mathcal{M}$ such that $\square_{\mathcal{M}} = (\triangle_{\mathcal{M}})^{\mathrm{tran}}$ holds.

To simplify the situation, we will now add new constant symbols to the signature. First, add constant symbols indexed by the meta numbers $c_0, c_1, c_2, \ldots$, and then add one more constant symbol $c_\infty$. Let $T_2$ be the set of statements asserting that these constant symbols are distinct; so $c_i \neq c_j$ for distinct $i, j \in \mathbb{N} \cup \{\infty\}$.

Let $\varphi(u, v)$ be the formula $u \triangle v \wedge \forall s\,(u \triangle s \rightarrow s = v)$, which says that $v$ is a unique element that $\triangle$-relates to $u$ on the right. Now do the following:

(1) For each $n \in \mathbb{N}$, construct a model $\mathcal{M}_n$ satisfying $\varphi(c_0, c_1), \ldots, \varphi(c_n, c_{n+1})$, as well as $T_1 \cup T_2 \cup \{c_0 \square c_\infty\}$. (To make $T_1$ hold, you need to guarantee that $\square_{\mathcal{M}_n} = (\triangle_{\mathcal{M}_n})^{\text{tran}}$.)

(2) Apply the compactness theorem to show that $T_1 \cup T_2 \cup \{c_0 \square c_\infty\}$ together with $\{\varphi(c_n, c_{n+1})\}_{n \in \mathbb{N}}$ is satisfiable.

(3) Explain how the previous two parts show that no set of axioms can define the transitive closure in the given signature (or, really, in any signature).

# 12 Total orders and monotone maps

In the previous section we studied the fundamental conditions that encapsulate what it means to put an “order” on a set of objects. The resulting definitions allowed situations where two objects could be completely unrelated, and so we called these “partial” orders. One can picture partial orders as allowing branching along different paths, as in the diagram from Example 11.15.

In the present section we define “total” orders. These are orderings where every pair of objects are ordered, and so there is no branching. The prototypical example is $\mathbb{N}$ under its usual ordering. One can picture total orders in a straight up-and-down manner; thus, total orders are also called “linear” orders.

In this section we will also define certain maps between ordered sets that let us tell when two ordered sets look the same. As a consequence, we prove that there is really only one type of total order on a finite set.

## 12.A Totality through trichotomy

We start with the following basic definitions, which capture the idea that any pair of objects must be ordered.

**Definition 12.1.** Let $R$ be a binary relation on a set $M$.

We say $R$ is *trichotomous* if $\forall x, y \in M\ (xRy \vee yRx \vee (x = y))$.

We say $R$ is *connective* if $\forall x, y \in M\ (xRy \vee yRx)$.

The three options in trichotomy are mutually exclusive if and only if the relation is asymmetric. Thus, one reason asymmetry is often explicitly included in the list of axioms for strict partial orders (even though it could be dropped) is to make it easier to observe this mutual exclusion.

Also notice that connectivity implies trichotomy, but not conversely. It is often used in place of trichotomy when handling nonstrict partial orders because it is slightly shorter to write down and doesn’t require any reference to the equality relation.

Trichotomy is the condition we need to define a total order. We will state our definition using the strict version of the order, but it works just as well for the nonstrict version.

**Definition 12.2.** Let $<$ be a strict partial order on a set $M$. We say $<$ is a *total order* on $M$ if $<$ is trichotomous. In this case we say $(M, <)$, or sometimes just $M$, is a *totally ordered* set.

Total orders and partial orders are often just called orders, especially when there is no need to emphasize trichotomy (or lack thereof).

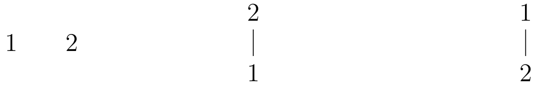
**Warning 12.3.** All total orders are partial orders. The word “partial” does not necessarily imply lack of trichotomy or connectivity.

**Example 12.4.** Let $M = \{1\}$. There is exactly one partial order on $M$, and thus exactly one total order on $M$. It is $\leq\, = \{(1,1)\}$. (The strict version is $<\, = \emptyset$.) △

**Example 12.5.** Let $M = \{1,2\}$. There are three partial orders on $M$; they are (using the reflexive versions) exactly

$$\{(1,1),(2,2)\},\ \{(1,1),(2,2),(1,2)\},\quad \text{and}\quad \{(1,1),(2,2),(2,1)\}.$$

Pictorially, these three posets are:



Only the last two of these partial orders are total orders. Up to renaming the elements of $M$, we see that the two total orders are the same. △

## 12.B Maps between ordered sets

There are many times it is natural to map between ordered structures. For instance, in calculus every function $f\colon \mathbb{R} \to \mathbb{R}$ is a map between ordered structures. When $f$ is continuous, understanding where $f$ is increasing and decreasing helps us find the critical points and local extrema. Maps on ordered sets are thus extremely useful in optimization problems, generally.

In Example 12.5 we have an example of two orders on a set that are, colloquially, the same after renaming the elements. The following definition will give us the language to make this mathematically precise.

**Definition 12.6.** Let $(M, \leq)$ and $(N, \leq')$ be two nonstrict partial orders. We say that a function $f\colon M \to N$ is *monotone* (sometimes called *isotone*) if whenever $a \leq b$, for some $a, b \in M$, then $f(a) \leq' f(b)$.

If $f\colon M \to N$ is a bijection, we say $f$ is an *order isomorphism* when both $f$ and $f^{-1}$ are monotone.

The use of nonstrict partial orders in Definition 12.6 is intentional, as the following example illustrates.

**Example 12.7.** Let $M = \{1,2\}$ and let $\leq\, = \{(1,1),(2,2),(1,2)\}$. The function $f\colon M \to M$ where $f(1) = f(2) = 2$ is monotone. Notice that $1 < 2$ but $f(1) \not< f(2)$. In other words, the function $f$ is nondecreasing, but it is not strictly increasing. △

**Example 12.8.** Let $M = \{1,2\}$. Also take $\leq\, = \{(1,1),(2,2),(1,2)\}$ and $\leq'\, = \{(1,1),(2,2),(2,1)\}$. The bijection $f\colon M \to M$ where $f(1) = 2$ and $f(2) = 1$ is an order isomorphism between $(M, \leq)$ and $(M, \leq')$. △

**Advice 12.9.** Think of the word "monotone" as short for "nondecreasing". If you want the map to have the stronger property that it respects strict inequalities, rather than just nonstrict ones, call it a "strictly increasing" map. Never use "increasing" without also using "strictly".

**Warning 12.10.** Some authors use the word "monotone" to mean "strictly increasing". This is no longer standard.

While not every monotone map is strictly increasing, it turns out that order isomorphisms are strictly increasing, as the following proposition shows.

**Proposition 12.11.** *Let $(M, <)$ and $(N, <')$ be two partial orders. If $f\colon M \to N$ is an injective monotone map, then $f$ is strictly increasing, meaning*

$$a < b \text{ implies } f(a) <' f(b),$$

*for any $a, b \in M$. In particular, order isomorphisms are strictly increasing.*

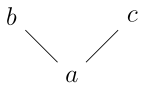
*Proof.* Assume $a < b$ for some $a, b \in M$. Since $f$ is monotone, we have $f(a) \leq' f(b)$. Since $f$ is injective, but $a \neq b$, we must have $f(a) <' f(b)$, as desired. □

Every condition on a poset that can be expressed in terms of just the poset structure, which doesn't depend on the names of the elements, is preserved under order isomorphisms. For instance, if $M$ has a chain $a_1 < a_2 < a_3$ of length three, and $f\colon M \to N$ is an order isomorphism, then $N$ similarly has a chain $f(a_1) < f(a_2) < f(a_3)$ of length three. (We may drop the dash on the symbol $<'$ when there is no danger in confusing the different partial orders.) Exercise 14.8 gives additional examples of this principle.

For two posets to be order isomorphic, they must have the same cardinality. For finite totally ordered sets the converse is true. Before we prove that fact we first introduce some useful notation and examples.

## 12.C Maximal versus greatest

In posets, elements do not need to be related. Thus, there may be multiple elements that are as big as possible. This happens in the poset



where the two elements $b$ and $c$ are as big as possible. The following definitions help differentiate between two closely related notions of "bigness".

**Definition 12.12.** Let $<$ be a strict partial ordering on a set $M$.

An element $x \in M$ is *maximal* if $\forall y \in M\,(x \not< y)$.

An element $x \in M$ is *greatest* (or *largest*) if $\forall y \in M\,((y = x) \vee (y < x))$.

Minor modifications of these definitions work just as well for the corresponding nonstrict version of the poset. For instance, an element $x \in M$ is maximal when

$$\forall y \in M\,(x \leq y \to x = y).$$

We leave it to the reader to write the "greatest element" condition using the nonstrict version of the partial order.

**Advice 12.13.** To a mathematician, the word "maximal" means that there is nothing bigger. But "greatest" means that everything else is smaller. The second condition is stronger than the first.

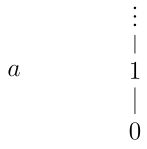
The motivated student might now attempt to define what it means for an element of a poset to be *minimal*, and also what it means to be the *least* (or *smallest*) element. The idea should be that there is nothing smaller than a minimal element, but all elements are greater than the least element (except itself).

There are some immediate connections between greatest and maximal elements.

**Proposition 12.14.** *In a poset, a greatest element is a unique maximal element. In particular, there can only ever be one greatest element. The converse can fail in the sense that a unique maximal element may not be a greatest element.*

*Proof.* Let $(M, <)$ be a poset, and let $x \in M$ be a greatest element. We wish to show that $x$ is maximal, so let $y \in M$ be arbitrary. If $x = y$ then $x \not< y$ by irreflexivity. So now consider when $x \neq y$. Since $x$ is a greatest element, then $y < x$. By asymmetry we get $x \not< y$, as needed. This same computation also shows that $y$ cannot be a maximal element, so $x$ is unique in this regard.

To see that the converse can fail, consider the poset $\mathbb{N} \cup \{a\}$ where $\mathbb{N}$ is ordered as usual and there are no other relationships. This poset can be pictured as follows.



The element $a$ is the unique maximal element, but it isn't a greatest element. □

In a poset, generally there can be many maximal elements. Can you think of an example?

We previously used bigness terminology in the book without comment. For example, in Exercise 11.15 we discussed *greatest* common divisors using the divisibility partial order on $\mathbb{N}$. Similarly, the reflexive closure of a binary relation $R$ on a set $M$ was defined as the *smallest* reflexive relation containing $R$. In other words, under the partial ordering relation $\subseteq$ on $\mathscr{P}(M \times M)$, which is the set of all binary relations on $M$, then $R^{\text{ref}}$ is the least element of the subset

$$\{S \in \mathscr{P}(M \times M) \;:\; S \text{ is a reflexive relation on } M \text{ containing } R\}.$$

We also referred to $\omega$ as the *smallest* inductive set.

Interestingly, the collection of partial ordering relations forms a poset under set inclusion. In mathematical settings it is commonly the case that the very subject one is studying is reflected back into itself. (This is also the case of formal set theory, for instance.)

There is some care that needs to be taken when differentiating between maximal and minimal elements, as the following example illustrates.

**Example 12.15.** Let $R$ be a strict partial order relation on $M$. It is not hard to show that $R^{-1}$, the inverse relation, is also a strict partial order relation on $M$. The maximal elements for $R$ are exactly the minimal elements of $R^{-1}$, and vice versa.

More concretely, consider the usual order relation $<$ on $\mathbb{N}$. The element 0 is the *least* element of $\mathbb{N}$ under this ordering, but it is the *greatest* element of $\mathbb{N}$ under the inverse ordering $<^{-1} = >$. In other words, the definition of a least element for a binary relation means "*least on the left*". △

**Advice 12.16.** As in the previous example, if a symbol like $>$ is used, and the smaller side is supposed to occur on the left, it is better to use $<^{-1}$ for the relation because it is standard to think of the "smaller" side of $>$ as occurring on the right. You can also use more precise terminology, like "$<^{-1}$-least", as needed.

## 12.D Finite totally ordered sets

In "Transition" we proved that every nonempty finite subset of $\mathbb{R}$ has a greatest element. This fact is true much more generally, using essentially the same proof.

**Lemma 12.17.** *If $<$ is a total ordering on a nonempty finite set $M$, then there is a greatest element.*

*Proof.* If $|M|$ has exactly one element, then that element is the greatest (and least) element, automatically. So we may assume $|M| \geq 2$, and also we inductively assume that each totally ordered set of smaller cardinality has a greatest element. Now, fix $x \in M$ and consider the set $M - \{x\}$, under the same relation $<$ but now restricted to $(M - \{x\})^2$. It is easy to check that the restricted relation is a total ordering on $M - \{x\}$. By our inductive hypothesis, there is a greatest element $y \in M - \{x\}$.

**Case 1**: Suppose $x < y$. Then $y$ is the greatest element of $M$.

**Case 2**: Suppose $y < x$, which is the only other case by trichotomy. Since $y$ is a greatest, hence maximal, element of $M - \{x\}$, we have $z \leq y$ for each $z \in M - \{x\}$. Thus, by transitivity, $z < x$ for each $z \in M - \{x\}$. □

We are now ready to show that for each finite cardinality, there is—up to order isomorphism—only one type of totally ordered set of that cardinality.

**Theorem 12.18.** *Let $(M, \leq)$ and $(N, \leq')$ be total orders, with $|M| = |N|$ finite. Then $M$ is order isomorphic to $N$.*

*Proof sketch.* The case when $M$ has cardinality 0 is obvious. So we may assume $|M| \geq 1$, and that the theorem is true for smaller cardinalities.

By Lemma 12.17, we may fix greatest elements $x \in M$ and $y \in N$. By restricting our orders, we know that $M - \{x\}$ and $N - \{y\}$ are order isomorphic, say by a bijection $f\colon M - \{x\} \to N - \{y\}$. Extend $f$ to the function $f \cup \{(x, y)\}\colon M \to N$. We leave it as an exercise to show that $f$ is monotone; and then by a symmetric argument $f^{-1}$ is also monotone. Hence $f$ is an order isomorphism. □

## 12.E Exercises

**Exercise 12.1.** Find a strict total ordering $\lhd$ on $\mathbb{N}$, such that $(\mathbb{N}, \lhd)$ has both a least and a greatest element.

**Exercise 12.2.** Let $(M, R)$ be a strict poset. Show that $(M, R^{-1})$ is also a strict poset.

**Exercise 12.3.** Prove that the reflexive closure of a trichotomous relation is trichotomous, as is the irreflexive reduction.

**Exercise 12.4.** Prove that a relation $R$ on a set $M$ is connective if and only if it is trichotomous and reflexive.

**Exercise 12.5.** How many partial orders are there on $M = \{1, 2, 3\}$? How many up to order isomorphism? Draw them all.

**Exercise 12.6.** How many total orders are there on a finite set of size $n$?

**Exercise 12.7.** Give an example of a monotone bijection between two partially ordered sets that is not an order isomorphism. (Is this possible if the two sets are totally ordered?)

**Exercise 12.8.** Fill in the gap in the proof of Lemma 12.17, by showing the following: If $<$ is a total ordering on a set $M$, and if $S \subseteq M$, then the restricted relation $<|_{S \times S}$ is a total ordering of $S$.

**Exercise 12.9.** Finish the proof of Theorem 12.18. (Is $f$ unique?)

**Exercise 12.10.** Prove, for each meta number $n \geq 1$, that "$\prec$ is a strict total order on a domain with $n$ elements" is first order definable.

# 13 Standard constructions with ordered sets

Just as sets can be combined together in multiple ways, so too can ordered sets be combined together to form new ordered sets. In this section we will introduce a few of the most common constructions.

## 13.A Summing ordered sets

Let $(M_1, \leq_1)$ and $(M_2, \leq_2)$ be any two totally ordered sets. In many situations it can be useful to describe a total order on the disjoint union of $M_1$ and $M_2$. One of the most natural orders is to just put $M_2$ above $M_1$. For example, suppose that $(M_1, \leq_1)$ is a ranking of your most embarrassing experiences, while $(M_2, \leq_2)$ is a list of your proudest moments. When ranking these together, you would probably just automatically place all the elements of $M_2$ above those in $M_1$.

To make this precise, we take the disjoint union to be $(\{1\} \times M_1) \cup (\{2\} \times M_2)$. We order this set by a relation, denoted $\leq$, defined using the three rules:

(1) $(1, x) \leq (1, y)$ holds if $x \leq_1 y$ (for any $x, y \in M_1$),
(2) $(2, x) \leq (2, y)$ holds if $x \leq_2 y$ (for any $x, y \in M_2$), and
(3) $(1, x) \leq (2, y)$ always holds (for any $x \in M_1$ and $y \in M_2$).

**Definition 13.1.** The ordered set described above is the *sum* of two ordered sets, written $M_1 + M_2$.

**Example 13.2.** The ordered set $\mathbb{N}+\mathbb{N}$ consists of two disjoint copies of $\mathbb{N}$, one above the other. That disjoint union is $(\{1\} \times \mathbb{N}) \cup (\{2\} \times \mathbb{N})$. We can write the elements of this set, in order (starting at the smallest), as

$$(1,0), (1,1), (1,2), \ldots, (2,0), (2,1), (2,2), \ldots. \qquad \triangle$$

**Example 13.3.** We can define the ordered set $\mathbb{Z}$ (or, at least an order isomorphic copy) in terms of the ordered set $(\mathbb{N}, \leq)$. Let $\mathbb{N}^*$ be a copy of the natural numbers, but under the inverse relation $\leq^{-1}$. Then the set we want is just $\mathbb{N}^* + \mathbb{N}$. (Does this differ from $\mathbb{N} + \mathbb{N}^*$?) △

We can generalize by taking the ordered sum of any finite number of disjoint ordered sets. Even more generally, let $(I, \leq_I)$ be any ordered indexing set, and suppose that for each $i \in I$ we have an ordered set $(M_i, \leq_i)$. We can then take $\sum_{i\in I} M_i$ to be the set

$$\{(i, x) : i \in I \text{ and } x \in M_i\}.$$

Define an ordering $\leq$ on this set by the rule that $(i, x) \leq (j, y)$ holds if and only if both (1) $i \leq_I j$, and (2) if $i = j$ then $x \leq_i y$.

**Example 13.4.** To test your understanding of this concept, construct the ordered set $\mathbb{N} + \mathbb{N} + \mathbb{N} + \cdots = \sum_{n\in\mathbb{N}} \mathbb{N}$. △

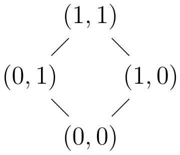
**Example 13.5.** The order type for the ordered set that will be used in the hint for Exercise 14.11 is just $X_{i_0} + X_{i_1} + X_{i_2} + \cdots = \sum_{n\in\mathbb{N}} X_{i_n}$. △

## 13.B Multiplying ordered sets

Let $(X, \leq)$ and $(Y, \leq)$ be two totally ordered sets. In this subsection we have no need to envision $X$ and $Y$ as disjoint, and so there is no need for subscripts, nor different symbols for the orders on $X$ and $Y$, leaving it to the reader to tell which relation is being used from context. (If you need to distinguish between the two orders, just write $\leq_X$ and $\leq_Y$, respectively.)

There are many different orders that can be given to the direct product $X \times Y$. The choice is usually made from the context of the situation. One of the most natural choices is the *product order*. Here we say $(a, b) \leq (c, d)$ exactly when $a \leq c$ and $b \leq d$. Unfortunately, this is usually not a total order, only a partial order.

**Example 13.6.** Order the set $X = \{0, 1\}$ as usual. The product (partial) order on the Cartesian product $X^2$ is pictured below.

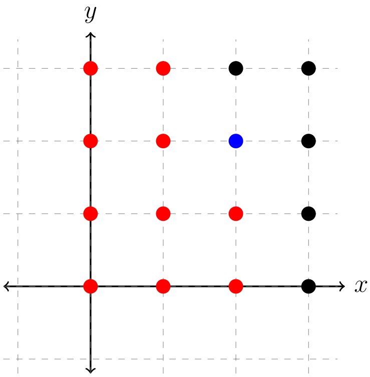


Can you picture $\mathbb{N} \times \mathbb{N}$? Does it have any maximal elements? △

Another popular ordering is the *lexicographical order*. This is the ordering where $(a, b) \leq (c, d)$ holds exactly when $a \leq c$, and if $a = c$ then $b \leq d$. This ordering derives its name from the fact that dictionary words are ordered by comparing their first letters, and if the first letters agree we look at their second letters, and so on. Similarly, here we compare the first coordinates, and if those agree we compare the second coordinates.

**Example 13.7.** Consider the ordered set $X = \{0, 1\}$ as before. The lexicographical order gives us $(0, 0) < (0, 1) < (1, 0) < (1, 1)$.

Next consider $\mathbb{N} \times \mathbb{N}$, pictured as the following infinite array of points in the Cartesian plane.



We can think of the lexicographical ordering as progressively going up each column, in order. In other words, one point $p$ is smaller than another point $q$ exactly when $p$ occurs strictly leftward from $q$ (and possibly up or down), or in the same column but downward. The red points are below the blue one, and the black points are above the blue one. △

When $(X, \leq) = (Y, \leq)$, there is another, little appreciated, order on $X \times Y$ that we will call the *canonical order*. Here we put $(a, b) \leq (c, d)$ exactly when the following three conditions hold:

(1) $\max(a, b) \leq \max(c, d)$,
(2) if $\max(a, b) = \max(c, d)$, then $a \leq c$, and
(3) if $\max(a, b) = \max(c, d)$ and $a = c$, then $b \leq d$.

**Remark 13.8.** The canonical order on $\mathbb{N} \times \mathbb{N}$ gives us

$$(0,0) < (0,1) < (1,0) < (1,1) < (0,2) < (1,2) < (2,0) < (2,1) < (2,2) < \dots,$$

and can be pictured as "filling in" larger and larger squares in the plane. The canonical order on $\mathbb{N} \times \mathbb{N}$ is order isomorphic to the standard order on $\mathbb{N}$. In particular we have $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$. You probably proved this equality previously by enumerating $\mathbb{N} \times \mathbb{N}$ in some other *ad hoc* way (perhaps by filling in larger and larger triangles in the plane). ▲

One can generalize the canonical order to arbitrary finite direct products of the same ordered set, but the definition gets more complicated. Similarly, one can generalize the lexicographical order to arbitrary finite direct products, by comparing entries one at a time.

For infinite direct products, problems can arise. For example, consider the set $X = \{0, 1\}$. Take the infinite direct product of this set indexed over $\mathbb{Z}$,

$$S = \prod_{i \in \mathbb{Z}} X = \cdots \times X \times X \times X \times \cdots,$$

and consider the sequence $s_0 \in S$ with 1's in the even entries and 0's in the odd entries. Oppositely, let $s_1 \in S$ have 1's in the odd entries and 0's in the even entries. Should we consider $s_0$ or $s_1$ to be larger? Because $\mathbb{Z}$ has no place to "start" there is no easy way to make this choice. If we instead use the index set $\mathbb{N}$, then there is a natural place to start, and thus we have a very nice order relation on $\mathbb{N}$-indexed products, which we will now describe.

## 13.C Orders on products over the natural numbers

For each $n \in \mathbb{N}$ let $(X_n, \leq)$ be a totally ordered set. Think of $\prod_{n \in \mathbb{N}} X_n$ as a set of sequences $(x_0, x_1, x_2, \ldots)$ where each $x_n \in X_n$. We order these sequences by the "compare first difference" relation. Thus, $(x_0, x_1, \ldots) < (y_0, y_1, \ldots)$ holds exactly when $x_n < y_n$ for the first number $n \in \mathbb{N}$ where $x_n \neq y_n$ (when such an $n$ exists). This generalizes the usual lexicographical order, and it is a total ordering.

## 13.D Exercises

**Exercise 13.1.** Draw a picture of the totally ordered set $\mathbb{Z} + \mathbb{Z}$, where each copy of $\mathbb{Z}$ has its usual ordering.

**Exercise 13.2.** Explicitly verify that the binary “order” relation defined on the sum of two totally ordered sets is a total order.

**Exercise 13.3.** If $(M_1, <_1)$ and $(M_2, <_2)$ are posets, and we define $M_1 + M_2$ in a similar manner as above, is it a poset?

**Exercise 13.4.** Explicitly describe the set $\mathbb{N} + \mathbb{N} + \mathbb{N} + \cdots$ in Example 13.4, using set-builder notation. Do the same for the total ordering on this set.

**Exercise 13.5.** Let $X = \{0, 1\}$, ordered as usual. Draw a picture for the product partial order on $(X \times X) \times X$.

**Exercise 13.6.** Show that the product order induced by two partial orders is a partial order.

**Exercise 13.7.** Show that the lexicographical order induced by two total orders is a total order. (Is the lexicographical order induced by two partial orders a partial order?)

**Exercise 13.8.** List the next ten ordered pairs in the canonical order on $\mathbb{N} \times \mathbb{N}$, after the nine already listed in Remark 13.8. Next, drawing $\mathbb{N} \times \mathbb{N}$ as a square grid, find a way to *visually* represent the canonical ordering.

**Exercise 13.9.** Show that the canonical order on the product of a totally ordered set with itself is a total order.

**Exercise 13.10.** (This is an important exercise.) Explicitly define an order isomorphism from $\mathbb{N} \times \mathbb{N}$ under the canonical order to $\mathbb{N}$ under the standard order.

**Exercise 13.11.** Let $(X, <)$ be a totally ordered set. Define a “canonical” order on $X^3 = X \times X \times X$. (Can you give a definition that works for $X^n$, for any (nonmeta!) natural number $n \geq 2$? Hint: Think of $X^n$ as the set of functions from $n \to X$, treating $n \in \omega$ as a set. Technically, ordered $n$-tuples are not the same as $n$-sequences, but we will ignore that issue.)

**Exercise 13.12.** Prove that $(0, 0, 0, \ldots) \in \prod_{n\in\mathbb{N}} \mathbb{N}$ is the least element under the “compare the first difference” relation. Does $\prod_{n\in\mathbb{N}} \mathbb{Z}$ have a least element? (Prove your answer.)

**Exercise 13.13.** Let $S \subseteq \prod_{n\in\mathbb{N}} \mathbb{N}$ be the set of sequences where there are only finitely many nonzero entries. Show that the restriction of the “compare the last difference” relation to $S$ totally orders $S$.

**Exercise 13.14.** Using the usual order on $\mathbb{Z}$, describe the canonical order on $\mathbb{Z} \times \mathbb{Z}$.

# 14 Recognizing nonisomorphic orders

It can take practice to recognize whether or not two different ordered sets are order isomorphic. In this section, we develop a few standard techniques that help in this task. Throughout this section (and in the exercises), let $(M, \leq)$ and $(N, \leq)$ be two arbitrary partially ordered sets. We will always be able to tell, from context, which of the two orders is being referenced, so we can safely use the same symbol for the two orders.

## 14.A Cardinality

An order isomorphism $f \colon M \to N$ is a monotone *bijection* whose inverse is also monotone. In particular, if $|M| \neq |N|$, then $M$ and $N$ cannot be order isomorphic.

**Example 14.1.** The sets $\mathbb{N}$ and $\mathbb{R}$ are not order isomorphic under their usual orders (or, in fact, under *any* partial orderings) since they have different cardinalities. △

Cardinality can be exploited in other ways, as in the following example.

**Example 14.2.** We will show that the sets $\mathbb{N}$ and $\mathbb{Z}$ are not order isomorphic under their usual orders. Since $\mathbb{N}$ and $\mathbb{Z}$ are both countable, it is not enough to look at the cardinalities of these two sets. However, given any element $x \in \mathbb{N}$ we see that the set of smaller elements in $\mathbb{N}$ is

$$\{z \in \mathbb{N} \,:\, z < x\} = \{0, 1, \ldots, x-1\},$$

which is finite. On the other hand, for any $y \in \mathbb{Z}$, the set of smaller elements in $\mathbb{Z}$ is

$$\{z \in \mathbb{Z} \,:\, z < y\} = \{y-1, y-2, \ldots\},$$

which is always (countably) infinite. If $f \colon \mathbb{N} \to \mathbb{Z}$ were an order isomorphism sending $x \mapsto y$, then it would (bijectively) send the elements that are smaller than $x$ to those that are smaller than $y$. This is impossible, since the cardinalities of those two sets do not match. △

At this point, you might try to show that the sets $\mathbb{Z}$ and $\mathbb{Q}$ are not order isomorphic under their usual orders, by using a cardinality argument. (Hint: There are infinitely many rational numbers between 0 and 1. Is a similar fact true for $\mathbb{Z}$?)

## 14.B Properties on special elements

An element in an ordered set $M$ may stand out as being special because of a property it enjoys. If another ordered set $N$ lacks a similar element, then $M$ and $N$ cannot be order isomorphic.

**Example 14.3.** The element $0 \in \mathbb{N}$ is very special, since it is the least element of $\mathbb{N}$. On the other hand, $\mathbb{Z}$ has no least element. Thus, this gives us another way to see that $\mathbb{N}$ and $\mathbb{Z}$ are not order isomorphic. △

In the previous example, we implicitly used the fact that any order isomorphism must send a least element to a least element. But how do we know that an order isomorphism does that? Here is the proof.

**Proposition 14.4.** *Let $(M, \leq)$ and $(N, \leq)$ be partially ordered sets, and let $f \colon M \to N$ be an order isomorphism. If $a \in M$ is the least element of $M$, then $f(a) \in N$ is the least element of $N$.*

*Proof.* Assume that $a \in M$ is the least element of $M$. To show that $f(a)$ is the least element of $N$, let $b \in N$ be an arbitrary element; our goal is to show that $f(a) \leq b$.

Since $f$ is surjective, there exists some $a' \in M$ such that $f(a') = b$. Since $a$ is the least element of $M$, we have $a \leq a'$. Since $f$ is monotone, we have $f(a) \leq f(a') = b$, as desired. □

The previous proposition is not accidental. Any property that can be described in first order logic (in the language whose signature consists of a single binary relation symbol $\prec$) must be respected by an order isomorphism. This means that an order isomorphism sends minimal elements to minimal elements, which you are asked to prove in Exercise 14.7. This also means that if two elements $a, a' \in M$ have no elements between them, then $f(a), f(a') \in N$ also have no elements between them; this is generalized in Exercise 14.8. Apart from those two exercises, you may now take it as granted that order isomorphisms respect any order properties that can be described in first order logic.

**Example 14.5.** Consider $\mathbb{Z}$ under its usual ordering. Any element $x \in \mathbb{Z}$ has the property that there exists a smaller element $y \in \mathbb{Z}$ such that there are no elements between $y$ and $x$. (Just take $y = x - 1$.) This is a first order property, expressed by the formula

$$\forall x \, \exists y \, (y \prec x \land \neg \exists z \, (y \prec z \land z \prec x)).$$

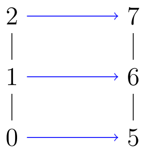
We might call any such element $y$ the "immediate predecessor" of $x$.

On the other hand, no element of $\mathbb{Q}$ has an immediate predecessor (under its usual ordering). So $\mathbb{Q}$ cannot be order isomorphic to $\mathbb{Z}$. △

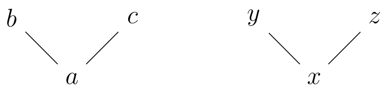
## 14.C Building isomorphisms

The previous subsections discussed some common tactics used to tell when there does not exist an order isomorphism. This subsection instead introduces some techniques for building order isomorphisms when they exist.

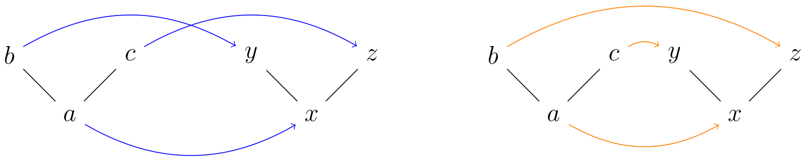
Consider the sets $S = \{0, 1, 2\}$ and $T = \{5, 6, 7\}$ under their usual orders. To build an isomorphism $f \colon S \to T$, we could start by noting that the least element of $S$ must map to the least element of $T$. Thus, we must have $0 \mapsto 5$. Now that this is determined, we see that $f$ must take the next smallest element of $S$ to the next smallest element of $T$, so $1 \mapsto 6$. Finally, we must have $2 \mapsto 7$. Thus, there is exactly one order isomorphism, $f = \{(0, 5), (1, 6), (2, 7)\}$, pictured as follows.



Next, consider the sets $U = \{a, b, c\}$ and $V = \{x, y, z\}$, partially ordered as in the following diagrams.



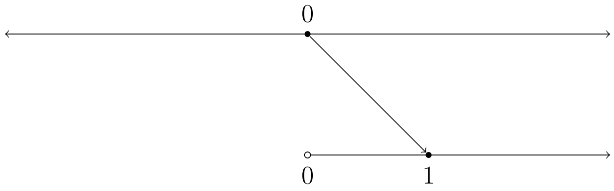
Looking at these pictures, we know that $U$ and $V$ are order isomorphic. Notice that an order isomorphism must send $a \mapsto x$, since the least element of $U$ must go to the least element of $V$. However, there is nothing to differentiate between $b$ and $c$, nor between $y$ and $z$. Thus, there are two different order isomorphisms $f_1 = \{(a, x), (b, y), (c, z)\}$ and $f_2 = \{(a, x), (b, z), (c, y)\}$, drawn in blue and orange below.



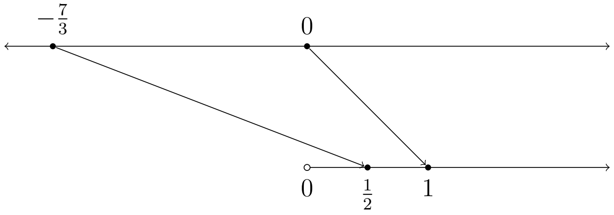
There are situations where all the elements of an ordered set are indistinguishable. For instance, consider the set $\mathbb{Q}$ under its usual ordering. There are a lot of different order isomorphisms from $\mathbb{Q}$ to itself. To see this, let $c \in \mathbb{Q}$ be any fixed rational number. The shifting map $g\colon \mathbb{Q} \to \mathbb{Q}$ given by the rule $g(x) = x + c$ is an order isomorphism. As $c$ was arbitrary, this proves that every element of $(\mathbb{Q}, <)$ looks like every other element. (However, the elements of $\mathbb{Q}$ look very different from one another if we are also allowed to talk about addition and multiplication.)

In some cases we can use the fact that all elements look alike to construct an order isomorphism almost randomly. We will end this section by building an order isomorphism from $X = \mathbb{Q}$ to $Y = \mathbb{Q}_{>0}$.

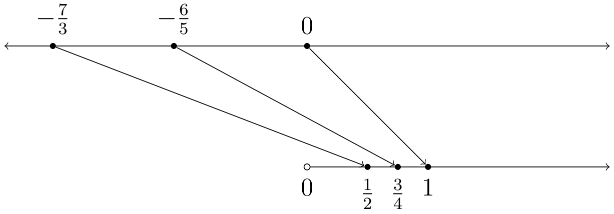
We start by sending a random element of $X$ to a random element of $Y$. For simplicity, let us start with $0 \mapsto 1$. If we view $X$ as a subset of the real line, and $Y$ as a subset of the open interval $(0, \infty)$, then we can picture this situation as follows.



Now pick a second random element of $X$, say $-7/3$. Since $-7/3 < 0$, then (to keep the function monotone) the number $-7/3$ has to map to some element of $Y$ that is smaller than 1, say $1/2$. This gives us following picture.



If we pick a third random element of $X$, there are three possibilities; it could occur in the interval $(-\infty, -7/3)$, or in $(-7/3, 0)$, or in $(0, \infty)$. To keep the function monotone, we must have that element map into the interval $(0, 1/2)$, or in $(1/2, 1)$, or in $(1, \infty)$, accordingly. So, for instance, if our third point is $-6/5 \in (-7/3, 0)$, then a possible output would be $3/4 \in (1/2, 1)$. This gives the following picture.



Now, there are four intervals on the top, and four corresponding intervals on the bottom. As we pick random rational elements from the top, we can continue to pick rational elements on the bottom from the corresponding intervals.

If we continue this way, forever, what would prevent the resulting function from being an order isomorphism? A big problem is that our process of picking random elements might not cover every element of $X$. For instance, perhaps we keep randomly picking negative numbers, and thus we never touch the positive half of $X$.

Even more troublesome is the fact that our "random" process is not mathematical. We need a formal method of picking the next element of $X$. One way to do this is fix, once and for all, a countable list of the elements of $X$. (The fact that the elements of $\mathbb{Q}$ can be put in a countable list was proved in "Transition".) Let $\underline{x} = (x_0, x_1, \ldots)$ be such a listing of the elements of $X = \mathbb{Q}$. As we run over the elements in this list, we are guaranteed to not miss any element of $X$. It is important to notice that the elements of $X$ are *not* listed in their order from $X$. (For instance, $X$ has no least element, but $\underline{x}$ has a first element in its list!)

Dealing with the set $Y$ is a little more complicated. Since $\mathbb{Q}_{>0}$ is countable, we can similarly fix a nonrepeating list $\underline{y} = (y_0, y_1, y_2, \ldots)$ of its elements. But we have no way to guarantee that these elements come in the same "order" as the elements of $\underline{x}$.

Our map $f\colon \mathbb{Q} \to \mathbb{Q}_{>0}$ cannot just send $x_n \mapsto y_n$, for then it wouldn't be monotone. For example, suppose

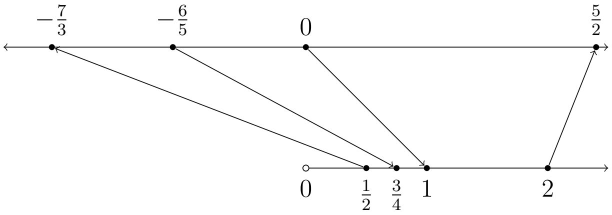
$$\underline{x} = \left(0, -\frac{7}{3}, -\frac{6}{5}, \frac{5}{2}, \frac{1}{2}, \dots\right) \text{ and } \underline{y} = \left(1, \frac{1}{2}, 2, 3, \frac{3}{4}, \dots\right). \tag{14.6}$$

There is no problem with sending $0 \mapsto 1$, and then $-7/3 \mapsto 1/2$, but we cannot send $-6/5 \mapsto 2$ for then the function would no longer be monotone. Instead, what we will do is send $-6/5$ to the *first* element in $\underline{y}$ that has not already been mapped to, and which belongs to the appropriate interval (so that monotonicity is preserved). So, in this example, we see that $-6/5$ does not map to 1 or $1/2$ (because those have already been mapped to), nor does it map to 2 or 3 (because they do not belong to the correct interval), but it maps to $3/4$.

To test your understanding using the enumerations in (14.6), try to discover where $5/2 \in X$ maps in $Y$. (Do you have enough information to tell where $1/2 \in X$ maps in $Y$?)

In the next section, we will describe this process more formally, and show that the resulting map is an order isomorphism. In the meantime, in the exercises below you may take it for granted that $\mathbb{Q}$ and $\mathbb{Q}_{>0}$ are order isomorphic. Before ending this section, we mention one additional trick that makes checking surjectivity easier. Instead of continually working through the elements of $\underline{x}$, instead we can alternate between the elements of $\underline{x}$ and those of $\underline{y}$. This guarantees that we go through both lists completely.

For example, using the enumerations in (14.6) we would first send $0 \in X$ to $1 \in Y$, just as before. But at the next step, we would look at $1/2 \in Y$ and figure out the first (unused) element of $X$ that could go there, which still (accidentally) happens to be $-7/3$. Then we switch back to $X$, and send $-6/5 \mapsto 3/4$. Then we would switch back to $Y$, and figure out which element of $X$ could go to $2 \in Y$, which is $5/2 \in X$.



## 14.D Exercises

In the exercises of this section, subsets of $\mathbb{R} \cup \{-\infty, \infty\}$ are assumed to be ordered as usual. We will continue to let $M$ and $N$ be arbitrary partially ordered sets.

**Exercise 14.1.** Prove or disprove that $\mathbb{Q}$ is order isomorphic to $\mathbb{R}$.

**Exercise 14.2.** Prove or disprove that $\mathbb{N} \cup \{-1\}$ and $\mathbb{N}$ are order isomorphic. (As an additional problem, think about what happens if you replace $-1$ with $-\infty$.)

**Exercise 14.3.** An *initial segment* of $\mathbb{N}$ is a set of the form

$$S_n = \{x \in \mathbb{N} : x < n\} = \{0, 1, \ldots, n-1\},$$

for some $n \in \mathbb{N}$. (Note that $S_0 = \emptyset$.)

Prove that distinct initial segments of $\mathbb{N}$ are never order isomorphic.

**Exercise 14.4.** Prove that $(-\infty, -1] \cup [1, \infty)$ is not order isomorphic to $\mathbb{R}$.

**Exercise 14.5.** Prove that $\mathbb{Z}$ and $\mathbb{Z} + \mathbb{Z}$ are not order isomorphic.

**Exercise 14.6.** Prove that $\mathbb{Q}$ and $\mathbb{Q} + \mathbb{Q}$ are order isomorphic.

**Exercise 14.7.** Prove that if $f\colon M \to N$ is an order isomorphism and $a \in M$ is a minimal element of $M$, then $f(a)$ is a minimal element of $N$.

**Exercise 14.8.** Prove that if $f\colon M \to N$ is an order isomorphism, and $a \in M$ is a point with no immediate predecessor in $M$ (i.e., for every point $a' \in M$ with $a' < a$, there is always a point $a'' \in M$ with $a' < a'' < a$), then $f(a)$ is similarly a point with no immediate predecessor in $N$.

Then prove that if $a$ is a point with no immediate predecessor such that there are exactly $n$ smaller points than $a$ also without immediate predecessors, for some natural number $n \in \mathbb{N}$, then $f(a)$ has the same property.

**Exercise 14.9.** Consider the standard order relation $<$ on $\mathbb{N}$, where we treat $<$ as a set of ordered pairs. Prove that there is a *least* subset of $<$ (under the $\subseteq$ relation) whose transitive closure is $<$. (Is the same true for the standard order on $\mathbb{N} \cup \{\infty\}$? This has some bearing on Exercise 11.16.)

**Exercise 14.10.** Let

$$\underline{x} = \left(0, 1, -1, 2, -2, \frac{1}{2}, -\frac{1}{2}, 3, -3, \frac{3}{2}, -\frac{3}{2}, \frac{1}{3}, -\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}, \ldots\right)$$

be a (partially specified) enumeration of $\mathbb{Q}$. Similarly, let

$$\underline{y} = \left(1, 2, \frac{1}{2}, 3, \frac{3}{2}, \frac{1}{3}, \frac{2}{3}, 4, \frac{4}{3}, \frac{1}{4}, \frac{3}{4}, 5, \frac{5}{2}, \frac{5}{3}, \frac{5}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \ldots\right)$$

be a (partially specified) enumeration of $\mathbb{Q}_{>0}$. Describe, for as many pairs as possible, the order isomorphism $f\colon \mathbb{Q} \to \mathbb{Q}_{>0}$ obtained by sending each successive element of $\underline{x}$ to the first available (and appropriate) element of $\underline{y}$. Also describe, for as many pairs as possible, the order isomorphism $g\colon \mathbb{Q} \to \mathbb{Q}_{>0}$ obtained by alternately sending the next available elements of $\underline{x}$ to the next available element of $\underline{y}$ and vice versa.

**Exercise 14.11.** (This is a harder exercise.) Show that the cardinality of the set of isomorphism types of total orders on a countable set is the continuum. (Hint: Let $X_0$ be $\mathbb{N}$ under the usual ordering. Let $X_1$ be $\{-\infty\} \cup \mathbb{Z}$ under its usual ordering. Now, given a sequence $(i_0, i_1, i_2, \ldots)$ where $i_m \in \{0, 1\}$, construct an ordered set consisting of $X_{i_0}$ followed by $X_{i_1}$, then by $X_{i_2}$, and so forth. Use Exercise 14.8.)

# 15 Density

The standard order on $\mathbb{N}$ is sparse, in the sense that for any given $n \in \mathbb{N}$ there is always a “next smallest element” $n+1$. In other words, there are no elements $x \in \mathbb{N}$ such that $n < x < n+1$. The opposite is true in $\mathbb{Q}$, where there is never an immediate successor.

In this section we will study orders that are not sparse, such as the usual order on the open interval $(0, 1)$. It is the idea of being able to form better and better approximations to physical systems that initially led to the construction of the real numbers, and we will see that in a strong sense this property uniquely defines the real numbers. In this section we will also formalize the “back-and-forth” construction that was introduced in the previous section.

## 15.A Defining density

Once again we are interested in studying order properties. We begin by defining a notion that is opposite that of sparseness.

**Definition 15.1.** Given a totally ordered set $(M, <)$, we say that the order is *dense* if for any $a, b \in M$ with $a < b$, then there is some $c \in M$ with $a < c < b$.

Moreover, we say that a subset $S \subseteq M$ is *dense in* $M$ if the element $c$ above can always be chosen from $S$.

The following is the prototypical example of a dense order.

**Example 15.2.** The usual order relation on $\mathbb{Q}$ is dense. Indeed, suppose that $a, b \in \mathbb{Q}$ with $a < b$. Then $c = a + \frac{b-a}{2}$ is a rational number between $a$ and $b$. (The element $c$ is the midpoint of the open interval $(a, b)$.)

The subset $\mathbb{Z} \subseteq \mathbb{Q}$ is not dense in $\mathbb{Q}$. To see this, note that there is no integer between $a = \frac{1}{3}$ and $b = \frac{2}{3}$.

Let $S \subseteq \mathbb{Q}$ be the set of rational numbers that can be written so that the denominator is a power of 10 (such as $\frac{32}{10} = 3.2$ and $\frac{-71}{100} = -0.71$). In other words, $S$ is the set of rational numbers with finite decimal expansions (unlike $\frac{1}{3} = 0.333\ldots$). In Exercise 15.1 you are asked to show that $S$ is dense in $\mathbb{Q}$. △

**Example 15.3.** The usual order relation on $\{-\infty\} \cup \mathbb{Q}$ is dense. To prove that, let $a, b \in \{-\infty\} \cup \mathbb{Q}$ with $a < b$. If $a \in \mathbb{Q}$, then the same argument as in the previous example works. In the other case, when $a = -\infty$, then we can take $c = b - 1$ to be our needed rational element between $a$ and $b$.

Notice that $\{-\infty\} \cup \mathbb{Q}$ has a least element, namely $-\infty$. On the other hand, $\mathbb{Q}$ has no least element. Thus, these two ordered sets are not order isomorphic by Proposition 14.4.

By similar reasoning using case analysis, $\mathbb{Q} \cup \{\infty\}$, and $\{-\infty\} \cup \mathbb{Q} \cup \{\infty\}$ are also dense orders. We now have four examples of countable dense orders. Can you see why none of these four ordered sets is order isomorphic to any of the others? △

The usual order relation on $(0,1)\cap\mathbb{Q}$ is also dense. Does that give us a fifth type of countable dense order? Perhaps surprisingly, the answer is no. By using the next theorem we can construct an order isomorphism between $\mathbb{Q}$ and $(0,1)\cap\mathbb{Q}$. The proof method is called a "back-and-forth" construction. It is a useful proof technique that appears in other contexts, but this is perhaps the easiest example of how to use it.

**Theorem 15.4.** *Every countable, dense, totally ordered set without a least or greatest element is order isomorphic to any other.*

*Proof.* Let $(M,<)$ and $(N,<)$ be two such ordered sets. Let $(a_0,a_1,\ldots)$ be an enumeration of the elements of $M$, and similarly let $(b_0,b_1,\ldots)$ be an enumeration of $N$. (This is where we use countability.)

We will start by recursively defining *partial functions* $f_0,f_1,f_2,\ldots$ from $M$ to $N$, satisfying some nice properties. By a partial function we just mean a function defined on part of the intended domain, which in this case is $M$.

The 0th stage of the recursion is simple; just take $f_0=\{(a_0,b_0)\}$. Notice that $f_0$ is an order isomorphism from its domain to its range.

For the $(n+1)$th stage of our recursion, suppose that we have already constructed $f_0\subseteq f_1\subseteq\ldots\subseteq f_n$ such that

(1) $f_n$ is an order isomorphism from its domain to its range,
(2) $f_n$ contains only a finite number of ordered pairs,
(3) $a_0,\ldots,a_n\in\operatorname{dom}(f_n)$, and
(4) $b_0,\ldots,b_n\in\operatorname{ran}(f_n)$.

We will not immediately construct a partial function $f_{n+1}$ that satisfies these four properties (with $n$ replaced by $n+1$). Rather, we first construct a partial function $f'_{n+1}$ satisfying only the first three properties.

Let $u_1<u_2<\cdots<u_m$ be the ordering of $\operatorname{dom}(f_n)$, and let $v_1<v_2<\cdots<v_m$ be the ordering of $\operatorname{ran}(f_n)$. This implies

$$f_n=\{(u_i,v_i)\,:\,1\le i\le m\}.$$

**Case 1**: Suppose $a_{n+1}>u_m$. Since $N$ has no greatest element, there must be some element $b_k>v_m$. Letting $k$ be the smallest integer such that this happens, put $f'_{n+1}=f_n\cup\{(a_{n+1},b_k)\}$.

**Case 2**: Suppose $a_{n+1}<u_1$. Since $N$ has no least element, there is a smallest integer $k$ such that $b_k<v_1$ and we take $f'_{n+1}=f_n\cup\{(a_{n+1},b_k)\}$.

**Case 3**: Suppose $u_i<a_{n+1}<u_{i+1}$ for some $1\le i\le m-1$. Since $N$ is dense, there is a smallest integer $k$ such that $v_i<b_k<v_{i+1}$, and we take $f'_{n+1}$ exactly as before.

**Case 4**: Suppose $a_{n+1}\in\operatorname{dom}(f_n)$. Then we just take $f'_{n+1}=f_n$.

In all four cases our new partial function $f'_{n+1}$ is still an order isomorphism from its domain to its range, it still contains only a finite number of ordered pairs, and it is now defined on $a_{n+1}$. If $b_{n+1}\in\operatorname{ran}(f'_{n+1})$, then we can just take $f_{n+1}=f'_{n+1}$. However, if $b_{n+1}\notin\operatorname{ran}(f'_{n+1})$, then we need to make one further adjustment. By applying the same idea as in Cases 1–3, we can take $f_{n+1}=f'_{n+1}\cup\{(a_\ell,b_{n+1})\}$ where

$\ell$ is the smallest integer such that $a_\ell$ fits between the points in $\mathrm{dom}(f'_{n+1})$ in exactly the same way that $b_{n+1}$ fits between the points of $\mathrm{ran}(f'_{n+1})$.

We have now constructed our sequence of partial functions $f_0 \subseteq f_1 \subseteq f_2 \subseteq \dots$. Let $f = \bigcup_{n\in\mathbb{N}} f_n$ be their union. Thus, $f$ is a collection of ordered pairs whose first coordinates come from $M$ and whose second coordinates come from $N$.

We will next prove that every element $a \in M$ appears exactly once as the first coordinate of a pair in $f$. Since $(a_0, a_1, \dots)$ is an enumeration of all the elements of $M$, we know that $a = a_n$ for some $n \in \mathbb{N}$. Thus, $a \in \mathrm{dom}(f_n)$ by property (3), and hence there is some $b \in N$ such that $(a, b) \in f_n$. Therefore $(a, b) \in f$. Now, suppose that there is another pair $(a, b') \in f$, for some (possibly different) $b' \in N$. As $f$ is the union of the partial functions, we have $(a, b') \in f_m$ for some $m \in \mathbb{N}$. Without loss of generality, we may assume $n \leq m$ and so $f_n \subseteq f_m$. Thus, $(a, b), (a, b') \in f_m$. Since $f_m$ is a function, we must have $b = b'$.

By a similar argument, each element of $B$ occurs exactly once as a second coordinate of a pair in $f$. Thus, $f$ is a bijection. Finally, we show that $f$ is monotone (and by a symmetric argument $f^{-1}$ will also be monotone, so $f$ will be an order isomorphism). Let $a, a' \in M$ with $a \leq a'$. For a large enough $n \in \mathbb{N}$, we know that both $a$ and $a'$ are in the domain of $f_n$. Since $f_n$ is an order isomorphism, we have $f_n(a) \leq f_n(a')$. But $f$ agrees with $f_n$ when it is defined, so $f(a) \leq f(a')$. □

**Remark 15.5.** To translate the proof of Theorem 15.4 into formal set theory one needs recursion in the formal theory. We will introduce formal recursion in a future chapter. ▲

## 15.B Order properties defined in terms of subsets

The notion of density in Definition 15.1 was generalized to apply to subsets of the domain of definition. There are other important order-theoretic conditions that are defined in terms of subsets. For instance, the notions of "greatest" and "least" elements can be stated in terms of subsets.

**Definition 15.6.** Let $(M, \leq)$ be a totally ordered set and let $S \subseteq M$.
An element $x \in S$ is the *greatest element of* $S$ if $\forall y \in S\, (y \leq x)$.
An element $x \in M$ is an *upper bound on* $S$ if $\forall y \in S\, (y \leq x)$.

One defines a *least element of* $S$ and a *lower bound on* $S$ similarly.

The difference between an upper bound on a subset and a greatest element of that subset is subtle; the element in question is allowed to belong to the ambient set in the former case. For instance, $(0, 1)$ has no greatest element, but it has many upper bounds in the ambient containing set $\mathbb{R}$. Notice that $\mathbb{R}$ also has no greatest element, but it has an upper bound in $\mathbb{R} \cup \{\infty\}$.

If a set has a greatest element, then that element will be the least possible upper bound (in any ambient containing set). But as the example of $(0, 1) \subseteq \mathbb{R}$ shows, sometimes there is a least possible upper bound even when the subset has no greatest element. Generally, we make the following definitions.

**Definition 15.7.** Let $(M, \leq)$ be a totally ordered set, and let $S \subseteq M$.

We say $x \in M$ is a *least upper bound* (for $S$ in $M$) if it is the least element of the set of all upper bounds for $S$ in $M$ (under the order induced from restricting $\leq$ to the set of upper bounds). In that case we write $x = \text{lub}_M(S)$.

One can define *greatest lower bounds* similarly. Try that now.

**Example 15.8.** Consider the set $S = (0, 1)$. In the larger set $M = \mathbb{R}$, there are many upper bounds. The smallest of these upper bounds is 1, and so $\text{lub}_{\mathbb{R}}(S) = 1$. Similarly, $S$ has a greatest lower bound of 0 in $\mathbb{R}$. △

**Example 15.9.** The set $\{0, 6, 15\}$ has many upper bounds in $\mathbb{N}$. The smallest upper bound is 15, and so $\text{lub}_{\mathbb{N}}(\{0, 6, 15\}) = 15$. Similarly, $\text{glb}_{\mathbb{N}}(\{0, 6, 15\}) = 0$. △

**Example 15.10.** The set $\mathbb{N}$ has no upper bounds in $\mathbb{R}$, and so it has no least upper bound. However, it has a greatest lower bound of 0. △

An important fact about the usual order relation on $\mathbb{R}$ is that every nonempty subset $S \subseteq \mathbb{R}$ that has *some* upper bound in $\mathbb{R}$ then has a least upper bound in $\mathbb{R}$. This fact is sometimes referred to as the "least upper bound property" on $\mathbb{R}$. To talk about this property more generally, we make the following definition.

**Definition 15.11.** If every nonempty subset of a totally ordered set $(M, \leq)$ that has an upper bound in $M$ also has a least upper bound in $M$, then we say that $\leq$ is *complete*.

Quite surprisingly, completeness is the same whether defined in terms of least upper bounds or in terms of greatest lower bounds; see Exercise 15.9.

Every finite totally ordered set is complete, as the following example demonstrates.

**Example 15.12.** Let $(M, \leq)$ be a totally ordered set. If a subset $S$ has a greatest element $x$, then $x = \text{lub}_M(S)$. Indeed, given any upper bound $y \in M$ for the set $S$, then we must have $y \geq x$ since $x \in S$. But $x$ itself is an upper bound on $S$, so it is the least such upper bound.

In particular, a total ordering on a finite set is complete, since any nonempty subset has a greatest element by Lemma 12.17. △

**Example 15.13.** The standard order on $\mathbb{N}$ is complete. To see this, let $S \subseteq \mathbb{N}$ be a nonempty, bounded subset. Since $S$ is bounded, it is finite. So, by the previous example, its greatest element is also its least upper bound in $\mathbb{N}$. △

**Question 15.14.** Is the standard ordering on $\mathbb{Z}$ (or $\mathbb{Q}$) complete? ▲

The completeness of the order relation on $\mathbb{R}$ is treated in some settings as an *axiom* of the structure $(\mathbb{R}, \leq)$, because of the following result. (In the following result, think of $M$ as a copy of $\mathbb{R}$, and think of the "countable, dense subset" of $M$ as $\mathbb{Q}$.)

**Theorem 15.15.** *There is an order isomorphism between any two unbounded, complete, total orderings that have countable, dense subsets.*

*Proof sketch.* (Some of the details of this proof are left as Exercise 15.10.) Let $(M, \leq)$ and $(N, \leq)$ be any two such ordered sets. Let $P \subseteq M$ and $Q \subseteq N$ be countable, dense subsets. Applying Theorem 15.4, there is an order isomorphism $f \colon P \to Q$. We extend this to an order isomorphism $g \colon M \to N$ by the rule

$$g(x) = \mathrm{lub}_N(\{f(y) \,:\, y \in P \text{ and } y \leq x\}). \qquad \square$$

**Advice 15.16.** Think of the previous proof as telling you that once you know how $\mathbb{Q}$ maps under a monotone map, then you can figure out how the rest of $\mathbb{R}$ should map, when the codomain is complete.

It is true that completeness is a property on total orderings, but it is not directly a first order property expressible in the signature $(\prec)$, because we quantify over subsets of the domain. Of course, completeness can be expressed in second order logic, since quantification over subsets of the domain of definition is allowed in that case. This gives us an intuitive reason to believe the stronger claim that there is no first order theory for "the order relation on $\mathbb{R}$" in the signature $(\prec)$. This stronger claim is true, and not too difficult to prove with the right machinery developed, but we won't prove it here.

## 15.C Exercises

**Exercise 15.1.** Let $S \subseteq \mathbb{Q}$ be the set of those rational numbers that have finite decimal expansions. Prove that $S$ is dense in $\mathbb{Q}$, under the usual order on $\mathbb{Q}$.

**Exercise 15.2.** Prove that every countable, dense, totally ordered set with a greatest element but without a least element is order isomorphic to any other.

**Exercise 15.3.** Give an example of infinite subsets $S \subseteq T \subseteq \mathbb{R}$, each given their usual order (as subsets of $\mathbb{R}$), such that $S$ is not dense in $T$, but $S$ is dense in itself. (Can you find an example where $T$ is dense in itself? How about where $T$ is not dense in itself?)

**Exercise 15.4.** Let $(M, <)$ be a totally ordered set, and let $S \subseteq M$ be a dense subset of $M$. Prove that $S$ is a dense subset of itself.

**Exercise 15.5.** In the proof of Theorem 15.4, condition (2) required a partial function to only contain a finite number of ordered pairs. In what places was this assumption implicitly *used* in the proof? (Hint: Note that Case 1 of the proof works just as well without the finiteness condition. We could rephrase that case as asserting that $a_{n+1}$ is greater than all the elements of the domain of $f_n$.)

**Exercise 15.6.** Answer Question 15.14, for both $\mathbb{Z}$ and $\mathbb{Q}$.

**Exercise 15.7.** Let $(M, <)$ be an arbitrary totally ordered set, and let $S \subseteq T \subseteq M$. Suppose that $S$ has a least upper bound in $T$ and also in $M$. Write $x = \text{lub}_T(S)$ and $y = \text{lub}_M(S)$. Prove that $y \le x$, and give an example showing that the inequality may be strict.

**Exercise 15.8.** Consider the signature $(\prec, R)$ where $\prec$ is a binary relation symbol and $R$ is a unary relation symbol. If we interpret $R(x)$ as saying "$x$ belongs to the fixed subset $S$ of the domain" then write a first order formula that means "$y$ is a least upper bound for $S$".

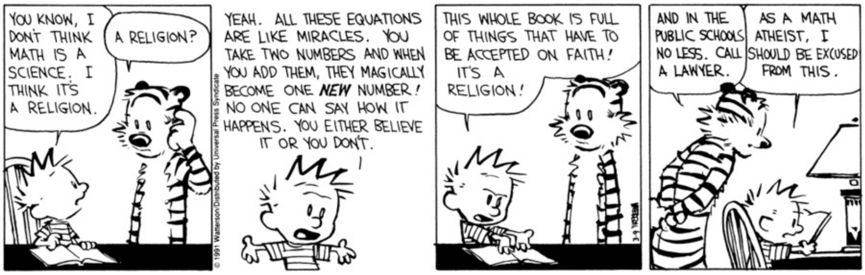
**Exercise 15.9.** Let $(M, \le)$ be a totally ordered set. Prove that $\le$ is complete if and only if every nonempty subset of $M$ that has a lower bound has a greatest lower bound.

**Exercise 15.10.** Finish the proof of Theorem 15.15, as follows:

1. Prove that $P$ (and similarly $Q$) is dense in itself and unbounded.
2. Prove that for each $x \in M$, the set $\{y \in P : y \le x\}$ is nonempty and has an upper bound. (Thus, the definition of $g$ makes sense.)
3. Prove that $g$ is bijective.
4. Prove that $g$ is monotone. (Do we need to prove that $g^{-1}$ is also monotone?)

# 16 Defining common ordered sets of numbers

How do we know that the number 0 exists? We have been taught about it for most of our lives, but have we merely been indoctrinated into believing that it exists? The ancient Greek philosophers raised a classical paradox by asking “How can not being be?” But even more fundamentally, we might ask why any number exists at all. Is mathematics a religion, in that is asks us to take certain existence claims on faith?


YOU KNOW, I DON'T THINK MATH IS A SCIENCE. I THINK IT'S A RELIGION.
A RELIGION?
© 1991 Watterson/Distributed by Universal Press Syndicate
YEAH. ALL THESE EQUATIONS ARE LIKE MIRACLES. YOU TAKE TWO NUMBERS AND WHEN YOU ADD THEM, THEY MAGICALLY BECOME ONE NEW NUMBER! NO ONE CAN SAY HOW IT HAPPENS. YOU EITHER BELIEVE IT OR YOU DON'T.
THIS WHOLE BOOK IS FULL OF THINGS THAT HAVE TO BE ACCEPTED ON FAITH! IT'S A RELIGION!
AND IN THE PUBLIC SCHOOLS NO LESS. CALL A LAWYER.
AS A MATH ATHEIST, I SHOULD BE EXCUSED FROM THIS.

*Calvin and Hobbes*, comic from 1991, by Bill Watterson.

As this is a book about formalizing mathematics, perhaps you’d be tempted to assert zero’s existence by assuming (religiously or not) the axiom of empty set and taking zero to be $\underline{0} = \emptyset$. However, that would be confusing the meta zero with the formal zero.

From a pragmatic point of view, you have seen how useful the notion of zero is. But does utility really have any bearing on ontology? Are other number concepts real only when they are useful? Is zero any more or less imaginary than $i \in \mathbb{C}$?

Many modern mathematicians avoid these philosophical musings, and work as follows. Suppose, for sake of argument, that we have already agreed to the existence of a set of positive whole numbers $W = \{1, 2, 3, \ldots\}$. Assuming that there is some other symbol, say $\star$, not in $W$, then we can take $\mathbb{N} = \{\star\} \cup W$.

Moreover, if $<$ is the usual ordering of $W$, then we can extend that ordering to $\{\star\} \cup W$, by using the new relation $< \cup (\{\star\} \times W)$. This forces $\star$ to be the first element of $\mathbb{N}$, strictly below each of the elements of $W$.

If we also have access to the addition and multiplication functions on $W$, then we can extend them to $\{\star\} \cup W$ by adjoining the rules

$$\begin{array}{ccc} \star + \star = \star, & \star + n = n, & n + \star = n, \\ \star \cdot \star = \star, & \star \cdot n = \star, & n \cdot \star = \star, \end{array}$$

for each $n \in W$.

After we take these steps, we see that $\star$ acts just like 0 should act. Since the symbol didn’t matter, we will go ahead and replace $\star$ by 0. (Alternatively, think about what happens if we replace $0 \in \mathbb{N}$ with a random symbol.) Mathematicians would argue that the specific symbol being used for zero is irrelevant, if it is not already being used; only the properties that the symbol enjoys are important.

What about the existence of more complicated numbers and sets? In previous sections we have used $\mathbb{Q}$ and $\mathbb{R}$ repeatedly; can their existence likewise be justified?

For the rest of this section we will assume that $\mathbb{N}$, and its usual order $<$, exist. This assumption will allow us to construct the more complicated sets $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$, as well as their respective orderings. It should be noted that the specific construction we use is not terribly important, just as the specific symbol used for zero is not important. Our construction will be order isomorphic to any other construction, and so it gives a concrete way to say that $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ exist as ordered structures, among many alternative "isomorphic" existences.

## 16.A Defining the ordered integers

To construct the integers out of the natural numbers, we need a way to define negative numbers. When defining a negative number in a computer system, one easy way is to add an additional bit, the "sign bit", that tells the computer whether a number is positive or negative. Mathematicians likewise do this using ordered pairs. In other words, if we have access to two new symbols, like $+$ and $-$, then we can define the positive numbers as the ordered pairs $(+,1),(+,2),(+,3),\ldots$ and the negative numbers as the ordered pairs $(-,1),(-,2),(-,3),\ldots$. To avoid writing so many symbols, mathematicians write $(+,n)$ as $+n$, and $(-,n)$ as $-n$.

Actually, there is no need for *new* sign symbols. Any two symbols would work just as well for the signs. Thus, we could use 1 for negative and 2 for positive. (This is close to what is sometimes done in computer systems, except that the sign bit is 0 for positive integers.) Hence, we will take the negative integers to be

$$\mathbb{Z}_{<0} = \{1\} \times (\mathbb{N} - \{0\}) = \{(1,1),(1,2),(1,3),\ldots\},$$

and take the positive integers to be

$$\mathbb{Z}_{>0} = \{2\} \times (\mathbb{N} - \{0\}) = \{(2,1),(2,2),(2,3),\ldots\}.$$

This leaves the number zero, which is neither negative nor positive. As it doesn't matter which sign we use, we just pick the positive sign. Thus, we can define $\mathbb{Z}$ to be the set

$$\mathbb{Z} = \mathbb{Z}_{<0} \cup \mathbb{Z}_{\geq 0} = (\{1\} \times (\mathbb{N} - \{0\})) \cup (\{2\} \times \mathbb{N}).$$

Our next task is to define the usual order $<_{\mathbb{Z}}$ on $\mathbb{Z}$, given the usual order $<_{\mathbb{N}}$ on $\mathbb{N}$. The ordered structure we want is exactly

$$(\mathbb{N} - \{0\}, <_{\mathbb{N}}^{-1}) + (\mathbb{N}, <_{\mathbb{N}}),$$

where the addition of ordered sets was described in Subsection 13.A. This orders the negative numbers in the reversal of the way that the positive numbers are ordered, and it puts the negative numbers before all the nonnegative numbers. This is also the main reason it was convenient to take the "sign" of a negative number to be 1, and of a nonnegative number to be 2, because the addition of two ordered sets was accomplished in a similar manner.

As mentioned previously, there are many alternative constructions of (an order isomorphic copy of) the ordered set $(\mathbb{Z}, <_{\mathbb{Z}})$. Since we only care about the order of the elements in $\mathbb{Z}$, and not the symbols we use, there is a lot of freedom in constructing other versions of this ordered set; see Example 13.3 where two copies of 0 are included.

What if we do care about the names of the symbols? In particular, what if we want $\mathbb{N}$ to be a subset of $\mathbb{Z}$? This is easily accomplished by replacing $\mathbb{Z}_{\geq 0} = \{2\} \times \mathbb{N}$ with $\mathbb{N}$. (You can think of this as replacing a string of symbols, like $+4293$, with the new string 4293 where we dropped the positive sign symbol.) Of course, along with this replacement we also must accordingly modify the order relation on $\mathbb{Z}$, by using the new order relation

$$<_{\mathbb{Z}} = \{((1,m),(1,n)) \in \mathbb{Z}^2_{<0} \,:\, n <_{\mathbb{N}} m\} \cup (\mathbb{Z}_{<0} \times \mathbb{N}) \cup <_{\mathbb{N}} .$$

This construction can be repeated in formal set theory, by replacing $\mathbb{N}$ with $\omega$. Thus, we can correspondingly create a formal version of $\mathbb{Z}$ in set theory.

It is usually not important which isomorphic version of $\mathbb{Z}$ that one settles on using in everyday mathematical practice. This is because any other isomorphic copy satisfies exactly the same order properties in FOL. Most mathematicians introduce new symbols, like $+$ and $-$, whenever needed. However, it is sometimes convenient to know how to encode such symbols using ordered pairs.

## 16.B Defining the ordered rationals

By Theorem 15.4 we know that any two unbounded, countable, dense, totally ordered sets are order isomorphic. But do we know that there is such an ordered set? The answer is yes, after we define $\mathbb{Q}$.

To define $\mathbb{Q}$ we will encode fractions $m/n$ in their lowest terms. We would like to encode $m/n$ as the ordered pair $(m,n) \in S = \mathbb{Z} \times (\mathbb{N} - \{0\})$. However, the set $S$ itself is too big to be $\mathbb{Q}$, since there are pairs that represent fractions not in lowest terms, such at $(21, 14)$. So, we instead take

$$\mathbb{Q} = \{(m,n) \in S \,:\, \mathrm{GCD}(m,n) = 1\} \subseteq S.$$

This is a set, by an axiom of separation, once we can describe when a GCD is 1, which is easy after defining multiplication and addition on $\mathbb{Z}$. Defining these operations is simple, after we introduce formal recursion. Alternatively, see Exercise 16.4.

There is a technical note that should be made at this point. We want $\mathbb{Z}$ to be a subset of $\mathbb{Q}$, which is not currently the case. An easy way to fix this problem is to replace the "copy" of $\mathbb{Z}$ in $\mathbb{Q}$ with the actual thing. In other words, let

$$T = \{(m,n) \in S \,:\, \mathrm{GCD}(m,n) = 1 \text{ and } n \neq 1\},$$

which corresponds to $\mathbb{Q} - \mathbb{Z}$. As $\mathbb{Z}$ and $T$ are disjoint, we can take $\mathbb{Q} = \mathbb{Z} \cup T$.

Finally, we need to define the order relation $<_{\mathbb{Q}}$. Given $(a,b), (c,d) \in \mathbb{Q}$, we can say $(a,b) <_{\mathbb{Q}} (c,d)$ exactly when $ad <_{\mathbb{Z}} bc$.

## 16.C Defining the ordered reals

There are many constructions of $(\mathbb{R},<)$ in the literature. The most commonly taught construction is to take real numbers to be infinite decimal expansions. No matter how you define the reals, by Theorem 15.15 it will be order isomorphic to any other version. Here we will think of each real number as being determined by the set of rational numbers above and below it. These sets of rationals are called *Dedekind cuts*, so named after Dedekind who championed this construction of the reals, although the idea appears in the work of earlier mathematicians.

**Example 16.1.** One possible definition of the real number $\sqrt{2}$ is as the infinite decimal expansion $1.41421\ldots$. Alternatively, this number cuts the real line into the lower half $(-\infty,\sqrt{2})$ and the upper half $(\sqrt{2},\infty)$. The corresponding Dedekind cut is the pair of sets

$$(\mathbb{Q}\cap(-\infty,\sqrt{2}),\mathbb{Q}\cap(\sqrt{2},\infty)).$$

These two sets of rational numbers squeeze in on $\sqrt{2}$ from both sides. △

Formally, let $\mathbb{R}$ be the set of decompositions of $\mathbb{Q}$ into two nonempty subsets, where each element of the first set is smaller than each element of the second set, and the second set has no least element. In symbols,

$$\begin{aligned}\mathbb{R}=\{(A,B)\in(\mathscr{P}(\mathbb{Q})-\emptyset)^2 \ :\ & B \text{ has no least element, } A\cup B=\mathbb{Q},\\ & \text{and } x<_{\mathbb{Q}} y \text{ for any } x\in A \text{ and any } y\in B\}.\end{aligned}$$

Identify each rational number $r\in\mathbb{Q}$ with the pair $(\mathbb{Q}_{\leq r},\mathbb{Q}_{>r})\in\mathbb{R}$. The reason for assuming that the set $B$ has no least element is because otherwise we double count the rational numbers; we don't want to capture $r$ a second time with the pair $(\mathbb{Q}_{<r},\mathbb{Q}_{\geq r})$. (Why isn't this a problem for the irrational numbers?)

The order relation on $\mathbb{R}$ is extremely easy to define. We simply stipulate

$$(A,B)\leq_{\mathbb{R}}(C,D) \text{ exactly when } A\subseteq C.$$

We leave it to the reader to check that $(\mathbb{R},<_{\mathbb{R}})$ satisfies each of the hypotheses of Theorem 15.15.

## 16.D Exercises

**Exercise 16.1.** It was claimed that the construction of $\mathbb{Z}$ can be formalized by replacing $\mathbb{N}$ with $\omega$. Which of the axioms of formal set theory are used to make this happen?

**Exercise 16.2.** Check that no element of $\omega$ is an ordered pair of the form $(\underline{1},n)$ with $n\in\omega-\{\underline{0}\}$, so that there is no intersection between $\omega$ and the formal version of $\mathbb{Z}_{<0}$ given by $\{\underline{1}\}\times(\omega-\{\underline{0}\})$.

**Exercise 16.3.** In Subsection 16.B, the order relation on $\mathbb{Q}$ was defined using the "ordered pairs" definition of $\mathbb{Q}$ that doesn't technically include $\mathbb{Z}$. Change the definition of $<_{\mathbb{Q}}$ to work for the modified definition of $\mathbb{Q}$ that includes $\mathbb{Z}$. Further note that $<_{\mathbb{Q}}$ extends $<_{\mathbb{Z}}$ (i.e., as sets of ordered pairs, $<_{\mathbb{Q}}$ contains $<_{\mathbb{Z}}$).

**Exercise 16.4.** This exercise describes an alternative way to create an ordered set (which works in formal set theory) that is order isomorphic to $(\mathbb{Q}, <)$, without using any addition or multiplication facts in $\mathbb{Z}$.

(1) Let $S$ be the set of all functions $f\colon \omega \to \{\underline{0}, \underline{1}\}$ with finite support, meaning that $f(n) = \underline{0}$ whenever $n$ is large enough. Prove that $S$ is a countable set. (Can you do this using only the axioms of formal set theory?)

(2) Let $<$ be the "compare first difference" relation on $S$, which was defined in Subsection 13.C. Justify why $<$ is a set in formal set theory.

(3) Prove that $<$ is an unbounded, dense, total order when restricted to $S - \{f_0\}$, where $f_0 \in S$ is the constant function with output $\underline{0}$.

**Exercise 16.5.** Let $(\mathbb{R}, <_{\mathbb{R}})$ be as defined in Subsection 16.C. Finish the proof that this is an unbounded, complete, total order with a countable, dense subset.

**Exercise 16.6.** How do we modify the definition of $\mathbb{R}$ so that it contains $\mathbb{Q}$?

**Exercise 16.7.** Prove that the usual "infinite decimal expansion" definition of $\mathbb{R}$ is an unbounded, complete, totally ordered set with a countable, dense subset. (Note: This set does not contain $\mathbb{Q}$, since rational numbers are not defined as infinite decimals. How may we overcome this deficit?)

# 17 Well-founded relations

In this section we study the consequences of the following significant property for binary relations.

**Definition 17.1.** Let $R$ be a binary relation. We say that $R$ is *well-founded* if each nonempty set $S$ has an $R$-minimal element.

By an $R$-minimal element of $S$ we mean an element $x \in S$ that satisfies the condition $\forall y\,(y \in S \to \neg(yRx))$. This generalizes the previous definitions of "minimality" by not assuming any sort of order properties. However, we are still thinking of the left side of $R$ as the "strictly smaller" side.

**Example 17.2.** Let $R = \{(0,3),(1,1)\}$. If we take $S = \{3,4,5\}$, then every element of $S$ is an $R$-minimal element, since none of its elements appears as the first coordinate of any pair in $R$. The definition of well-foundedness does not presuppose that there is a *unique* $R$-minimal element of the set $S$.

If we take $S = \{0,1\}$, then this set also has an $R$-minimal element, which just happens to be unique. It is 0.

If we take $S = \{1\}$, then this set has no $R$-minimal element. Thus, $R$ is not a well-founded relation. Well-foundedness requires *all* nonempty sets to have $R$-minimal elements. △

**Example 17.3.** Let $R = \{(a,b),(b,a)\}$. It is not hard to show that every nonempty set $S$ has an $R$-minimal element, except $S = \{a,b\}$. △

**Example 17.4.** The usual strict order $<$ on $\mathbb{N}$ is well-founded; you were secretly asked to prove a formal version of this fact as part (4) of Exercise 3.8. This is called the "well-ordering property" of $\mathbb{N}$. Is the inverse relation $>$ on $\mathbb{N}$ well-founded? △

## 17.A Necessary conditions for well-foundedness

One might ask: Is there an easy way to check whether or not a relation is well-founded? There are some immediate simplifications that allow us to more easily check for this condition. First, we can limit our consideration to those sets contained in $\operatorname{ran}(R)$. (Any element $x \in S - \operatorname{ran}(R)$ is automatically $R$-minimal in $S$.) Also, if $S \cap \operatorname{dom}(R) = \emptyset$, then every element of $S$ is $R$-minimal. Well-founded relations also have two of the properties that we studied previously.

**Lemma 17.5.** *If $R$ is a well-founded relation, then $R$ is irreflexive and asymmetric.*

*Proof.* First we handle irreflexivity. Working contrapositively, let $(x,x) \in R$. Then taking $S = \{x\}$, we see that $S$ is a nonempty set without an $R$-minimal element.

Next, again working contrapositively, assume $(x,y),(y,x) \in R$. Then taking $S = \{x,y\}$, we see that $S$ is a nonempty set without an $R$-minimal element. □

Surprisingly, well-foundedness also behaves well with respect to transitivity.

**Lemma 17.6.** *A relation $R$ is well-founded if and only if its transitive closure is well-founded.*

*Proof.* ($\Leftarrow$): Let $S$ be a nonempty set. If $S$ has an $R^{\text{tran}}$-minimal element, then this is automatically an $R$-minimal element, because $R \subseteq R^{\text{tran}}$.

($\Rightarrow$): Contrapositively, assume $R^{\text{tran}}$ is not well-founded. Fix some nonempty set $S$ that does not have an $R^{\text{tran}}$-minimal element. Define the new set

$$T = \{y \in \operatorname{ran}(R) \ : \ \text{there is some } x \in S \text{ such that } xR^{\text{tran}}y\}.$$

It just remains to show that $T$ is nonempty and has no $R$-minimal element.

First note that $S \subseteq T$ (by the assumption on $S$), so $T \neq \emptyset$. Given any $y \in T$, fix some $x \in S$ such that $xR^{\text{tran}}y$. From the definition of the transitive closure of $R$, there is some $n \in \mathbb{N}$ and some elements $w_0, \ldots, w_n$, with $w_0 = x$, such that

$$w_0Rw_1,\ w_1Rw_2, \ldots, w_{n-1}Rw_n,\ w_nRy.$$

Then $w_nRy$ and $w_n \in T$ (even when $n = 0$). So $y$ is not $R$-minimal in $T$. □

An immediate corollary of the previous two lemmas is that any well-founded relation $R$ induces a unique well-founded partial order; namely, the transitive closure of $R$. This partial order does not need to be a total order as the following example easily illustrates.

**Example 17.7.** Let $R = \emptyset$ and let $M = \{1, 2\}$. The relation $R$ is a well-founded, partial order on $M$ that is not a total order. △

While we have discussed many different necessary conditions that a well-founded relation must satisfy, there is also an important condition related to decreasing "chains" of elements being related by $R$.

**Lemma 17.8.** *If there is a set $S = \{x_i \ : \ i \in \mathbb{N}\}$ with $(x_{n+1}, x_n) \in R$ for each $n \in \mathbb{N}$, then $R$ is not a well-founded binary relation.*

*Proof.* The nonempty set $S$ has no $R$-minimal element. □

We may informally think of the premise of the previous lemma as asserting that $\ldots Rx_2Rx_1Rx_0$. Surprisingly, the converse of Lemma 17.8 is valid, but only after assuming another axiom in our formal set theory. We will discuss this more fully at the end of this section, but the following piece of advice may be useful now.

**Advice 17.9.** Think of well-foundedness as saying that there cannot be an infinite "decreasing" sequence of related elements.

## 17.B Well-foundedness and trichotomy

We will now describe what happens when we additionally assume trichotomy.

**Definition 17.10.** A binary relation $<$ on a set $M$ that is both well-founded and a total ordering is called a *well-ordering*. The set $M$ is said to be *well-ordered* (by the relation $<$).

For finite sets, well-foundedness is automatic for total orders, as the following lemma shows.

**Lemma 17.11.** *Every total ordering on a finite set is a well-ordering.*

*Proof.* Let $(M, <)$ be a total ordering with $M$ finite. It suffices to show that every nonempty subset of $M$ has a minimal element. Any such subset is finite, so this follows from Lemma 12.17, but replacing "maximal" with "minimal". □

There are also infinite well-ordered sets. We already saw that $\mathbb{N}$ is well-ordered. Less trivially, $\mathbb{N} \cup \{\infty\}$ is well-ordered. Similarly, the sum $\mathbb{N} + \mathbb{N}$ is well-ordered. Is $\mathbb{Z}$ well-ordered? The answer is no, because there is an infinite decreasing chain $\ldots < -2 < -1 < 0$.

Any nonempty well-ordered set $M$ has a least element $m_0$. If this isn't the only element of $M$, then $M - \{m_0\}$ is nonempty, and thus also has a least element. In other words, $M$ will have a second smallest element $m_1$. If $M \neq \{m_0, m_1\}$, then there is a third smallest element. When $M$ is not finite, we can repeat this process to get an infinite list $m_0 < m_1 < m_2 < \ldots$. It is possible that this list doesn't completely exhaust $M$ either, and so there is a smallest element above these elements, and so forth. We will explore this idea more fully when we define ordinals in a future section.

It might appear that in order to guarantee that a relation is a well-ordering we need to posit that the relation is well-founded, transitive, and trichotomous. However, one of these conditions is implied by the other two.

**Proposition 17.12.** *If $R$ is a well-founded binary relation on $M$ that is trichotomous, then $R$ is transitive.*

*Proof.* Assume $a, b, c \in M$ with $aRb$ and $bRc$. We cannot have $a = c$, or else $R$ would not be well-founded by Example 17.3. (Can you find the infinite descending $R$-related sequence?) There is a similar problem if $cRa$ (and a similar infinite descending sequence). Thus, by trichotomy, we must have $aRc$. □

By Lemma 17.5 and the sentence immediately following Definition 12.1, we see that the three cases of trichotomy are always mutually exclusive for a well-founded relation. There is another form of uniqueness that well-orderings possess.

**Proposition 17.13.** *Let $R$ be a binary relation on a set $M$. The relation $R$ is a well-ordering of $M$ if and only if every nonempty subset $S \subseteq M$ has a least element.*

*Proof.* The forward direction is clear, since a minimal element of a set under a total ordering is a least element. For the backwards direction, assume that every nonempty subset $S \subseteq M$ has an $R$-least element. Tautologically, this tells us that $R$ is well-founded. Hence, $R$ is irreflexive and asymmetric by Lemma 17.5. If we can show that $R$ is trichotomous, then it will also be transitive by Proposition 17.12, and we'll be done. Suppose, by way of contradiction, that $R$ is not trichotomous. Fix $a, b \in M$ with $a \neq b$, as well as $\neg(aRb)$ and $\neg(bRa)$. The subset $S = \{a, b\}$ has two distinct $R$-minimal elements, contradicting the assumption. □

Thus, a well-ordering can be more simply described as a "uniquely well-founded relation".

If $(M, <)$ is a well-ordered set (or even just a well-founded poset), and $x \in M$ is some element, then we define the *initial segment* determined by $x$ as the set $I_{x,<} = \{y \in M : y < x\}$. Notice that the initial segment determined by the least element of $M$ is $\emptyset$. (In the well-ordered set $\mathbb{N} \cup \{\infty\}$, what is the initial segment of $x = \infty$?)

Initial segments are useful in comparing different well-orders on the same set. Let $(M, <_1)$ and $(M, <_2)$ be two well-orders. Putting $S = \{x \in M : I_{x,<_1} \neq I_{x,<_2}\}$, this set is empty if and only if $<_1$ and $<_2$ are the same well-order. If $<_1$ and $<_2$ are not the same, then $S$ is nonempty so it has both a $<_1$-least and $<_2$-least element; we can think of these elements as the points where the well-orders begin to differ.

## 17.C Decreasing sequences

We will now prove the converse of Lemma 17.8.

**Theorem 17.14.** *Let $R$ be a binary relation. If $R$ is not well-founded, then there exists a set $\{x_i : i \in \mathbb{N}\}$ with $(x_{n+1}, x_n) \in R$ for each $n \in \mathbb{N}$.*

*Proof.* Assume $R$ is not well-founded, and fix some nonempty set $S$ that has no $R$-minimal element. As $S$ is nonempty, we may pick some $x_0 \in S$. Since $x_0$ is not an $R$-minimal element of $S$, there exists some $x_1 \in S$ with $x_1 R x_0$. Similarly, since $x_1$ is not an $R$-minimal element of $S$, there exists some $x_2 \in S$ with $x_2 R x_1$. Repeating this process, we construct the set we seek. □

At first glance it might be unclear how this theorem uses any new set theoretic machinery. Thus, let us clarify a few issues.

First, an essential part of the proof is the claim that given a nonempty set $S$ we can "pick" an element from $S$. From a semantic point of view, this is what existential quantification precisely means. We have used this idea, without comment, whenever we've used the words "fix" or "take" or "put".

It is instructive to explain what is happening from the syntactic point of view. We reach a point in the deduction of the form $\vdash \exists x\, \varphi(x)$. At that point we add a new constant symbol $c$ to the language $\mathscr{L}$, to form a new language $\mathscr{L}(c)$. We then assert $\vdash \varphi(c)$ and continue the deduction in the new language $\mathscr{L}(c)$. If the last line of the deduction has no instance of $c$, and we never used universal generalization to quantify over a free variable of $\varphi(c)$, then the deduction is valid and could have been deduced without the introduction of the new constant symbol $c$ (but probably with a much longer deduction) because of the following result:

**Metatheorem 17.15.** *Suppose that $\exists x\, \varphi(x)$ and $\psi$ are statements of a FOL language $\mathscr{L}$ where $x$ is a free variable in $\varphi$. Further suppose that $T$ is a collection of statements from $\mathscr{L}$, and that $c$ is a constant symbol not in $\mathscr{L}$.*

*If $T \cup \{\varphi(c)\} \vdash_{\mathscr{L}(c)} \psi$ can be deduced without applying universal generalization to quantify over any free variables of $\varphi(c)$, then $T \cup \{\exists x\, \varphi(x)\} \vdash_{\mathscr{L}} \psi$.*

*Proof.* We work directly. The assumption about not applying universal generalization to free variables in $\varphi(c)$ allows us to use the Deduction Theorem 10.7, and so

$$T \vdash_{\mathscr{L}(c)} \varphi(c) \to \psi.$$

In the proof of this deduction, replace $c$ everywhere by a variable $y$ that does not appear anywhere in the proof; this gives a proof of $T \vdash_{\mathscr{L}} \varphi(y) \to \psi$. By universal generalization, we have $T \vdash_{\mathscr{L}} \forall y\, (\varphi(y) \to \psi)$. By a logical tautology (left to the reader), we have $T \vdash_{\mathscr{L}} (\exists x\, \varphi(x)) \to \psi$. By modus ponens, $T \cup \{\exists x\, \varphi(x)\} \vdash_{\mathscr{L}} \psi$. □

In the proof of Theorem 17.14, the ability to pick $x_0 \in S$ lets us claim that there exists a function $f\colon \{\underline{0}\} \to S$. The element $x_0$ is simply $f(\underline{0})$; equivalently, $f = \{(\underline{0}, x_0)\}$. In other words, $\{\underline{0}\} \times S$ is nonempty whenever $S$ is nonempty. This claim is rather undeniable, from a semantic perspective.

What leads to trouble in the proof of Theorem 17.14 is not the picking of each individual element, but doing an infinite number of picks all at once. In the last sentence of the proof it says "repeating this process" but there is no rule of logic that tells us we can repeat our picking process an infinite number of times.

You might be asking: "Doesn't that phrase just mean that we should apply induction?" It is true that similar phraseology is often used for induction, but the phrase isn't being used that way here. It is being used to say that we should repeat a meta process an infinite number of times.

Even so, doesn't induction imply that we can pick an element $x_n$, for each $n \in \mathbb{N}$? The answer is actually complicated. We must be very careful because the process of "picking" is a meta process, and so when we pick $x_0$, $x_1$, and so forth, the subscripts are a count in the metatheory, *not* in the formal set $\omega$ of natural numbers. So we have no way to collect all of the $x_n$ together in a formal set.

What happens if we try to rephrase the process of picking as an internal process in our formal set theory? Above, we showed that a single pick corresponds to the existence of a function $f\colon \{\underline{0}\} \to S$. Similarly, when we pick $x_{\underline{0}}$ and $x_{\underline{1}}$ satisfying

$x_{\underline{1}} R x_{\underline{0}}$, we can rephrase this in formal set there by saying that there exists a function $f\colon \{\underline{0}, \underline{1}\} \to S$ with $f(\underline{1}) R f(\underline{0})$. By (nonmeta, formal) induction, we *can* prove that for each (nonmeta) natural number $n \in \omega$, there is a function

$$f\colon \{\underline{0}, \underline{1}, \ldots, n\} \to S \text{ such that } f(m+1) R f(m) \text{ for each natural number } m < n.$$

What we have no way of doing is "completing" this process to obtain a function $g\colon \omega \to S$ with $g(m+1) R g(m)$ for each $m \in \omega$. This requires a new axiom.

You might be thinking: "If we are just a little more clever, can't we modify the induction argument to make the $f$'s compatible with each other all the way up to infinity, without using that new axiom?" The answer is no, but the proof is beyond the scope of this book. One can create a model of formal set theory that satisfies all of the usual axioms but fails this new axiom. We discuss this new axiom, and other related axioms, in greater detail in the next section.

## 17.D Exercises

**Exercise 17.1.** (This is an important exercise.) Prove that the sum of two well-ordered sets is well-ordered. (Is the sum of two well-founded sets well-founded?)

**Exercise 17.2.** If $(X, <)$ is a well-order, prove that the canonical order on $X^2$ is too.

**Exercise 17.3.** Give an example of a set $S$ that is well-ordered by the relation $\in$ on its elements, but $S$ is not a transitive set. (Recall Definition 3.4 for the definition of a transitive set. Try to make the set $S$ as simple as possible.)

**Exercise 17.4.** Prove that if $S$ is a transitive set on which $\in$ is a transitive relation, then any $x \in S$ has these same two properties. Give examples showing that if either of the two hypotheses is dropped, then the conditions in the conclusion can fail simultaneously.

**Exercise 17.5.** Prove that if $S$ is a transitive set on which $\in$ is well-founded, then the $\in$-relation on any $x \in S$ is also well-founded.

**Exercise 17.6.** In this exercise we will assume the existence of an unusual set $T = \{a_1, a_2, \ldots\}$, with $a_n = \{a_m : m > n\}$, and $a_i \neq a_j$ when $i \neq j$. This is a strange hypothesis. This is the type of set you should never encounter except possibly "in the wild". (We will later rule out this type of set by adjoining an additional axiom to our formal set theory.)

Show that the $\in$-relation on $T$ is a total ordering on $T$, but not a well-founded relation. Also show that $T$ is a transitive set. Show that $a_1 \in T$ also has these properties. (Thus the well-founded hypothesis in Exercise 17.5 is necessary, if $T$ exists.)

**Exercise 17.7.** (This may be a difficult exercise.) Prove that well-foundedness of a binary relation symbol $\prec$ is not axiomatizable in FOL using just the signature $(\prec)$. In other words, show that no collection of statements in that language holds exactly when $\prec$ is well-founded. (You might find Exercise 11.16 helpful. As bonus problems: (1) Show that this holds for any signature containing $\prec$. (2) Show that we can work under any theory $T$ that does not force $\prec$-sequences to be of bounded length.)

# 18 Axioms of choice

Consider the following seemingly obvious statements.

(1) Every infinite set has a countable subset.

(2) If $A$ and $B$ are sets, then either $|A| < |B|$, $|B| < |A|$, or $|A| = |B|$.

(3) A countable union of countable sets is countable.

If you carefully study the usual proofs of these facts, you will see that they somehow require an infinite number of choices. In this section we will introduce axioms in our formal set theory that formalize infinite picking processes. Also, we will use $\mathbb{N}$ (rather than $\omega$) for the formal set of natural numbers, and avoid using underline notation.

## 18.A Axiom of countable choice

Consider the following hypothetical situation. You are given a function $g$ whose domain is $\mathbb{N}$, and for each $n \in \mathbb{N}$ you know that $g(n) = S_n$ is a nonempty set. You would like to pick one element out of each of the sets $S_0, S_1, \ldots$. In other words, you wish to claim that there exists a function $f$ whose domain is $\mathbb{N}$, and for each $n \in \mathbb{N}$ it happens that $f(n) \in S_n$. The function $f$ represents the process of picking out the elements $f(0) \in S_0, f(1) \in S_1, \ldots$. We call it a *choice function*, because it allows us to choose an element from each $S_n$.

In some situations, claiming the existence of such a function $f$ is very easy. For instance, if $g(n) = \{0\}$ for each $n \in \mathbb{N}$, then there is only one option for $f$; we must take $f(n) = 0$ for each $n \in \mathbb{N}$. In other words the function $f = \mathbb{N} \times \{0\}$ satisfies the conditions we need, and it is the *only* function that does so.

In other situations it is still possible to claim the existence of such a function, because we can describe a rule that defines $f$. For instance, consider the function $g\colon \mathbb{N} \to \mathscr{P}(\mathbb{N})$ where $g(n)$ is the set of natural numbers with exactly $n$ prime factors (counting multiplicities). Thus

$$\begin{aligned} g(0) &= \{1\}, \\ g(1) &= \{2, 3, 5, 7, \ldots\}, \\ g(2) &= \{4, 6, 9, 10, \ldots\}, \end{aligned}$$

and so forth. One option to construct a choice function $f$ is to define $f(n)$ as the smallest element of $g(n)$.

However, if there is no way to choose elements from $S_n$ uniformly as $n$ varies, then while we may hope that a choice function $f$ should exist, there is nothing in our previous list of axioms of formal set theory that guarantees it does exist. The *axiom of countable choice* is exactly that guarantee. In symbols

$$\begin{aligned} \forall g \quad &\big((g \text{ is a function} \wedge (\mathrm{dom}(g) = \mathbb{N}) \wedge \emptyset \notin \mathrm{ran}(g)) \to \\ &\exists f\, (f \text{ is a function} \wedge (\mathrm{dom}(f) = \mathbb{N}) \wedge \forall n \in \mathbb{N}\, (f(n) \in g(n)))\big). \end{aligned} \tag{18.1}$$

The steps for turning this shorthand formula into an actual formula in the signature $(\in)$ are outlined in the exercises.

The axiom of countable choice has some immediate and important consequences. One of the most significant is the following result.

**Theorem 18.2.** *A countable union of countable disjoint sets is countable.*

*Proof.* Let $\{S_n : n \in \mathbb{N}\}$ be an arbitrary collection of disjoint countable sets. Because $S_n$ is countable, we may fix a bijection $h_n \colon S_n \to \mathbb{N} \times \{n\}$. Let $h = \bigcup_{n\in\mathbb{N}} h_n$, which is a bijection $\bigcup_{n\in\mathbb{N}} S_n \to \mathbb{N} \times \mathbb{N}$. There is a bijection $j \colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ (for instance, see Remark 13.8); composing $h$ with $j$ shows that the union is countable. □

Did you see where the axiom of countable choice was used? We use the axiom when we fix the bijections $h_n$; we fixed one such bijection for each $n \in \mathbb{N}$, and that's a countable number of choices. To be more precise, let $g$ be the function that takes a natural number $n$ to the set $A_n$ of all bijections $S_n \to \mathbb{N}\times\{n\}$. The axiom of countable choice guarantees the existence of a choice function $f$ such that $f(n) = h_n \in A_n$ for each $n \in \mathbb{N}$.

It is important to point out that often one does *not* need to use the axiom of countable choice to prove that a countable union of countable disjoint sets is countable. This is the case, for instance, if along with the sets $S_n$ you are given an explicit bijection $\mathbb{N} \to S_n$; now you don't need to pick a random bijection, but you can instead use the one given to you.

Perhaps an even more fundamental fact that we take for granted about sets is that every infinite set has a countable subset. This too requires an infinite amount of fixing. We will now give the proof, with some extra details included as parenthetical comments. It is recommended that after you've read through the proof that you do so a second time, but skip those comments.

**Theorem 18.3.** *Every infinite set has a countable subset.*

*Proof.* Let $S$ be an infinite set. For each (formal) natural number $n \in \mathbb{N}$, let $A_n$ be the set of all injective functions $h \colon n \to S$. (We know that $\mathscr{P}(\mathbb{N} \times S)$ is a set by our work in Subsection 4.A. Further, the condition "$h$ is an injective function with domain $n$ and codomain $S$" is expressible by a first order formula $\varphi(h, n, S)$; see Exercise 18.2 for hints how to do this. Thus, $A_n = \{h \in \mathscr{P}(\mathbb{N} \times S) : \varphi(h, n, S)\}$ is a set by an axiom of separation.) That each $A_n$ is nonempty is proved by induction.

Let $g$ be the function that takes each $n \in \mathbb{N}$ to $A_n$. (We can describe $g$ as the collection $\{(x, y) \in \mathbb{N} \times \mathscr{P}(\mathscr{P}(\mathbb{N} \times S)) : x \in \mathbb{N} \text{ and } y = A_x\}$, which is a set by separation. The fact that we can express $(x \in \mathbb{N}) \wedge (y = A_x)$ by a first order formula follows quickly from our work in the previous paragraph. Hereafter we will avoid explicitly spelling out these types of considerations, and assume that the definitions we give for functions are sufficient.)

Now, by the axiom of countable choice there is a function $f$ with $f(n) = h_n \in A_n$ for each $n \in \mathbb{N}$. Define a new function $p \colon \mathbb{N} \to S$ by the rule that $p(n)$ is the first value of $h_n$ that is not in $\{p(i) : i < n\}$. The function $p$ enumerates a countable subset of $S$, as desired. □

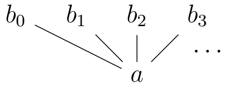
## 18.B Axiom of dependent choice

The proof given for Theorem 18.3 may seem overly complicated, since a countable subset of any infinite set $S$ could have been chosen as follows. First, since $S$ is infinite it must have some element $x_0 \in S$. Now $S - \{x_0\}$ is still infinite, so there exists some element $x_1 \in S - \{x_0\}$. Repeating this process we have the countable subset $\{x_0, x_1, \ldots\}$.

This idea does indeed simplify the proof of Theorem 18.3, but it uses a principle stronger than countable choice. We are not only making a countable number of choices, but each successive choice depends on the previous choices. This stronger principle is called the *axiom of dependent choice*, or *DC* for short. We will forgo writing out the formal statement of this axiom.

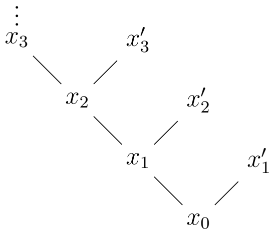
The axiom of dependent choice was the principle we used previously to prove Theorem 17.14, which says that any binary relation $R$ that is not well-founded has an infinite decreasing $R$-sequence. Here is one more interesting example of this principle in action.

A *tree* is a poset $(T, <)$ with a least element (called the *root*) and every initial segment in the tree is finite and linearly ordered. A *branch* in $T$ is a maximal linearly ordered subset. For example, $\mathbb{N}$ is a tree with exactly one, infinitely long branch. Some trees can be short but very wide, such as in the picture below.



Each branch in this tree is just $\{a, b_n\}$ for some $n \in \mathbb{N}$. (Can you think of a wider tree?) We think of the set of elements connected directly above a point as the *children* of that point (which is opposite of how things work in a family tree, but, chronologically speaking, the branches of an actual tree grow upwards).

Not all branches have to be the same length. Consider the following tree.



It has a branch of every positive length, as well as one infinite branch.

Trees with arbitrarily long branches do not necessarily have any infinite branches. However, we can guarantee that certain trees always have infinite branches.

**Lemma 18.4** (Kőnig's tree lemma)**.** *If $(T, <)$ is a tree with arbitrarily long branches, and each element has only finitely many children, then there is an infinite branch.*

*Proof.* Let $x_0$ be the root of the tree. Among the finitely many children of $x_0$, at least one of them belongs to arbitrarily long branches; fix such a point and call it $x_1$. Similarly, among the finitely many children of $x_1$, at least one of them belongs to arbitrarily long branches (since $x_1$ does); fix such a point and call it $x_2$. Repeating this process, we have constructed an infinite branch $\{x_0, x_1, x_2, \ldots\}$. □

It may seem surprising that a result about objects as simple as trees apparently requires a special choice axiom for its proof. There are many similar results in formal set theory.

## 18.C Axiom of choice

One might rightly ask why we have been limiting our choice axioms to only countable choices. Indeed, the full *axiom of choice*, or *AC* for short, makes no such restriction. It asserts that given any collection (countable or not) of nonempty sets, we can always pick one element from each of those sets. In symbols

$$\forall X \Big(\emptyset \notin X \rightarrow \exists f \Big(f \text{ is a function} \wedge (\operatorname{dom}(f) = X) \wedge \forall x \in X\,(f(x) \in x)\Big)\Big).$$

Any function $f$ whose existence is asserted in this axiom is called a *choice function* for the set $X$. The existence of a choice function for certain sets does not require the axiom of choice, as we saw previously. The axiom of choice is generally necessary only when there is no formulaic way to choose elements.

For many mathematicians, the axiom of choice is obvious—of course you can pick elements out of nonempty sets. For others of a more constructivist bent, the axiom is dubious—it asserts the existence of a choice function without explicitly detailing which ordered pairs should be its members. For many others, they accept it as a useful but not necessarily physically real principle—making infinitely many arbitrary choices is a purely mental task not reflected in reality. However you feel about AC, it is commonly used in mathematical arguments.

When the axiom was first introduced, many people immediately accepted it as a good principle. However, surprising consequences of the axiom started appearing. One of the most startling is the following result of Zermelo.

**Theorem 18.5.** *If AC holds, then every set has a well-ordering.*

*Proof sketch.* Let $S$ be a set, and take $X = \mathscr{P}(S) - \{\emptyset\}$. Fix a choice function $f$ for $X$. Using the choice function $f$ we will construct a well-ordering on $S$. The idea is to take the least element of $S$ to be $x_0 := f(S)$, to then take the second smallest element of $S$ to be $x_1 := f(S - \{x_0\})$, and repeat until we run out of elements of $S$ (which may be only after a very long time; note that standard induction doesn't work here because $S$ could be uncountable). We will formalize the idea of "repeating" this process in the next section, but Exercise 18.8 outlines another argument. □

Surprisingly, the converse of Theorem 18.5 is true as well.

**Proposition 18.6.** *If every set has a well-ordering, then AC holds.*

*Proof.* Let $X$ be a set, and assume $\emptyset \notin X$. Fix a well-ordering $<$ on $\bigcup X$. Define a choice function $f\colon X \to \bigcup X$ by taking $f(x)$ to be the $<$-least element of $x$. □

Hereafter, we will assume that AC holds, unless we explicitly say otherwise. We end this subsection with the following small remark: Almost every result in "Transition" can be proved without any choice axiom. However, Exercise 31.7 from "Transition" is an exception. That exercise asserts the following proposition, which is left for the reader to prove.

**Proposition 18.7.** *Let $A$ and $B$ be nonempty sets. There is an injection $A \to B$ if and only if there is a surjection $B \to A$.*

## 18.D Axiom of global choice

After some minor tweaks, the axiom of choice can be shortened to read

$$\forall X\, \exists f\, (f \text{ is a choice function for } X - \{\emptyset\}).$$

An interesting question is whether or not we can switch the order of those quantifiers. Could there exist a function $f$ that acts as a choice function for all collections of nonempty sets, all at once? If so, we say that there is a *global* choice function.

Any global choice function is a *class*, rather than a set, since its domain is a proper class. In our formal set theory we only quantify over sets, not classes, so it becomes difficult (thought not entirely impossible, with some minor modifications) to assert the existence of a global choice function. In other versions of formal set theory, such as those that do allow quantification over both classes and sets, the task of asserting the existence of a global choice function is trivial.

The axiom of global choice does not arise often, if at all, in mathematical disciplines outside of logic and set theory. We will never use it, nor will we accept it as an axiom of our formal set theory.

## 18.E Exercises

**Exercise 18.1.** Write a well-formed formula in the signature $(\in)$ that represents the shorthand formula $\forall x \in S,\ \varphi(x)$. Do the same for $\exists x \in S,\ \varphi(x)$.

**Exercise 18.2.** Let $x$, $a$, $b$, and $f$ be free variables throughout this exercise. Write a well-formed formula, in the signature $(\in)$, that represents the expression "$x = \{a\}$". Do the same for "$x = \{a, b\}$". We can now define the expression "$x$ is the ordered pair $(a, b)$" as:

$$\exists u\, \exists v\, ((x = \{u, v\}) \wedge (u = \{a\}) \wedge (v = \{a, b\})).$$

Now give formulas for the expressions "$x$ is an ordered pair", "$f$ is a relation (with $\mathrm{dom}(f) = a$ and $\mathrm{ran}(f) = b$)", and "$f$ is a function".

**Exercise 18.3.** Prove that a set $S$ is Dedekind-infinite if and only if $S$ contains a countable subset. (Hint: For the harder direction, let $f\colon S \to T \subsetneq S$ be an injective function. Fix some $x_0 \in S - T$, and consider the set $\{x_0, f(x_0), f(f(x_0)), \ldots\}$. Do we need any axiom of choice for this part of the exercise?) Conclude that the Dedekind-infinite sets are exactly the infinite sets, assuming AC.

**Exercise 18.4.** (This exercise is easy to do incorrectly.) Without using any choice axioms prove that if $S$ is a countable subset of $\mathbb{R}$, then $\mathbb{R} - S$ has a countable subset. (Hint: It might be easier to replace $\mathbb{R}$ with the set of all functions $\mathbb{N} \to \{0, 1\}$, i.e., the set of all $\mathbb{N}$-sequences of 0's and 1's.)

**Exercise 18.5.** Construct an example of a tree with arbitrarily long branches, but with no infinite branch.

**Exercise 18.6.** Justify why the subset found at the end of the proof of Kőnig's tree lemma, Lemma 18.4, is a branch (i.e., a *maximal* linearly ordered subset of the tree).

**Exercise 18.7.** Prove Proposition 18.7. Explicitly mention where you use AC.

**Exercise 18.8.** Fill in the details for a proof of Theorem 18.5, as follows. Let $W$ be the set of all well-orderings $(T, <_T)$ where $T \subseteq S$ and $x = f(S - I_{x,<_T})$ for each $x \in T$. (Recall that $I_{x,<_T}$ is the initial segment determined by $x$.)

(1) Prove that $W$ is not empty (even if $S$ is empty).
(2) Check that if $<_{T_1}, <_{T_2} \in W$, then one them extends the other. (Hint: You might find Proposition 20.6 and the idea behind its proof useful.)
(3) Defining $<$ to be $\bigcup W$, show that $<$ is a well-ordering of some subset of $S$.
(4) Show that if $<$ well-orders a *proper* subset of $S$, then we can extend to an even larger well-ordering, giving us a contradiction.

# 19 Ordinals

There are (at least) two types of numbers in the English language. The first list of numbers goes “zero, one, two, three” and so forth. These are the cardinal numbers, and they denote quantity. The other list of numbers goes “zeroth, first, second, third” and so forth. These are the ordinal numbers, and they denote order. In this section we generalize the notion of ordinal numbers to the infinite case. This can be confusing to students learning this subject for the first time, because infinite ordinals behave quite differently than infinite cardinals. We will point out differences as they arise.

## 19.A Motivating idea

There is exactly one set with zero elements, $\emptyset$. This is one reason we decided to take $0 = \emptyset$ in formal set theory. However, there are many sets with exactly one element. (Indeed, so many that they form a proper class.) Examples include $\{\mathbb{N}\}$, $\{8403\}$, and $\{\pi\}$. If you were asked to pick the simplest example of such a set, perhaps the set you would choose is $1 = \{\emptyset\} = \{0\}$.

Similarly, there are many sets with two elements. The simplest example seems to be $2 = \{\emptyset, \{\emptyset\}\} = \{0, 1\}$. When we get to sets with three elements, there are two sets that naturally vie for the title “best set with three elements”; namely, either

$$\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \text{ or } \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}. \tag{19.1}$$

They are both quite simple. At first glance, the set on the right seems to be the simpler of the two. However, there are two reasons to favor the set on the left.

**Difference #1**: The set on the right in (19.1) is constructed by the (meta)operation “adjoin an element with one more pair of set braces”. So we get the sequence of sets

$$S_0 = \emptyset,\ S_1 = \{\emptyset\},\ S_2 = \{\emptyset, \{\emptyset\}\},\ S_3 = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\},\ \ldots.$$

If we were to “reach infinity” we would have the set $S_\infty = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots\}$. At this point it seems that the process cannot be continued. We can no longer just increase the number of set braces and adjoin that element.

The set on the left in (19.1) is constructed by the (meta)operation “take the collection of all the previous sets, and put them together”. So we get the sequence of sets

$$0 = \emptyset,\ 1 = \{0\},\ 2 = \{0, 1\},\ 3 = \{0, 1, 2\},\ \ldots.$$

When we “reach infinity” we have the set $\mathbb{N} = \{0, 1, 2, \ldots\}$. We use the symbol “$\omega$” rather than “$\mathbb{N}$” when emphasizing this (meta)process (or to differentiate from the meta numbers). At this point the process *can* be continued. If we take the collection of all the previous sets we have constructed, then we can write $\omega + 1 = \{0, 1, 2, \ldots, \omega\}$. The process can be continued another time, and we have $\omega + 2 = \{0, 1, 2, \ldots, \omega, \omega + 1\}$, and so forth.

**Difference #2**: The set $3 = \{0, 1, 2\}$ on the left in (19.1) has the added advantage that the relation $\in$ on its elements is trichotomous and transitive, while $\in$ is neither trichotomous nor transitive for the set $S_3$ on the right. As $\in$ is also well-founded on 3, this means that the elements of 3 are well-ordered by $\in$.

Thus the set 3 does two jobs at once. It contains three elements, and hence it *counts* that cardinal number, but it is also an *ordered* set and so it also corresponds to an ordinal number.

**Warning 19.2.** It can be dangerous to overuse a symbol in this way, especially when the two concepts (ordinal and cardinal) are quite similar and yet distinct. We must be careful to make the distinction clear when the two concepts differ.

Recall that $\aleph_0$ is the symbol used for the cardinality of $\omega$. By Theorem 18.3, every infinite set has a countable subset, and thus $\aleph_0$ is the smallest infinite cardinality. (Without any choice axiom, this seemingly obvious fact is not necessarily true.) All the other infinite cardinalities are uncountable. On the other hand, we wrote $\omega$ above for the first infinite ordinal; its cardinality is just $\aleph_0$. However, the next infinite ordinal is $\omega + 1 = \{0, 1, 2, \dots, \omega\}$, which is still countable. Thus, the ordinals grow much more slowly than the cardinal numbers. Ordinals measure well-orderings, and there are many countable well-orderings that are not order isomorphic. (Is $\omega + \omega$ countable?)

## 19.B Defining the ordinals

We have used the word "transitive" in two distinct ways, both as a qualifier for relations and for sets. These notions are somewhat connected, which is why they share a name. Recall that by Definition 3.4, a set $S$ is transitive if whenever $x \in y$ and $y \in S$, then $x \in S$ (for arbitrary sets $x, y$). Thus, a set $S$ is transitive if and only if the transitive property for relations, $(xRy \wedge yRz) \to xRz$, holds in the case when $R = \in$ and $z = S$.

Our definition of an ordinal number should capture the (meta)operation of "collect together all the previously constructed ordinals". Thus, an ordinal should be a transitive set. Also, the elements of an ordinal should naturally be well-ordered. This leads us to make the following definition.

**Definition 19.3.** A set $\alpha$ is an *ordinal* if $\alpha$ is a transitive set and it is well-ordered by the membership relation $\in$ (restricted to $\alpha \times \alpha$).

It is common to denote ordinals using the initial letters $\alpha, \beta, \gamma, \dots$ of the lowercase Greek alphabet, and we will follow that convention. It is easy to check that 0 is an ordinal. Less trivially, each natural number is an ordinal; this quickly follows from Exercises 3.5 and 3.8. Of course, those same exercises also imply that $\omega$ is an ordinal. We denote the collection of all ordinals by Ord. (This will turn out to be a proper class.)

Just as we did with the natural numbers, we define a relation $<$ on Ord by saying that $\alpha < \beta$ exactly when $\alpha \in \beta$. Of course, for any set $S$ it holds that $S = \{x : x \in S\}$. Thus, as long as all the elements of $\beta$ are ordinals we will have $\beta = \{\alpha \in \mathrm{Ord} : \alpha < \beta\}$, which matches our motivating idea of constructing each ordinal as the set of the "previous" ordinals. The following lemma proves this claim.

**Lemma 19.4.** *If $\beta \in \mathrm{Ord}$ and $\alpha \in \beta$, then $\alpha \in \mathrm{Ord}$.*

*Proof.* We must show that $\alpha$ is a transitive set and every nonempty subset has a unique $\in$-minimal element.

First we show $\alpha$ is a transitive set. Suppose $x \in y$ and $y \in \alpha$, for some sets $x$ and $y$. Since $\beta$ is a transitive set, we have $y \in \beta$, and then $x \in \beta$. As $\in$ is a transitive relation on the elements of $\beta$, we have $x \in \alpha$.

Now, let $S$ be an arbitrary nonempty subset of $\alpha$. Each $x \in S$ is also an element of $\beta$ (by the argument in the previous paragraph). Thus $S \subseteq \beta$. As every nonempty subset of $\beta$ has a unique $\in$-minimal element, we are done. □

Intuitively, the first few ordinals should be

$$(19.5)\quad 0 < 1 < 2 < 3 < \ldots < \omega < \omega + 1 < \omega + 2 < \ldots < \omega + \omega < \omega + \omega + 1 < \ldots,$$

and we will build up the machinery to prove these facts, as well as show that the relation $<$ on the class Ord is a well-ordering of that class. First, we prove that ordinals are well-behaved with respect to intersections.

**Lemma Schema 19.6.** *If $C \neq \emptyset$ is a class of ordinals, then $\bigcap C \in \mathrm{Ord}$.*

*Proof.* Let $\alpha = \bigcap C$, which is a set by Exercise 2.9. We first show that $\alpha$ is a transitive set. Suppose $x \in y$ and $y \in \alpha$, for some sets $x$ and $y$. For each $\beta \in C$ we have $\alpha \subseteq \beta$, and hence $y \in \beta$. Since each such $\beta$ is a transitive set, we have $x \in \beta$. As $\beta \in C$ was arbitrary, $x \in \alpha$.

Next, we show that every nonempty subset $S \subseteq \alpha$ has a unique $\in$-minimal element. Since $C$ is nonempty, fix some $\beta \in C$. As $\alpha \subseteq \beta$, we have $S \subseteq \beta$. Since $\beta$ is an ordinal, $S$ has the requisite property. □

We will now show that $\in$ is a trichotomous relation on Ord.

**Proposition 19.7.** *Let $\alpha, \beta \in \mathrm{Ord}$.*

(1) *If $\alpha \subsetneq \beta$, then $\alpha \in \beta$.*
(2) *Either $\alpha = \beta$, $\alpha \subsetneq \beta$ or $\beta \subsetneq \alpha$.*
(3) *Either $\alpha = \beta$, $\alpha \in \beta$, or $\beta \in \alpha$.*

*Proof.* (1) Let $y$ be the $\in$-minimal element of $\beta - \alpha$. For any $x \in y$, then $x \in \beta$, as $\beta$ is a transitive set. From the minimality condition on $y$, we have $x \in \alpha$. Thus $y \subseteq \alpha$.

We will now prove that $\alpha \subseteq y$. Let $x \in \alpha$. If $y = x$ then $y \in \alpha$, which contradicts the definition of $y$. Similarly, if $y \in x$ then $y \in \alpha$ (by transitivity of $\alpha$), which is impossible. So, since $\beta$ is well-ordered by $\in$, and $x, y \in \beta$, we must have $x \in y$. Thus $\alpha \subseteq y$. This proves that $\alpha = y \in \beta$.

(2) By Lemma Schema 19.6, $\gamma := \alpha \cap \beta \in \mathrm{Ord}$. Suppose, by way of contradiction, that $\gamma \neq \alpha$ and $\gamma \neq \beta$. Part (1) implies that $\gamma \in \alpha$ and $\gamma \in \beta$. But then $\gamma \in \gamma$, contradicting the fact that $\in$ is irreflexive on ordinals. (Why is it irreflexive?)

(3) This part follows from parts (1) and (2). □

**Theorem Schema 19.8.** *The relation $<$ on* Ord *is a well-ordering.*

*Proof.* We just proved that $<$ is trichotomous in Proposition 19.7. Thus, by both Lemma 17.5 and Proposition 17.12, it suffices to show that $<$ is a well-founded relation on the class Ord.

We interpret well-foundedness (on a proper class) as a schema of statements. Given any nonempty class $C$ of ordinals, let $\alpha = \bigcap C$. We know that $\alpha \in \mathrm{Ord}$ by Lemma Schema 19.6. We claim that $\alpha \in C$. If not, then for each $\beta \in C$ we have $\alpha \subsetneq \beta$, and hence $\alpha \in \beta$ by Proposition 19.7(1). Thus $\alpha \in \bigcap C = \alpha$, contradicting the fact that $\in$ is irreflexive on ordinals.

Finally, we claim that $\alpha$ is an $\in$-minimal element of $C$. If not, fix some $\beta \in C$ with $\beta \in \alpha$. As $\alpha = \bigcap C$, we have $\alpha \subseteq \beta$. Case 1: If $\alpha \neq \beta$, then Proposition 19.7(1) gives us $\alpha \in \beta$. By transitivity, $\alpha \in \alpha$, a contradiction as before. Case 2: If $\alpha = \beta$, then again we have $\alpha \in \alpha$, which is impossible. □

Notice that the proof of Theorem Schema 19.8 improves Lemma Schema 19.6, showing that the intersection of any nonempty class of ordinals is the unique minimal ordinal belonging to that class.

Now that we know Ord is well-ordered, it has a least element. Proposition 19.7 tells us that smaller ordinals are not only elements of the larger ones, but also subsets. Thus $0 = \emptyset$ is the least ordinal. It is left to the readers to justify to themselves that the first few ordinals are exactly as in (19.5).

## 19.C Exercises

Ordinals split into two types. An ordinal $\beta$ is a *successor ordinal* if $\beta = \alpha + 1 = \alpha \cup \{\alpha\}$ for some $\alpha \in \mathrm{Ord}$. The ordinals that are not successors are called *limit ordinals*. The next few exercises work out some easy properties of these concepts.

**Exercise 19.1.** Prove that 0 is a limit ordinal, and prove that all of the other (formal) natural numbers are successors.

**Exercise 19.2.** For any $\alpha \in \mathrm{Ord}$, prove that $\alpha + 1$ is the greatest lower bound on the class of ordinals strictly larger than $\alpha$. Thus, $\alpha + 1 = \bigcap\{\beta \in \mathrm{Ord} : \beta > \alpha\}$.

**Exercise 19.3.** Given $\beta \in \mathrm{Ord}$, prove that $\beta$ is a limit ordinal if and only if $\alpha < \beta$ implies $\alpha + 1 < \beta$ (for each $\alpha \in \mathrm{Ord}$).

**Exercise 19.4.** Prove that $\omega$ is the smallest nonzero limit ordinal.

**Exercise 19.5.** Prove that Ord is a proper class. (Hint: If not, it is an ordinal.)

**Exercise 19.6.** Prove that there are uncountably many countable ordinals. (More precisely, there are exactly $\aleph_1$ countable ordinals, where $\aleph_1$ is the cardinality of the first uncountable ordinal.)

We defined an ordinal as a transitive set that is well-ordered by the membership relation $\in$. From our work in Section 17, we know that a relation is a well-ordering if and only if it is well-founded and trichotomous. Thus, an ordinal is defined by three basic conditions:

(1) the set is transitive,
(2) $\in$ is a well-founded relation on the set, and
(3) $\in$ is trichotomous relation on the set.

Exercise 17.3 tells us that condition (1) does not follow from the other two parts of the definition. By Exercise 17.6 (and assuming the existence of an unusual set), we see that (2) does not follow from the other two parts of the definition of an ordinal. The next exercise shows that (3) similarly doesn't follow from the other two parts.

**Exercise 19.7.** Find an example of a transitive set $S$, on which $\in$ is a well-founded but not trichotomous relation.

**Exercise 19.8.** By Proposition 17.12, any trichotomous relation is transitive when it is well-founded. Trichotomy does not follow from well-foundedness and transitivity of the relation (for an easy example, see Example 17.7). However, show that any transitive set $S$ on which $\in$ is a well-founded and transitive relation is an ordinal. Thus the definition of ordinals can be weakened somewhat. (Hint: First consider the case when $S$ contains only ordinals. You may also find two exercises in Section 17 useful.)

**Exercise 19.9.** Let $S$ be any set of ordinals. Prove that $\bigcup S \in \mathrm{Ord}$. Also prove that this union is the least upper bound in Ord for the elements of $S$.

# 20 ZFC

There is one final axiom of formal set theory that we have not yet stated. It rarely, if ever, appears in arguments outside of set theory. Its main purpose is to simplify the structure of sets, allowing for easier proofs of metatheorems. It is the *axiom of regularity* and it asserts that nonempty sets have $\in$-minimal elements. Formally,

$$\forall S\,(S \neq \emptyset \rightarrow \exists x \in S\,(S \cap x = \emptyset)). \tag{20.1}$$

This means that we treat the membership relation $\in$ as if it is well-founded. Thus, the axiom of regularity rules out strange behaviors like $x \in x$. Even more generally, from our work in Section 17 we know that there can be no decreasing sequence, indexed by the formal natural numbers, of the form

$$\ldots \in x_2 \in x_1 \in x_0.$$

The $\in$-relation is a proper class. The axiom of regularity tells us that it is well-founded when restricted to any *set*. Is it also well-founded on proper classes? The answer is yes.

**Theorem Schema 20.2.** *Every nonempty class has an $\in$-minimal element.*

*Proof.* Let $C$ be a nonempty class, and let $S \in C$. If $S \cap C = \emptyset$ then $S$ is an $\in$-minimal element. So we may assume $S \cap C \neq \emptyset$.

Let $T_0 = S$ and recursively let $T_{n+1} = \bigcup T_n$ for each $n \in \omega$. The set $T = \bigcup_{n\in\omega} T_n$ is transitive, and it contains $S$. (It is called the *transitive closure* of $S$, not to be confused with the transitive closure of a relation.) Now, consider $T \cap C$, which is a nonempty set. By the axiom of regularity, we may fix an $\in$-minimal element $x \in T \cap C$.

Suppose, by way of contradiction, that $x$ is not $\in$-minimal in $C$, say $y \in x \cap C$. Since $T$ is transitive, $y \in T \cap C$, contradicting the minimality assumption on $x$. □

The axioms of formal set theory are the following.

(1) Extensionality.
(2) Empty set.
(3) Pairing.
(4) Union.
(5) Infinity.
(6) Power set.
(7) Choice.
(8) Separation (scheme).
(9) Replacement (scheme).
(10) Regularity.

The axioms of empty set and separation follow from the others, so they can be removed from the list, if wanted. In the signature $(\in)$, the first order theory of these axioms is denoted ZFC, which is short for "Zermelo-Fraenkel set theory with choice".

This is an infinite list of axioms, as replacement is an infinite scheme. It is a deep result, which we will not prove in this book, that ZFC has no finite axiomatization. Interestingly, there are stronger theories that can encode all of ZFC and that are finitely axiomatizable. If you are interested in learning about such systems, one especially nice example is NBG (short for "von Neumann, Bernays, Gödel set theory"), where classes are treated as formal objects alongside sets (and no longer correspond to formulas in the language).

We will describe the replacement axioms—and their consequences—more thoroughly in the next chapter. We end this section by describing a few more important facts about formal sets.

## 20.A Algebra with ordinals

Let $\alpha, \beta \in \mathrm{Ord}$. By Exercise 17.1, the sum of two well-ordered sets is well-ordered. So (by a small overuse of notation) we let $\alpha + \beta$ denote the unique ordinal with the same order type.

**Lemma 20.3.** *If $\alpha \in \mathrm{Ord}$, then $\alpha + 0 = \alpha$.*

*Proof.* By definition, $\alpha + 0$ as a set is $(\{1\} \times \alpha) \cup (\{2\} \times \emptyset) = \{1\} \times \alpha$. For any $(1, x), (1, y) \in \alpha + 0$, the order relation $(1, x) \leq (1, y)$ holds exactly when $x \in y$. Thus, the map $\alpha \to \{1\} \times \alpha$ given by $x \mapsto (1, x)$ is an order isomorphism. □

We leave it as an exercise that $\alpha + 1 = \alpha \cup \{\alpha\}$, which is called $\alpha$'s *successor*.

It is important to notice that addition is *not* commutative in general. For instance, we find $1 + \omega = \omega \neq \omega + 1$. However, addition is associative. Proving this fact is left for the exercises.

Multiplication is thought of as repeated addition. Informally, $\alpha \cdot \beta$ is the ordinal that we get by replacing each point in $\beta$ by a copy of $\alpha$. So, $\omega 2 = \omega + \omega$; we add together 2 copies of $\omega$. On the other hand

$$2\omega = \underbrace{2 + 2 + 2 + \cdots}_{\omega \text{ times}} = \omega.$$

So we see that multiplication is not generally commutative. However, just as for addition, it will be associative. Formally, we let $\alpha \cdot \beta$ denote the unique ordinal with the same order type as the lexicographical order on $\beta \times \alpha$ (which will be a well-ordered set).

We will see, in a later section, that there is an easier way to define ordinal addition and multiplication, even though the definition will be less "concrete".

## 20.B Order isomorphisms

The goal of this subsection is to show that every well-ordered set is order isomorphic to exactly one ordinal. Thus the ordinals are a convenient and, in some sense, "canonical" way to represent well-orders.

First, we will investigate the strictly increasing maps from a well-ordered set to itself. They are quite restricted.

**Lemma 20.4.** *If $(M, <)$ is a well-ordering, and $f\colon M \to M$ is strictly increasing, then $x \le f(x)$ for each $x \in M$.*

*Proof.* Let $S = \{x \in M : f(x) < x\}$. If $S$ is empty, we are done. Otherwise $S$ has a minimal element by well-foundedness. Call such an element $y$, so we have $f(y) < y$. Applying $f$ a second time, we have $f(f(y)) < f(y)$. This contradicts the minimality hypothesis on $y$. □

**Corollary 20.5.** *There is at most one order isomorphism between two well-ordered sets.*

*Proof.* Suppose that $(M, <)$ and $(N, <)$ are two well-ordered sets. Let $f, g\colon M \to N$ be order isomorphisms. Notice that $g^{-1} \circ f\colon M \to M$ is also an order isomorphism. If we can show that any order isomorphism $h\colon M \to M$ is the identity map, then $g^{-1} \circ f$ is the identity map, and so $f = g$ as desired.

Let $x \in M$. By Lemma 20.4, we have $x \le h(x)$. But $h^{-1}$ is also an order isomorphism, so we also have $x \le h^{-1}(x)$, or in other words $h(x) \le x$. Putting these inequalities together, we get $h(x) = x$. □

We are now ready to prove a type of trichotomy among well-ordered sets. The idea of the proof will be to compare two well-ordered sets, starting with their least elements, building up an order isomorphism one element at a time, until we run out of elements in one set or the other.

**Proposition 20.6.** *Let $(M, <)$ and $(N, <)$ be well-orderings. Exactly one of the following cases holds:*

1. *$M$ and $N$ are order isomorphic*
2. *$M$ is order isomorphic to an initial segment of $N$.*
3. *$N$ is order isomorphic to an initial segment of $M$.*

*Proof.* If cases (1) and (2) both hold, then $N$ would be order isomorphic to an initial segment of itself, contradicting Lemma 20.4. The other joint possibilities are similarly contradictory. This proves the uniqueness claim.

We now show at least one case holds. Let $f = \{(m, n) \in M \times N : I_{m,<} \cong I_{n,<}\}$. In other words, $f$ is the set of pairs $(m, n)$ where the initial segment determined by $m \in M$ is order isomorphic to the initial segment determined by $n \in N$.

We claim that $f$ is an injective relation. Suppose $(m_1, n), (m_2, n) \in f$ for some $m_1, m_2 \in M$ and some $n \in N$. We then have

$$I_{m_1,<} \cong I_{n,<} \cong I_{m_2,<}.$$

Without loss of generality, $m_1 \le m_2$. The initial segment $I_{m_2,<}$ is well-ordered, and so cannot be isomorphic to an initial segment of itself by Lemma 20.4. Thus $m_1 = m_2$.

By a symmetric argument we see that $f^{-1}$ is injective. Equivalently, $f$ is a function.

Next, we claim that $f$ is a strictly increasing map. Let $(m_1, n_1), (m_2, n_2) \in f$ with $m_1 < m_2$. Fix order isomorphisms $f_i \colon I_{m_i,<} \to I_{n_i,<}$ for $i \in \{1, 2\}$. Also let $g \colon I_{m_1,<} \to I_{m_2,<}$ be the inclusion map. The map $f_2 \circ g \circ f_1^{-1}$ is an order isomorphism of $I_{n_1,<}$ to an initial segment of $I_{n_2,<}$. Applying Lemma 20.4 once more, we cannot have $n_1 \geq n_2$. Thus $n_1 < n_2$, as desired.

By a similar argument, $f^{-1}$ is strictly increasing. We have now shown that $f$ is an order isomorphism from its domain to its range.

Next, we claim that $\operatorname{dom}(f)$ is either all of $M$ or an initial segment. If $\operatorname{dom}(f)$ is not all of $M$, then $M - \operatorname{dom}(f)$ is a nonempty subset of $M$, and hence has a minimal element $x \in M - \operatorname{dom}(f)$. It suffices to show that $\operatorname{dom}(f) = I_{x,<}$.

First, if $y \in I_{x,<}$ then $y \in \operatorname{dom}(f)$ by minimality of $x$, and so $I_{x,<} \subseteq \operatorname{dom}(f)$. For the reverse containment, supposing $y \in \operatorname{dom}(f)$ then there is an order isomorphism $g \colon I_{y,<} \to I_{f(y),<}$. If $x \leq y$ then $g$ restricts to an order isomorphism $I_{x,<} \to I_{g(x),<}$, contradicting the fact that $x \notin \operatorname{dom}(f)$. Thus $y < x$, and so $\operatorname{dom}(f) \subseteq I_{x,<}$.

By a similar argument, we see that $\operatorname{ran}(f)$ is either all of $N$ or an initial segment.

We are now done, except when $\operatorname{dom}(f) = I_{m,<}$ and $\operatorname{ran}(f) = I_{n,<}$, for some $m \in M$ and $n \in N$. But then $f$ itself is an order isomorphism $I_{m,<} \to I_{n,<}$ by what we just showed, and so $(m, n) \in f$, contradicting the fact that $\operatorname{dom}(f) = I_{m,<}$. □

The proof that every well-ordered set is order isomorphic to an ordinal is quite similar, as follows.

**Theorem 20.7.** *If $(M, <)$ is a well-ordering, then there exists a unique ordinal $\alpha$ that is order isomorphic to $M$. The order isomorphism is also unique.*

*Proof.* Uniqueness of the order isomorphism will be a consequence of Corollary 20.5. Uniqueness of the ordinal $\alpha$ follows from Lemma 20.4, for if $M \cong \alpha$ and $M \cong \beta$, for some ordinals $\alpha < \beta$, then this would imply that $\beta$ is order isomorphic to an initial segment of itself.

All that remains is existence. Consider the collection

$$f = \{(m, \alpha) \, : \, I_{m,<} \cong \alpha \text{ and } \alpha \text{ is an ordinal}\}.$$

It is a set by an axiom of replacement. The fact that $f$ is a function follows from the uniqueness claims above. It is injective by the same argument as in the proof of Proposition 20.6.

Let $\beta$ be the least ordinal such that $\beta \notin \operatorname{ran}(f)$. By the same argument as in the proof of Proposition 20.6, $\operatorname{ran}(f) = \beta$ and $\operatorname{dom}(f)$ is either all of $M$ or an initial segment $I_{m,<}$ for some $m \in M$. If $\operatorname{dom}(f) = M$ we are done. In the other case, $f \colon I_{m,<} \to \beta$ is an order isomorphism. But then $(m, \beta) \in f$, contradicting the fact that $\beta \notin \operatorname{ran}(f)$. □

## 20.C Philosophy of sets

We began this book with the intent to formally define sets. Now that we have described the standard formalization, we should take a moment to reflect on how closely this formal notion matches our informal idea of "collection".

One thing is clear; we are no longer studying arbitrary collections. First of all, by the axiom of regularity we have explicitly rejected some collections that, at least *a priori*, are perfectly logical objects of study. Second, the very idea of "classes" suggests that sets are only the very small collections, which might in turn be collected into even larger collections. Those larger collections cannot be sets, on pain of Russell's paradox, but they still seem to act like sets.

In fact, our use of first order logic to formalize our informal notion of collection seems to implicitly disregard Russell's paradox. How can we on one hand assert that the totality of all collections cannot itself be collected, yet on the other hand require that semantics are provided by a total collection (i.e., a model)? Does it even make sense to talk about the totality of all collections, for doesn't that very notion collect them? This calls into question the idea of universal quantification itself. But as these are philosophical questions, not yet satisfactorily addressed by mathematicians, we leave them unanswered.

## 20.D Exercises

**Exercise 20.1.** Turn the shorthand version of the axiom of regularity, given in (20.1), into an actual formula in the signature $(\in)$.

**Exercise 20.2.** Show that ordinal arithmetic is not distributive from the right, by finding ordinals $\alpha, \beta, \gamma$ such that $(\alpha + \beta) \cdot \gamma \neq (\alpha \cdot \gamma) + (\beta \cdot \gamma)$.

**Exercise 20.3.** Prove that ordinal addition is associative. (Hint: Prove that the ordinals $(\alpha+\beta)+\gamma$ and $\alpha+(\beta+\gamma)$ both have the same order type as the lexicographical order on $(\{0\} \times \alpha) \cup (\{1\} \times \beta) \cup (\{2\} \times \gamma)$.)

**Exercise 20.4.** Prove that if $\alpha, \beta \in \mathrm{Ord}$ and $\beta \neq 0$ is a limit ordinal, then $\alpha + \beta$ is a limit ordinal.

**Exercise 20.5.** A binary class relation $R$ on a class $M$ is *set-like* if for each set $y \in M$, the class $\{x : xRy\}$ is a set. (The class of $R$-smaller elements is a set.) Prove that $\in$ is a set-like binary relation on the class of all sets $V$.

**Exercise 20.6.** A binary class relation $R$ on a class $M$ is *extensional* if for each pair of sets $x, y \in M$, we have

$$\{z : zRx\} = \{z : zRy\} \to x = y.$$

Prove that $\in$ is an extensional relation on $V$.

The previous two exercises are important because of the following fact. Any well-founded, set-like, extensional relation $R$ on a class $M$ is "isomorphic" to the membership relation $\in$ restricted to a (unique!) transitive class $N$, called the *Mostowski collapse* of $M$. This generalizes the work in Subsection 20.B from well-orderings (on sets) to well-founded relations (on classes).

**Exercise 20.7.** Does the relation $\{(4,5),(4,6),(6,7)\}$ have a Mostowski collapse? If so, then find it. If not, explain why not.

**Exercise 20.8.** Does the relation $\{(0,1),(0,2),(1,2),(1,3),(3,4)\}$ have a Mostowski collapse? If so, then find it. If not, explain why not.

# Chapter IV

# Induction, Recursion, and Computing

*The road to hell is paved with intractable recursions, bad equilibria, and information cascades.* Brian Christian

In this chapter we generalize induction to the ordinals. This allows us to transfer induction arguments to arbitrary well-ordered sets.

This higher form of induction in turn provides for transfinite recursion—building objects using an infinitely long process. For instance, given a set $S$, then using just the axioms of union, pairing, and power set, we can construct the sets

$$\{S\},\ \{S, \mathscr{P}(S)\},\ \{S, \mathscr{P}(S), \mathscr{P}(\mathscr{P}(S))\}, \ldots,$$

but we can't seem to complete the infinite construction of putting all of these sets together to form

$$\{S, \mathscr{P}(S), \mathscr{P}(\mathscr{P}(S)), \ldots\}.$$

Transfinite recursion lets us construct sets like this, and more.

On the other hand, other forms of recursion occur in logic and computing. We give a brief introduction to those topics as well.

## 21 Transfinite induction

The principle of mathematical induction is a claim about the natural numbers. It asserts the following: If an open sentence $P(x)$ is true at 0, and when it is true at $n$ this implies it is also true at $n+1$, then $P(x)$ is true at all natural numbers. However, as most students recognize, the principle applies in other situations. For instance, we can apply induction to the set of positive integers; the only change is that the base case is now 1. There are even more exotic examples.

What really makes these examples work is the fact that the order relation on $\mathbb{N}$ is well-founded. This naturally leads us to the following fundamental theorem.

**Theorem Schema 21.1** (Transfinite induction). *Let $C$ be a subclass of* Ord. *Suppose that the following three conditions hold:*

(1) $0 \in C$,

(2) *if* $\alpha \in C$, *then* $\alpha + 1 \in C$, *and*

(3) *if* $\alpha \neq 0$ *is a limit ordinal and* $\beta \in C$ *for each* $\beta < \alpha$, *then* $\alpha \in C$.

*In this case,* $C = \mathrm{Ord}$.

*Proof.* If $C \neq \mathrm{Ord}$, let $\alpha$ be the smallest ordinal not in $C$ (which exists since Ord is well-founded). If $\alpha = 0$, then this contradicts (1). If $\alpha$ is a successor ordinal, this contradicts (2). If $\alpha$ is a nonzero limit ordinal, this contradicts (3). □

Transfinite induction is like standard induction, but there is an extra case to handle the limit ordinals. As a starting example, let us prove, by transfinite induction, that every ordinal is the sum of a limit ordinal and a finite ordinal, in a unique way.

**Proposition 21.2.** *If* $\alpha \in \mathrm{Ord}$, *then there exists a unique limit ordinal* $\beta$ *and a unique natural number* $n \in \omega$ *such that* $\alpha = \beta + n$.

*Proof.* **(Existence)**: We work by transfinite induction on $\alpha \in \mathrm{Ord}$. For the base case $\alpha = 0$, the decomposition $\alpha = 0 + 0$ works. We next do the inductive step at successor ordinals. Suppose that $\alpha = \beta + n$ is such a decomposition, and we will find a decomposition for $\alpha + 1$. We see that

$$\alpha + 1 = (\beta + n) + 1 = \beta + (n + 1).$$

Finally, if $\alpha$ is a nonzero limit ordinal, then $\alpha = \alpha + 0$ works.

**(Uniqueness)**: In this part of the proof we will use the fact that $\beta + n$ is a successor ordinal if $n \geq 1$. (It is the successor of $\beta + (n - 1)$.) Again work by transfinite induction. If $\alpha$ is a limit ordinal, then it is not a successor ordinal, by definition, and so the only possible decomposition is $\alpha = \alpha + 0$. (Note: This covers the case when $\alpha = 0$.) To check uniqueness of such decompositions for successor ordinals, suppose that $\alpha + 1 = \beta + n$ is some possible decomposition. Clearly $n > 0$, otherwise $\alpha + 1$ would be a limit ordinal. Now $\alpha = \beta + (n - 1)$, and our inductive hypothesis says that this decomposition is unique. So there was only one choice for $\beta$ and one choice for $n - 1$ (and hence for $n$). □

It is quite common to combine parts (1) and (3) of Theorem 21.1 into a single check over all limit ordinals (zero or not). When doing so, it can be helpful to the reader to explicitly point out that the zero ordinal is being included. We did just that in the second half of the proof of Proposition 21.2.

It is also possible to combine all three parts of Theorem 21.1 into a single check. You just have to check for every ordinal $\alpha$ (either a limit or a successor) that

$$\text{if } \beta \in C \text{ for each } \beta < \alpha, \text{ then } \alpha \in C.$$

In a sense, this is a form of "strong" transfinite induction. However, in actual mathematical practice, the proof of this statement when $\alpha$ is a successor ordinal is usually different than when $\alpha$ is a limit ordinal. Therefore, those two cases are often kept separate.

## 21.A Informal transfinite recursion

Recursion just means repeating a process or rule. For instance, the Fibonacci numbers $\{F_n : n \in \mathbb{N}\}$ are defined by a simple, repeated rule: Take $F_0 = 0$, $F_1 = 1$, and put $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$.

"Transfinite recursion" is a fancy name for infinitely repeating a process. To motivate the idea, we will prove by transfinite recursion that for every set $S$ we have $|S| < |\mathscr{P}(S)|$. This result was proved in "Transition" by very different methods, using the notion of "barber sets".

**Theorem 21.3.** *If $S$ is a set, then $|S| < |\mathscr{P}(S)|$.*

*Proof.* Let $f\colon \mathscr{P}(S) \to S$ be any function. We will show that $f$ is not injective, by finding two distinct subsets of $S$ that have the same image under $f$.

We construct a sequence of subsets of $S$ and another sequence of elements of $S$, indexed by the ordinals, as follows. First, we take $X_0 := \emptyset$ and $s_0 := f(\emptyset) = f(X_0)$. Next we take $X_1 := \{s_0\}$ and $s_1 := f(\{s_0\}) = f(X_1)$. After that we take $X_2 := \{s_0, s_1\}$ and $s_2 := f(\{s_0, s_1\}) = f(X_2)$. Similarly $X_3 := \{s_0, s_1, s_2\}$ and $s_3 := f(\{s_0, s_1, s_2\}) = f(X_3)$. Continuing in this way, it should be clear how to define $X_n$ and $s_n$ for each $n \in \omega$. What happens at $\omega$? We follow exactly the same idea, and take

$$X_\omega := \{s_n : n \in \omega\} \text{ and } s_\omega := f(X_\omega).$$

Then we take $X_{\omega+1} := \{s_0, s_1, s_2, \dots, s_\omega\}$ and $s_{\omega+1} := f(X_{\omega+1})$, and so forth.

We can write this formally for a general ordinal $\alpha$. If for each ordinal $\beta < \alpha$ the elements $s_\beta$ have been constructed, then we let

$$X_\alpha := \{s_\beta : \beta < \alpha\} \text{ and } s_\alpha := f(X_\alpha).$$

This recursively defines $X_\alpha$ and $s_\alpha$ with respect to the previously constructed terms in the sequence.

We are now ready to finish the proof. Since $S$ is a set, but Ord is a proper class, the list of elements we constructed must have repetitions. Let $\beta$ be the smallest ordinal such that $s_\beta = s_\alpha$ for some $\alpha > \beta$. From this minimality condition, we see that $s_\beta \notin X_\beta$. However, $s_\beta \in X_\alpha$. Thus $f$ maps the distinct sets $X_\beta \neq X_\alpha$ to the same element $s_\beta = s_\alpha$. □

(To aid in understanding the previous proof, you might attempt Exercise 21.5 now. You might also try the case when $S = \mathbb{N}$, and $f\colon \mathscr{P}(S) \to S$ is the function where $f(X) = |X|$ when $X$ is finite and $f(X) = 0$ otherwise.)

Transfinite recursion is helpful in constructing many different objects, functions, and rules. For example, we defined ordinal addition using Definition 13.1, but there is an alternative definition using transfinite recursion. Since we are defining the addition symbol for ordinals, we will (temporarily) use the notation $\alpha^* = \alpha \cup \{\alpha\}$ in place of the usual $\alpha + 1$.

$$\alpha + \beta = \begin{cases} \alpha & \text{if } \beta = 0 \text{ is the smallest ordinal,} \\ (\alpha + \gamma)^* & \text{if } \beta = \gamma^* \text{ is a successor ordinal,} \\ \mathrm{lub}_{\mathrm{Ord}}(\{\alpha + \gamma \,:\, \gamma < \beta\}) & \text{if } \beta \neq 0 \text{ is a limit ordinal.} \end{cases} \tag{21.4}$$

Taking $\beta = 1$ in this definition, we have $\alpha + 1 := (\alpha + 0)^* = \alpha^*$, so our notations agree.

We can similarly define ordinal multiplication, as follows.

$$\alpha \cdot \beta = \begin{cases} 0 & \text{if } \beta = 0 \text{ is the smallest ordinal,} \\ (\alpha \cdot \gamma) + \alpha & \text{if } \beta = \gamma + 1 \text{ is a successor ordinal,} \\ \mathrm{lub}_{\mathrm{Ord}}(\{\alpha \cdot \gamma \,:\, \gamma < \beta\}) & \text{if } \beta \neq 0 \text{ is a limit ordinal.} \end{cases} \tag{21.5}$$

One benefit of transfinite recursive definitions is that they often naturally lead to an intuitive understanding of a concept. For instance, it becomes very easy to define ordinal exponentiation.

$$\alpha^\beta = \begin{cases} 1 & \text{if } \beta = 0 \text{ is the smallest ordinal,} \\ (\alpha^\gamma) \cdot \alpha & \text{if } \beta = \gamma + 1 \text{ is a successor ordinal,} \\ \mathrm{lub}_{\mathrm{Ord}}(\{\alpha^\gamma \,:\, \gamma < \beta\}) & \text{if } \beta \neq 0 \text{ is a limit ordinal.} \end{cases} \tag{21.6}$$

There are many "down-to-earth" examples of transfinite recursion. For instance, we can use it to prove that every vector space has a basis. The idea is simple: Keep picking elements that are linearly independent of the previously chosen elements, until you run out of options. (If you do not know what linearly independent elements of a vector space are, read the paragraph after the proof below.) We make this formal as follows.

**Theorem 21.7.** *Every vector space has a linearly independent subset that is maximal under set inclusion.*

*Proof.* Let $V$ be a vector space, and use the axiom of choice to fix a well-ordering $<$ on $V$. For each $\alpha \in \mathrm{Ord}$ let $v_\alpha \in V$ be the $<$-smallest element of $V - \{v_\beta : \beta < \alpha\}$ such that the set $\{v_\beta : \beta \leq \alpha\}$ is linearly independent, if any such element exists, otherwise stop the recursion. This process does eventually stop since $V$ is a set but Ord is a proper class. Letting $\theta$ be the ordinal when this process stops, we see that $\{v_\beta : \beta < \theta\}$ is a maximal linearly independent subset of $V$. □

Analyzing this proof we realize that the same argument works in a much broader context. Pretend for a moment that you do not know what "vector space" and "linear independence" mean, except that the latter is a property of subsets of the former type of set. Indeed, before reading further, replace those two terms with "set" and "interesting subset" respectively, and reread the proof above.

There are only two things we need to know about interesting subsets for this modified theorem to continue to hold. First, for limit ordinals $\alpha$, if $\{v_\gamma : \gamma < \beta\}$ is an interesting subset for each $\beta < \alpha$, then $\{v_\gamma : \gamma < \alpha\}$ is also interesting (else the subset we constructed may not be interesting). Second, no uninteresting set should be contained in an interesting set (else the subset we constructed in the last sentence may not be maximal). Both of these conditions hold for linear independence, since a set is linearly independent if and only if every finite subset has that same property.

**Remark 21.8.** There is a result in algebra, called Zorn's lemma, that is a merger of transfinite recursion with the axiom of choice. It is often used, in place of direct transfinite recursion, to prove results like Theorem 21.7. The benefit of transfinite recursion over Zorn's lemma is that it provides the intuition behind how an object is constructed in stages (perhaps with choices involved at each step). Zorn's lemma is favored by some mathematicians, because one is not required to explicitly (only implicitly) posit a well-ordering on the set in question. ▲

## 21.B Formal transfinite recursion

In the previous subsection we described, informally, how to repeat a process indexed by the ordinals. What does this mean formally? By a process, what we really mean is a **class** function. But what, exactly, is a class function? Even more fundamentally, formally what is a class, exactly? We will answer these questions now.

The language of formal set theory using first order logic, in the signature $(\in)$, only has set variables. Technically speaking, there are no objects called classes; there are only sets. Thus, a class is not a formal object. We want to think about classes as collections $\{x : \varphi(x)\}$, where $\varphi(x)$ is a well-formed FOL formula. We can call $\varphi(x)$ the *underlying formula* defining the class. For instance, we think of the class of all sets, $V$, as the collection $\{x : x = x\}$, determined by the formula $x = x$.

One might rightly ask: If classes are not formal objects, how do we use them in formal proofs? The answer is that classes are a convenient shorthand. Whenever we talk informally about these collections, we could instead formally use the underlying formula $\varphi(x)$ determining the class.

**Warning 21.9.** In formal set theory, the only classes available are those determined by well-formed formulas. There are some metatheoretical collections that are not formal sets, nor are they classes of formal sets, because they cannot be described by first order formulas.

**Warning 21.10.** There are other versions of formal set theory that have classes as objects. We won't discuss those systems in this book, but they are not too different than the formal system we have described.

Now, a *class function* is just a class that behaves like a function. More precisely, it is determined by a FOL formula that acts like a function. Let $x$, $y$, and $p$ be variables. We will think of $x$ as ranging over the inputs of our class function. We will think of $y$ as ranging over the outputs of our function. The variable $p$ will act as a parameter—by plugging in different values for $p$ we will get different (but related) functions. Let $\varphi(x, y, p)$ be a well-formed FOL formula, where $x$, $y$, and $p$ are free. It is the underlying formula for a class function exactly when ZFC proves

$$(\varphi(x, y_1, p) \wedge \varphi(x, y_2, p)) \to (y_1 = y_2). \tag{21.11}$$

Informally, but very suggestively, we might write $F_p = \{(x, y) : \varphi(x, y, p)\}$ to denote our class function.

**Example 21.12.** Let $\varphi(x, y)$ be the formula $\forall z\,(z \in y \leftrightarrow z \subseteq x)$. The reader is entrusted with the task of converting this shorthand formula to a formula in the signature $(\in)$, if so desired. There is no parameter $p$ in this example. Now, the class function $F = \{(x, y) : \varphi(x, y)\}$ is simply the power set function. In other words, $F(x) = \mathscr{P}(x)$ for each set $x$. △

**Example 21.13.** Let $\varphi(x, y, p)$ be the formula that says

$$(p \text{ is a function}) \wedge ((x, y) \in p).$$

Let $F_p = \{(x, y) : \varphi(x, y, p)\}$. If $p$ is a function, then $F_p$ is just the function $p$ itself. On the other hand, if $p$ is not a function, then $F_p = \emptyset$. △

**Example 21.14.** Let $\varphi(x, y)$ be the formula that says

$$y \in \mathrm{Ord} \wedge ((\forall z \in x)(z \in \mathrm{Ord} \wedge z \leq y)) \wedge ((\forall w \in \mathrm{Ord})(((\forall z \in x)(z \leq w)) \to (y \leq w))).$$

Note that $z \leq y$ just means $(z \in y) \vee (z = y)$. Also, $(\forall S \in T)\,\psi$ is shorthand for $\forall S\,(S \in T \to \psi)$. By Exercise 21.11, this shorthand formula can be converted to a full formula. The class function defined by $\varphi(x, y)$ is the function $y = \mathrm{lub}_{\mathrm{Ord}}(x)$, defined only when $x$ is a set of ordinals. △

We are now ready to state the *replacement axioms*. Informally, they read:

If $F$ is a class function and $S$ is a set, then $\{F(s) : s \in S\}$ is a set.

Formally, for each formula $\varphi(x, y, p)$ we get a corresponding axiom of replacement

$$\begin{aligned}&\big(\forall x\,\forall y\,\forall z\,((\varphi(x,y,p) \wedge \varphi(x,z,p)) \to (y = z))\big)\\ &\to \big(\forall S\,\exists T\,\forall y\,(y \in T \leftrightarrow (\exists x \in S)\,\varphi(x,y,p)\big).\end{aligned} \tag{21.15}$$

We cannot combine these axioms into a single axiom because we have no way to quantify over formulas $\varphi$ in FOL. We only can quantify over the formal set variables. In ZFC, we accept these axioms as part of our formal set theory.

One may wonder why we only used a single parameter $p$, and not more. This is because we can encode more parameters by letting $p$ be an ordered pair, or triple, and so forth.

The proof of transfinite recursion fundamentally uses the replacement axioms (but not any choice axioms). In some sense the converse is also true; the replacement axioms are corollaries of (some forms of) transfinite recursion. We will not prove these claims, as the proofs are quite technical. However, it is important to mention that these results hint at the deep importance of transfinite recursion as a fundamental (even axiomatic) principle in mathematics.

**Advice 21.16.** Do not focus on the technical details of transfinite recursion. To do it, you just need to tell the reader how you construct the next term in your sequence. This usually consists of describing what to do at successor and limit ordinals.

There are two technical points to be aware of regarding transfinite recursion. First, just because a process works at the first infinite ordinal $\omega$, that does not immediately imply that it will work at every infinite ordinal. Many errors in transfinite recursion happen at the limit ordinals, because authors are sloppy in explaining what happens in the general case. Second, and more importantly:

**Warning 21.17.** Formal transfinite induction works for processes that can be defined by FOL formulas. There are some metaprocesses that cannot be so defined.

## 21.C Exercises

**Exercise 21.1.** What is the smallest ordinal that you do not have a notation for? (Give a description of it that your classmates can understand.) If you call that ordinal $\zeta$, what is the new smallest ordinal that you do not have a notation for?

In the proof of Proposition 21.2, there were some implicit claims that we took for granted. The next three exercises clarify these points.

**Exercise 21.2.** Suppose that $\beta = \alpha + 1$ is a successor ordinal. Prove that $\alpha$ is the maximal element of $\{\gamma \in \mathrm{Ord} : \gamma < \beta\}$. (Hint: You may find Exercise 19.2 useful.)

**Exercise 21.3.** For any two ordinal numbers $\alpha, \alpha' \in \mathrm{Ord}$, prove that if $\alpha+1 = \alpha'+1$ then $\alpha = \alpha'$.

**Exercise 21.4.** Explain how the previous two exercises justify the use of $n - 1$ in the proof of Proposition 21.2.

**Exercise 21.5.** Let $S = \mathbb{N}$, and define a function $f\colon \mathscr{P}(S) \to S$ by the rule

$$f(X) = \begin{cases} |X| + 1 & \text{if } |X| < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Describe the sequence of sets $X_0, X_1, X_2, \ldots$ and the sequence of elements $s_0, s_1, s_2, \ldots$ that arise in the proof of Theorem 21.3.

**Exercise 21.6.** Prove by transfinite induction that ordinal arithmetic is left distributive.

**Exercise 21.7.** Prove by transfinite induction that the two definitions we have given for ordinal addition agree. (Can you do the same for ordinal multiplication?)

**Exercise 21.8.** Let $\alpha < \beta$ be ordinals. Prove that there exists a unique $\gamma \in \mathrm{Ord}$ such that $\alpha + \gamma = \beta$.

**Exercise 21.9.** Prove the following "left division algorithm" for ordinals. Given $\alpha, \beta \in \mathrm{Ord}$ with $\beta \neq 0$, then there exist unique $\gamma, \delta \in \mathrm{Ord}$ with $\delta < \beta$ such that $\alpha = \beta \cdot \gamma + \delta$.

**Exercise 21.10.** Prove *Cantor's normal form theorem*: Every $\alpha \in \mathrm{Ord}$ has a unique representation as

$$\alpha = \sum_{i=1}^{m} \omega^{\beta_i} n_i = \omega^{\beta_1} n_1 + \omega^{\beta_2} n_2 + \cdots + \omega^{\beta_m} n_m$$

for some natural number $m$, some positive integers $n_1, \ldots, n_m$, and some ordinals $\beta_1 > \beta_2 > \cdots > \beta_m$. (In the base case, when $\alpha = 0$, then $m = 0$ and the sum is empty.)

**Exercise 21.11.** Find a well-formed FOL formula in the signature $(\in)$, with one free variable $x$, that says "$x$ is an ordinal".

**Exercise 21.12.** Prove that the separation axioms are a consequence of the replacement axioms. (Hint: Let $\varphi(x)$ be a formula. For each set $S$ we want to prove that $\{x \in S : \varphi(x)\}$ is a set. Consider the class function $F = \{(x, x) : \varphi(x)\}$.)

# 22 Formalities and recursion

In this section we give some further examples of recursion. We provide examples of what would be expected when constructing objects using recursion, and also explain what parts of recursion can be "swept under the rug". We also finally give the formal definition of cardinality.

## 22.A Transitive closure

Let $R$ be a binary set relation. To define the transitive closure of $R$, as we did previously, an author would be expected to write only the following: Take $S_0 := R$, and for natural numbers $n > 0$ we recursively take $S_{n+1} := R \circ S_n$. Finally, we define $R^{\text{tran}} := \bigcup_{n \in \mathbb{N}} S_n$.

**Question**: Is $n$ being used as a meta natural number, or as a formal natural number?

**Answer**: It depends on the setting.

**Question**: Are there two versions of the transitive closure, one formal and one meta?

**Answer**: Yes. How very astute of you to see that!

**Question**: How can we tell the difference between them?

**Answer**: Authors usually rely on context. If $R$ belongs to a formal system, then the construction should be the formal one.

**Question**: When doing formal recursion, don't we need to describe the meta process using a FOL formula to avoid running afoul of Warning 21.17?

**Answer**: Not usually, because most mathematicians automatically recognize when a process can be so formalized. But you have hit upon the main issue that leads to mistakes when doing recursion. Sometimes the construction is not actually defined. Perhaps a simple example will elucidate this point.

Suppose that a parent of three children says "The youngest needs to clean their room, and the next youngest needs to clean their room and wash their own dishes, and so forth". The oldest child is justified in not knowing exactly what is expected of them (without further clarification from the parent). There is no formula to describe what the parents have said.

The recursive definition of the transitive closure does not have this issue. Each $R^n$ is clearly defined, using the previous terms in the sequence and other standard constructions that we have already defined.

**Question**: If I was interested in formally describing the transitive closure how would I go about giving the corresponding FOL formula?

**Answer**: You would need to describe a class function $\varphi(x, y, R)$, where if you input the set $x = \{(0, R^1), (1, R^2), \ldots, (n-1, R^n)\}$ that encodes the previous steps of the recursion, then the formula should force an output of $y = R^{n+1}$.

## 22.B Sets of power sets

Let us show that a formal set $\{S, \mathscr{P}(S), \mathscr{P}(\mathscr{P}(S)), \ldots\}$ can be defined. Put $S_0 := S$, and for any integer $n \geq 0$ put $S_{n+1} := \mathscr{P}(S_n)$. The set we desire is just $\{S_n : n \in \omega\}$ (which exists by an axiom of replacement). That's all that is needed.

The process of taking power sets stops at $\omega$. Indeed, what would an infinite iteration of the powerset operation

$$\cdots(\mathscr{P}(\mathscr{P}(\mathscr{P}(S))))\cdots$$

even mean? But there is a closely related process that we can continue indefinitely. For the base case, take $T_0 := S$. For successor ordinals, if we have constructed $T_\alpha$ then take $T_{\alpha+1} := T_\alpha \cup \mathscr{P}(T_\alpha)$. Finally, for any nonzero limit ordinal $\alpha$, if we have constructed $T_\beta$ for each $\beta < \alpha$, take $T_\alpha := \bigcup_{\beta<\alpha} T_\beta$.

**Proposition 22.1.** *If $\beta < \alpha$ are ordinals, then $T_\beta \subseteq T_\alpha$ and $|T_\beta| < |T_\alpha|$.*

*Proof idea.* This is left as an exercise. Induct on $\alpha$. □

An immediate consequence of this proposition is that cardinalities never end. There are at least as many cardinalities as ordinals, so they form a proper class. Technically we have not formally defined cardinals, so we will do that now.

## 22.C Cardinal numbers

**Definition 22.2.** A set $\kappa$ is a *cardinal* if it is an ordinal for which no smaller ordinal is bijective with it.

While we use the first few lowercase Greek letters $\alpha, \beta, \gamma, \ldots$ to denote ordinals, we will use the later letters $\kappa, \lambda, \ldots$ when emphasizing that an ordinal is also a cardinal.

Each natural number is a cardinal, and $\omega$ is a cardinal too. When emphasizing that $\omega$ is a cardinal, we write $\aleph_0$, and think of it as the first infinite cardinal. The next infinite cardinal is $\aleph_1$ and its underlying ordinal is written as $\omega_1$. In other words, $\omega_1$ is the first ordinal that follows all of the (uncountably many) countable ordinals.

We should show that every set is in bijection with a unique cardinal. Let $S$ be a set and let $C$ be the class of ordinals that are in bijection with $S$. As $S$ can be well-ordered by the axiom of choice, then by Theorem 20.7 there is at least one ordinal in bijection with $S$. So $C \neq \emptyset$. By Theorem Schema 19.8, $C$ has a unique minimal element. That unique minimal element is what we call the *cardinality* of $S$. (Since $C$ is a set, rather than a proper class, cardinality is defined without using the "schema" aspect of Theorem Schema 19.8.)

The collection of all cardinals is a proper class. There is no largest cardinal, by Theorem 21.3, and also by Proposition 22.1. The infinite cardinals are well-ordered, so we write $\aleph_\alpha$ for the $\alpha$th infinite cardinal, where $\alpha \in \mathrm{Ord}$.

Another reason for using different symbols for cardinals and ordinals is because we can define addition and multiplication of cardinals, and these are different functions than ordinal addition and multiplication.

**Definition 22.3.** Let $\kappa$ and $\lambda$ be cardinals. We define $\kappa+\lambda$ to be the cardinality of the disjoint union of $\kappa$ and $\lambda$. Similarly, we define $\kappa \cdot \lambda$ to be the cardinality of the direct product $\kappa \times \lambda$.

One doesn't often talk about the sum or product of just two infinite cardinals, because these functions are somewhat trivial.

**Theorem 22.4.** *If* $\alpha, \beta \in \mathrm{Ord}$*, then* $\aleph_\alpha + \aleph_\beta = \aleph_\alpha \cdot \aleph_\beta = \aleph_{\max\{\alpha,\beta\}}$.

*Proof.* Without loss of generality we may assume $\alpha \leq \beta$. It is clear that

$$\aleph_\beta \leq \aleph_\alpha + \aleph_\beta \leq \aleph_\beta + \aleph_\beta = 2 \cdot \aleph_\beta \leq \aleph_\alpha \cdot \aleph_\beta \leq \aleph_\beta \cdot \aleph_\beta,$$

so by the Schröder-Bernstein Theorem it suffices to find an injection $\aleph_\beta^2 \to \aleph_\beta$.

Consider the canonical order on $\omega_\beta^2$. We claim that under this ordering, $\omega_\beta^2$ is order isomorphic to $\omega_\beta$. By way of contradiction suppose not, and let $\beta$ be the least such ordinal. (It cannot be the case that $\beta = 0$ by Exercise 13.10. The argument below can be viewed as an extension of that exercise.)

First, notice that $\omega_\beta^2$ cannot be order isomorphic to an initial segment of $\omega_\beta$, for then $\omega_\beta$ would inject into an initial segment of itself, which is impossible (as it is a cardinal). So there exists some $(\gamma, \delta) \in \omega_\beta^2$ such that the initial segment $S$ generated by $(\gamma, \delta)$ is order isomorphic to $\omega_\beta$. In particular, $\aleph_\beta = |S| \leq \max\{|\gamma^2|, |\delta^2|\}$.

We will consider the case when $\gamma$ and $\delta$ are not finite, leaving the other possibility to the reader. As $\gamma < \omega_\beta$, the inductive hypothesis implies that $|\gamma^2| = |\gamma| \cdot |\gamma| = |\gamma|$, and similarly for $\delta$. Thus

$$\aleph_\beta \leq \max\{|\gamma|, |\delta|\} < \aleph_\beta,$$

giving the needed contradiction. □

**Corollary 22.5.** *Let $S$ be any infinite set. If $S^{<\infty}$ is the collection of all finite sequences of elements from $S$, then $|S| = |S^{<\infty}|$.*

*Proof.* Let $S^n$ be the set of all sequences of length $n$ from the elements in $S$. We see that $S^{<\infty} = \bigcup_{n\in\omega} S^n$, and that this union is disjoint. By Theorem 22.4 and induction on $n$, we have $|S^n| = |S|$ for each $n \geq 1$. (Trivially, $|S^0| = 1$, since there is only one empty sequence.) Thus

$$|S| \leq |S^{<\infty}| = \left|\bigcup_{n\in\omega} S^n\right| \leq \aleph_0 \cdot |S| = |S|.$$

The result now follows from the Schröder-Bernstein theorem. □

## 22.D Meta counting can be recursive

Some recursive constructions are in the metatheory. For the remainder of this section we will let $\mathbb{N}$ denote the metamathematical natural numbers, and we will let $\omega$ be the corresponding formal set. We have an injective map $\mathbb{N} \to \omega$ given by the rules

$$0 \mapsto \emptyset,\ 1 \mapsto \{\emptyset\},\, 2 \mapsto \{\emptyset, \{\emptyset\}\}, \ldots,$$

but to simplify notation we will instead write $n \mapsto \underline{n}$. Thus, if $n$ is a meta number, then $\underline{n}$ is the formal set that naturally encodes that number. This map is defined recursively (in the metatheory) by the rule that $0 \mapsto \emptyset$ and for any $n \geq 0$ if $n \mapsto \underline{n}$ then $n+1 \mapsto \underline{n} \cup \{\underline{n}\}$.

This "underline" map is just one of the many meta maps that can be constructed recursively. We will describe one more important example.

Let $\mathscr{L}$ be a FOL language with signature $(\in)$. We will assume that the variables are $x_0, x_1, \ldots$, for simplicity. Let $\mathrm{String}_{\mathscr{L}}$ be the (metamathematical) collection of all strings of symbols in the language $\mathscr{L}$. We know that $\mathrm{String}_{\mathscr{L}}$ is countable, so there exists some bijection $\mathrm{String}_{\mathscr{L}} \to \mathbb{N}$. We will construct such a bijection, recursively.

The symbols in the language $\mathscr{L}$ are parentheses, the universal quantifier, negation and disjunction symbols, the equality and $\in$ symbols, and variables. Let $\mathrm{Symb}_{\mathscr{L}}$ be the collection of these symbols. Consider the map $f\colon \mathrm{Symb}_{\mathscr{L}} \to \mathbb{N}$ given by the rules:

$$(\mapsto 0,\ )\mapsto 1,\ \forall \mapsto 2,\ \neg \mapsto 3,\ \vee \mapsto 4,\ = \mapsto 5,\ \in \mapsto 6,\ x_n \mapsto 7+n. \tag{22.6}$$

Let $\mathbb{N}^{<\infty}$ be the set of finite sequences of natural numbers. The map $f$ induces a new map $f^{<\infty}\colon \mathrm{String}_{\mathscr{L}} \to \mathbb{N}^{<\infty}$. For instance, we would have

$$)x_2\forall = \mapsto (1, 9, 2, 5).$$

This new, induced map $f^{<\infty}$ is defined by recursion over the length of strings. First, the empty string of symbols maps to the empty sequence of numbers. Next, we recursively define it on strings of length $n+1$ by the rule

$$f^{<\infty}(s_0 s_1 \cdots s_{n-1} s_n) = (f^{<\infty}(s_0 s_1 \cdots s_{n-1}), f(s_n)).$$

To test your understanding of this map, you might attempt Exercise 22.7 now.

Next, one can recursively define a bijection $g\colon \mathbb{N}^{<\infty} \to \mathbb{N}$; see Exercise 22.8. So the map $\# = g \circ f^{<\infty}$ defines a bijection $\#\colon \mathrm{String}_{\mathscr{L}} \to \mathbb{N}$.

## 22.E Subtleties and the use-mention distinction

Consider the following two sentences:

The cow mooed softly.
"Cow" has three letters.

In the first sentence we use the word "cow" to refer to the bovine animal. In the second sentence we mention the word "cow" without referring to the animal. This illustrates what is known as the "use-mention distinction"; the first sentence uses the word, whereas the second sentence only mentions it.

Sometimes we can tell whether a word is being used or mentioned from context. Other times quotation marks help. Some authors use single quote marks, others use double quote marks (as we have done above), and others invent new symbols such as the square quote marks $\ulcorner$ and $\urcorner$.

Notice that in the previous sentence we did not *use* the square quote marks, we only *mentioned* them. So, should we have technically written: "$\ulcorner\ulcorner\urcorner$ and $\ulcorner\urcorner\urcorner$"? Most languages are just not well-equipped to handle these issues. Moreover, humans are rarely perfect at making the use-mention distinction explicit—such a person could rightly be accused of pedantry.

Let's consider a more mathematical example. We have viewed $\omega$ in two different ways: firstly, as a set in our formal set theory, and secondly, as a (derived) constant symbol of our formal set theory. Technically speaking, the *language* of formal set theory is a bunch of symbols, and involves no sets. We only evaluate the variables and constant symbols as sets inside a model.

Thus, when we say that there is a map $\mathbb{N} \to \omega$, it might be more appropriate to say that for any model $\mathcal{M}$ of set theory there is a map $\mathbb{N} \to \omega_{\mathcal{M}}$. As the model varies, so too does the map; we thus get an entire spectrum of maps. However, the map varies in a natural way, because $0 \mapsto \emptyset_{\mathcal{M}}$, $1 \mapsto \{\emptyset\}_{\mathcal{M}}$, and so forth. So, it might be even more appropriate to say that there exists a map from $\mathbb{N}$ to the (meta) collection of (possibly derived) constant symbols $c$ such that $\models_{\mathrm{ZFC}} c \in \omega$.

Even if ZFC is *consistent*, meaning that we cannot derive a contradiction, the underline map need not be surjective. Indeed, we could add a new constant symbol $c$ to the language, and adjoin to ZFC the new axioms $c \in \omega$ and $c \neq \underline{n}$ (for each $n \in \mathbb{N}$). This new theory is consistent by the compactness theorem. Any model $\mathcal{M}$ of this new theory will also be a model of ZFC, and $\mathbb{N} \to \omega_{\mathcal{M}}$ is not surjective.

## 22.F Exercises

The first two exercises are philosophical problems, and may have multiple answers.

**Exercise 22.1.** Give an explanation for why it does not contradict Exercise 11.16 that the transitive closure of a binary relation $R$ can be defined in formal set theory.

**Exercise 22.2.** Suppose a budding mathematician writes: Take $x_0 := 0$ and recursively define $x_{n+1}$ to be the smallest interesting number greater than $x_n$.

Is this a valid form of recursion? Justify your answer. If not, what is needed to fix this construction?

**Exercise 22.3.** Let $S = \emptyset$ and define the sets $T_\alpha$, for $\alpha \in \mathrm{Ord}$, as in Subsection 22.B. Describe $T_0$, $T_1$, $T_2$, $T_3$, $T_\omega$, and $T_{\omega+1}$.

**Exercise 22.4.** Prove Proposition 22.1.

**Exercise 22.5.** Handle the cases when $\gamma$ or $\delta$ is finite in the proof of Theorem 22.4.

**Exercise 22.6.** Let $S$ be any infinite set. Prove that the set of all finite subsets of $S$ has the same cardinality as $S$.

**Exercise 22.7.** What is the value, under the map $f^{<\infty}$, for the string of symbols "$x_7(\forall\vee\vee x_8$"? What string of symbols maps to $(5, 4, 3, 2, 1)$?

**Exercise 22.8.** Define an explicit bijection $g\colon \mathbb{N}^{<\infty} \to \mathbb{N}$ that can be evaluated by a computer. (Hint: Consider $(n_0, n_1, \dots, n_k) \mapsto 2^{n_0}3^{n_1} \cdots p_k^{n_{k-1}} p_{k+1}^{n_k+1}$.)

**Exercise 22.9.** Using the bijection you defined in the previous exercise, find $\#^{-1}(0)$ and $\#^{-1}(1)$. (As a bonus problem, find the smallest natural number $n$ such that $\#^{-1}(n) \in \mathrm{Form}_{\mathscr{L}}$.)

**Exercise 22.10.** (This is a tricky exercise if you don't use the axiom of choice.) Write $2^\kappa$ for the cardinality of $\mathscr{P}(\kappa)$, for any cardinal $\kappa$. Show that $\aleph_{\alpha+1} < 2^{2^{\aleph_\alpha}}$ for each $\alpha \in \mathrm{Ord}$. (Hint: Every well-ordering of $\aleph_\alpha$ is a subset of $\mathscr{P}(\aleph_\alpha^2)$.)

# 23 Computing functions

In this section we will define two classes of functions that have become important in mathematics, logic, computer science, and philosophy. They are called the *primitive recursive* and *computable* functions.

A word needs to be said about our presentation. Recall that when giving an axiomatization for first order logic, a choice is made for the set of logical connectives, the axioms, and the rules of inference. One of the most parsimonious presentations uses only one logical connective (the negation of disjunction), one axiom, and one rule of inference. That presentation is difficult to understand, let alone use, until a large number of techniques are developed to make it easier. On the other hand, a beginning student is usually taught logic using an extravagant presentation that uses a large number of connectives, axioms, and rules of inference. However, that approach makes it difficult to prove meta results because of the large number of cases to check, at least until one shows that the presentation could have been simplified.

In this book we took a middle road, by giving a small number of connectives, axioms, and rules of inference, which were still understandable. We left the task of *proving* that our axiomatization satisfies all of the usual logical rules to a future course in pure logic, taking it for granted that such a mundane task could be performed.

Likewise, we will make claims in this section that are not fully proven, and perhaps even impossible to state precisely without significant work. Our presentation is very compact and simple. We follow the work of Daniel E. Severin that was published in 2008.

## 23.A Primitive recursive functions

The motivating question for this section is: What types of functions $\mathbb{N} \to \mathbb{N}$ can computers evaluate? Some obvious examples come to mind.

The *zero function*, given by the rule $Z(x) = 0$, is perhaps the simplest. More generally, for any fixed $n \in \mathbb{N}$ the constant function $C_n(x) = n$ is computable. The *identity function*, given by the rule $I(x) = x$, may also compete for the role of the simplest example. A centrally important example, which has appeared previously in this text, is the *successor function*, given by the rule $S(x) = x + 1$.

More complicated examples abound, such as:

$$\begin{array}{rcll} D(x) & = & 2x & \text{(the doubling function),} \\ Sq(x) & = & x^2 & \text{(the squaring function), and} \\ Pow(x) & = & 2^x & \text{(the power of 2 function).} \end{array}$$

Rather than list more and more examples of such functions, it is helpful to describe operations on functions that preserve computability. Using these operations, we may build up from a few initial functions other computable functions. Consider the following three basic operations:

(1) *Arithmetic subtraction*: Given $f, g$, construct the new function $f \dot{-} g$ given by the rule

$$x \mapsto \begin{cases} f(x) - g(x) & \text{if } f(x) \geq g(x), \\ 0 & \text{otherwise.} \end{cases}$$

(2) *Composition*: Given $f, g$, construct the new function $fg$ (also written $f \circ g$).

(3) *Iteration*: Given $f$, construct the new function $f^{\square}$ given by the rule

$$x \mapsto \underbrace{f \circ f \circ \cdots \circ f}_{x \text{ times}}(0).$$

Equivalently, take $f^{\square}(0) := 0$ and for each $x \in \mathbb{N}$ take $f^{\square}(x+1) := f(f^{\square}(x))$.

**Example 23.1.** Start with the successor function $S$. Using arithmetic subtraction we can create the zero function $S \dot{-} S = Z$. Since $S \dot{-} Z = S$ and $Z \dot{-} S = Z \dot{-} Z = Z$, arithmetic subtraction alone will produce no additional functions.

Next consider composition. Of course $ZZ = ZS = Z$. However, $SZ = C_1$ is a new function, and so is $SS$ (which sends $x \mapsto x + 2$). We can continue composing functions to create new functions. Additionally, with access to these new functions we can again use arithmetic subtraction to create new functions like $S \dot{-} C_1 = I$, and then $I \dot{-} C_1$ (a sort of *predecessor function*, which we will denote as $P$).

Now throw iteration into the mix. We find that $(SS)^{\square} = D$, since

$$(SS)^{\square}(x) = S^{2x}(0) = \underbrace{((\cdots((0+1)+1)\cdots+1)+1)}_{2x \text{ times}} = 2x.$$

Similarly, $S(SD)^{\square} = \mathit{Pow}$. $\triangle$

The types of functions described in the previous example have a special name.

**Definition 23.2.** A function built from $S$ using finitely many applications of arithmetic subtraction, composition, and iteration is called *primitive recursive*.

These functions may seem somewhat special, but in fact this class of functions is quite extensive. Indeed, we make the following:

**Claim 23.3.** *A function $f\colon \mathbb{N} \to \mathbb{N}$ is primitive recursive if and only if there exists an algorithm that determines the function without using unbounded searches.*

In common jargon, "while loops" are a form of unbounded search, but "for loops" are given explicit bounds. The forward direction of Claim 23.3 seems easy enough. The function $S$ is given by the pseudo-code

```
x:=UserInput,
Increment[x],
Print[x]
```

Similarly, given programs for computing $f$ and $g$, we can create new programs for computing $f \dot{-} g$, $fg$, and $f^{\square}$, using only "for loops", "if-then-else commands", as well as "increment" and "decrement" routines and other standard functionality (but no "while loops"). However, making this claim concrete requires one to define algorithms in a precise, mathematical manner, which we will not attempt here.

It also seems intuitively obvious that any primitive recursive function can be evaluated on a computer. However, that is technically false. Most (if not all) computers you encounter in everyday life have finite memory. It is thus impossible to store sufficiently large numbers, let alone evaluate functions on them. Those who study computation theory deal with this issue by positing an *idealized* computer that has unbounded memory. The exact form that this idealized computer takes may vary slightly, but ultimately it can be described mathematically.

Programs evaluating the primitive recursive functions can be implemented on such an idealized machine. As we will see shortly, there are functions that are not primitive recursive, but they can still be implemented on an idealized machine; they just require unbounded searches in their definitions.

**Remark 23.4.** Julia Robinson has shown that one may dispense with the arithmetic subtraction operation at the cost of replacing $S$ with two complicated functions. ▲

## 23.B Computable functions

Consider the situation where we have a surjective function $f\colon \mathbb{N} \to \mathbb{N}$. For each $n \in \mathbb{N}$ there are one or more inputs $m \in \mathbb{N}$ with $f(m) = n$. Thus, there is a *least* such input. Assuming that the values of $f$ can be found, then an easy algorithm exists to find that smallest $m$.

```
m:=0,
While[f(m)!=n,
   Increment[m]
],
Print[m]
```

This algorithm is unbounded; we cannot give, *a priori*, an upper bound on how far we need to search.

Given $f$ as above, we will write $f^{[-1]}\colon \mathbb{N} \to \mathbb{N}$ for the function given by the rule that $f^{[-1]}(x)$ is the smallest $y \in \mathbb{N}$ with $f(y) = x$. We call the operation $f \mapsto f^{[-1]}$ the *first-inversion* operation.

**Example 23.5.** Consider the integer square-root function $f(x) = \lfloor \sqrt{x} \rfloor$, which is surjective. The first few outputs, as ordered by the inputs, are

$$0, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, \ldots$$

This function increases by 1 at each square, and remains unchanged otherwise.

The first-inversion of $f$ is the usual squaring function. △

Both the integer square-root function and the squaring function are primitive recursive. However, since first-inversion involves unbounded searches, we suspect that applying first-inversion to a primitive recursive function may not necessarily keep it primitive recursive. Thus, we make the following definition (to be justified shortly).

**Definition 23.6.** A function built from $S$ using finitely many applications of arithmetic subtraction, composition, iteration, and first-inversion (when defined) is called *computable*.

**Warning 23.7.** Computable functions are called *recursive* in many places. We will avoid that terminology.

**Remark 23.8.** Julia Robinson has shown that both arithmetic subtraction and iteration can be dispensed with in Definition 23.6, if $S$ is replaced by two complicated functions. In another direction, she has shown that every computable function is of the form $fg^{[-1]}$, for some primitive recursive functions $f$ and $g$, with $g$ surjective. ▲

Just as we did with primitive recursive functions, we make the following:

**Claim 23.9.** *A function $f\colon \mathbb{N} \to \mathbb{N}$ is computable if and only if there exists an algorithm that determines the function.*

It may seem strange, at first, that *every* algorithmic function arises from the successor function $S$, using only four simple operations. If you are familiar with coding, the following analogous situation may make the claim easier to believe: Computer programming languages often are based on only a few simple routines.

Claim 23.9 was first proved by Church and Turing, independently, using different (but, ultimately, equivalent) definitions of computable function than the one given here. Also, their idealized machines differed from one another; but their proofs of Claim 23.9 show that the machines are computationally equivalent. The *Church-Turing hypothesis* is the philosophical assertion that all deterministic (idealized) machines that can perform any unambiguous computation are computationally equivalent. Mathematicians have shown that many different idealized machines satisfy this hypothesis, and it is widely believed to hold true.

Our definition of first-inversion prevents it from being applied to all functions. For instance, for the successor function $S$ there is no value $y$ such that $S(y) = 0$. Thus $S^{[-1]}(0)$ is not defined. However, for any function $f\colon \mathbb{N} \to \mathbb{N}$, we will find it useful to think of $f^{[-1]}$ as a *partial* function that is well-defined on the image of $f$. One can thus define *computable partial functions* (in the obvious way) as those partial functions given by algorithms that may fail to print an output.

**Example 23.10.** The partial function $S^{[-1]}\colon \mathbb{N} \to \mathbb{N}$ is given by the rule

$$S^{[-1]}(x) = \begin{cases} \text{undefined} & \text{if } x = 0, \\ x - 1 & \text{if } x > 0. \end{cases}$$

This function differs from the predecessor function $P$ only at 0. △

## 23.C More general domains and codomains

You probably have firsthand experience with a computer doing computations on structures other than $\mathbb{N}$. For instance, your word processor may have given recommendations for changing grammar and spelling. Rather than developing a whole new definition of computability in terms of exotic structures like "the English language" we can incorporate them into our previous framework.

Let $S$ and $T$ be countable sets, and let $\varphi\colon \mathbb{N} \to S$ and $\psi\colon T \to \mathbb{N}$ be bijections. Given any function $f\colon S \to T$, we can ask if the function $f' = (\psi \circ f \circ \varphi)\colon \mathbb{N} \to \mathbb{N}$ is computable. If so, we can say that $f$ is computable, relative to $(\varphi, \psi)$.

The map $\varphi$ renames, or encodes, the elements of $S$ as natural numbers (and similarly $\psi$ renames $T$). If $S$ is countable, then there are uncountably many different choices for $\varphi$. We want the encodings to be simple and straightforward; so if $\varphi$ and $\psi$ can be defined in a recursive manner, then that often makes it easier to check if $f'$ is computable.

**Example 23.11.** When $S = \mathbb{N}^2$, there is an especially simple (and recursively defined) bijection $J\colon \mathbb{N}^2 \to \mathbb{N}$ given by the algebraic rule

$$J(x, y) = \frac{(x + y)^2 + 3x + y}{2}.$$

Computability of a function $f\colon \mathbb{N}^2 \to \mathbb{N}$ is *always* to be understood relative to $(J^{-1}, I)$. One can show that the computable functions $\mathbb{N}^2 \to \mathbb{N}$ are exactly the functions determined by algorithms where two inputs are allowed. △

We will give one more example of computability, in terms of set theory.

**Example 23.12.** Let $\mathscr{L}$ be the language of formal set theory, and let $\mathrm{String}_{\mathscr{L}}$ be the set of strings of symbols in this language. Recall that we recursively defined a bijection $\#\colon \mathrm{String}_{\mathscr{L}} \to \mathbb{N}$.

Consider the negation operation $N\colon \mathrm{String}_{\mathscr{L}} \to \mathrm{String}_{\mathscr{L}}$ that places a negation symbol at the front of any string. For instance $\neg(x_6 = x_7) \mapsto \neg\neg(x_6 = x_7)$.

Is $(\# \circ N \circ \#^{-1})$ a computable function? Yes; it is primitive recursive. As intuitive evidence towards this fact, the reader is invited to (1) find the value of this function at 5 (or any other random number), and (2) write a computer algorithm (without unbounded searches, but using standard functions) that encodes this function. (Showing that this function satisfies Definition 23.2 takes more work.) △

## 23.D Ackermann function

Define the *Ackermann function* $A\colon \mathbb{N}^2 \to \mathbb{N}$ by the three rules

$$A(0,y) = y+1, \qquad A(x+1,0) = A(x,1),$$

and

$$A(x+1,y+1) = A(x, A(x+1,y)).$$

This function is defined everywhere, since its values only depend on previously defined values (where "previous" means "smaller in the lexicographical order on $\mathbb{N}^2$"). For instance, we find

$$A(1,1) = A(0,A(1,0)) = A(0,A(0,1)) = A(0,2) = 3.$$

You might compute $A(2,2)$ to get a feeling for this function.

It becomes apparent that this function grows very fast. Indeed, it grows so quickly that we begin to believe it cannot be primitive recursive. That is indeed the case.

**Theorem 23.13.** *The Ackermann function is computable but not primitive recursive.*

The proof of this theorem is outlined in the exercises.

## 23.E Exercises

**Exercise 23.1.** Prove, for each $n \in \mathbb{N}$, that $C_n$ is primitive recursive.

**Exercise 23.2.** Prove that $S(SD)^{\square} = Pow$.

**Exercise 23.3.** Define a function $p\colon \mathbb{N} \to \mathbb{N}$ by the rules $p(0) = 0$, $p(1) = 1$, and $p(n)$ is the smallest prime factor of $n$ when $n \geq 2$. Justify the claim that $p$ is computable by describing a rough algorithm for computing $p$. (Does your algorithm use unbounded searches?)

**Exercise 23.4.** Prove that there exists a function $f\colon \mathbb{N} \to \mathbb{N}$ that is not computable.

**Exercise 23.5.** Let $X \subseteq \mathbb{N}$. Prove that if the characteristic function $\chi_X$ is computable (or, alternatively, primitive recursive), then so is $\chi_{\mathbb{N}-X}$.

**Exercise 23.6.** Compute $A(2,2)$.

**Exercise 23.7.** In this exercise we outline a proof that the Ackermann function is not primitive recursive.

Let us say that a function $f\colon \mathbb{N} \to \mathbb{N}$ is *majorized* by $A$ if there exists some constant $c_f \in \mathbb{N}$ (depending on $f$) such that

$$f(x) < A(c_f, x).$$

(1) Prove that $A$ is a strictly increasing function in both coordinates.
(2) Prove that $A$ majorizes $S$.
(3) Prove that if $A$ majorizes two functions $f, g\colon \mathbb{N} \to \mathbb{N}$, then it majorizes $f \dot{-} g$, $fg$, and $f^{\square}$.
(4) Prove that $A$ does not majorize $A \circ J^{-1}$.

# 24 Computers and decision problems

## 24.A Informal computing machines

Computer programs can accept a wide assortment of inputs. Usually those inputs are encoded as binary strings, but such an encoding is not strictly necessarily. Some programs will take numbers as inputs, others will take Chinese sentences, and others will take no input at all. Programs can even take other computer programs as inputs.

The output of a computer program often depends on the input. For instance, consider the program that immediately terminates if the input is 0, but prints "I will never end" to the screen, repeatedly, under any other input. Code for such a program might look like:

```
n:=UserInput,
While[n!=0,
   Print[I will never end]
]
```

A program is said to *halt* if it ceases activity. The code above leads to halting in only one case. Coders know that nonhalting is one of the biggest annoyances in testing code. Thus, it would be beneficial if there were some "super" program that could check if code was free from infinite loops, unending searches, or other nonhalting behaviors. This is the *halting problem*; it turns out that this task is impossible, as we will show in Section 26.

Technically, the code above is mere syntax. For it to run properly we need it to run on a machine that has been designed to interpret the code as intended. The machine thus provides the semantic meaning to the code's syntax. Turing proved that one only needs a single "universal" machine to run all programs (up to encoding in $\mathbb{N}$), rather than different machines for different programs—something that we take for granted nowadays.

For simplicity we fix our programming language to use only the symbols 0 and 1. This is essentially what humans do in practice, after using an ASCII table to convert strings of bits to a larger alphabet, and vice versa. We can view all finite strings of 0's and 1's as programs. Some of the programs will be gibberish, as interpreted by our machine. We can design our machine to halt on gibberish or continue forever, it makes no difference. However, we do assume that our machine can implement standard operations, such as flipping bits, implementing while loops, and so forth.

To this point we have made no extravagant assumptions. However, there are two implicit philosophical assumptions. First, to talk about halting we implicitly assume that a machine could continue running forever. Second, we assume that our machine has unbounded memory, otherwise there is no way to store arbitrarily large numbers or calculate complicated functions. While no computer built by humans has either of these properties, it seems at least conceivable that a machine could be built to repair itself and continue to add memory as necessary. There is the question of whether reality itself is finite, so the type of machine we are positing is at present merely a mathematical abstraction that resembles an actual computer.

## 24.B Decision problems

An old problem in number theory is to classify the natural numbers that are sums of three integer cubes, $n = x^3 + y^3 + z^3$. The first few natural numbers are easy:

$$\begin{aligned} 0 &= 0^3 + 0^3 + 0^3 \\ 1 &= 1^3 + 0^3 + 0^3 \\ 2 &= 1^3 + 1^3 + 0^3 \\ 3 &= 1^3 + 1^3 + 1^3 \end{aligned}$$

What about $n = 4$? There is a mindless algorithm that one can run to look for solutions to the equation $n = x^3 + y^3 + z^3$. Take any recursively defined well-ordering of $\mathbb{Z}^3$ of order type $\omega$, and just check each option $(x, y, z) \in \mathbb{Z}^3$, one at a time, in order. For a given value of $n$ it is possible that no solution would be found, and thus this algorithm would continue forever without stopping; such is the case for $n = 4$. In fact, one can easily check that every cube is $\equiv 0, 1, -1 \pmod 9$. Thus, sums of three cubes cannot be congruent to 4 or 5 modulo 9, and we may immediately rule out those two cases.

The next few examples are again quite easy to find.

$$\begin{aligned} 6 &= 2^3 + (-1)^3 + (-1)^3 \\ 7 &= 2^3 + (-1)^3 + 0^3 \\ 8 &= 2^3 + 0^3 + 0^3 \\ 9 &= 2^3 + 1^3 + 0^3 \\ 10 &= 2^3 + 1^3 + 1^3 \\ 11 &= 3^3 + (-2)^3 + (-2)^3 \end{aligned}$$

Some additional examples like $29 = 3^3 + 1^3 + 1^3$ are also easy to find. However,

$$42 \;=\; (-80538738812075974)^3 + 80435758145817515^3 + 12602123297335631^3,$$

took over a million hours of distributed computation to find, even with major improvements to the algorithm.

*Finding* solutions to the equation $n = x^3 + y^3 + z^3$, for different values of $n$, thus appears to be a difficult problem. However, it is conjectured that it is much easier to *decide* whether or not a solution exists.

**Conjecture 24.1.** *A natural number $n$ is a sum of three integer cubes if and only if $n \not\equiv 4, 5 \pmod 9$.*

The previous discussion motivates the following definitions.

**Definition 24.2.** Let $S \subseteq \mathbb{N}$. We say that $S$ is *semi-decidable* if there is a computer program such that when given $n \in S$ as input it will print 1 and halt, and when given $n \in \mathbb{N} - S$ as input either prints 0 and halts, or does not print anything and never terminates. If the program can be chosen to always halt, then we say $S$ is *decidable*.

**Remark 24.3.** Semi-decidable sets are often called *recursively enumerable* in the literature, and we will do the same. Decidable sets are sometimes called *recursive*, but we won't use that terminology to avoid overload. ▲

**Example 24.4.** The set of odd natural numbers is decidable. The reader is left with the easy task of coming up with an algorithm that when run on an idealized computer will correctly decide whether or not a natural number is odd. △

**Example 24.5.** The "mindless algorithm" we described at the beginning of this subsection shows that the set of natural numbers that are sums of three integer cubes is semi-decidable. It is currently an open problem whether this set is decidable, but the set is decidable if Conjecture 24.1 is true. △

There is an interesting characterization of decidable sets in terms of the semi-decidable sets.

**Proposition 24.6.** *A set $S \subseteq \mathbb{N}$ is decidable if and only if both $S$ and $\mathbb{N} - S$ are semi-decidable.*

*Proof.* ($\Rightarrow$): Let $A$ be an algorithm that correctly determines whether or not $n \in \mathbb{N}$ belongs to $S$. This algorithm clearly shows that $S$ is semi-decidable. Let $A'$ be the new algorithm where we run $A$ and then switch the end values before printing them. This new algorithm shows that $\mathbb{N} - S$ is semi-decidable.

($\Leftarrow$): Let $B_1$ be an algorithm which shows that $S$ is semi-decidable, and similarly let $B_2$ be an algorithm for $\mathbb{N} - S$. Given any $n \in \mathbb{N}$ as an input, we can run a new algorithm $A$ that alternates between checking the next step of $B_1$ and the next step of $B_2$, on that input. Either $B_1$ or $B_2$ will eventually print 1, at which point we stop and have $A$ print 1 or 0 accordingly. □

Let $S$ be a semi-decidable set, and let $P$ be a computer program demonstrating this fact. We can think of $P$ as a partial function $g_P \colon \mathbb{N} \to \mathbb{N}$, where

$$g_P(n) = \begin{cases} 1 & \text{if } P \text{ prints 1 when given input } n, \\ 0 & \text{if } P \text{ prints 0 when given input } n, \\ \text{undefined} & \text{otherwise.} \end{cases} \tag{24.7}$$

Modifying the program $P$ to enter an infinite loop if it was going to print 0, we can even guarantee that $S$ is exactly the *domain* of the partial function $g_P$. We claim, without proof, that those partial functions $\mathbb{N} \to \mathbb{N}$ that can be implemented by a computer algorithm—with the possibility of failing to halt—are exactly the computable partial functions. Thus $g_P$ is computable.

Semi-decidability is related to computability in another way.

**Theorem 24.8.** *A nonempty set $S \subseteq \mathbb{N}$ is semi-decidable if and only if it is the* range *of a computable, total function $f \colon \mathbb{N} \to \mathbb{N}$.*

*Proof.* ($\Leftarrow$): Assume $f\colon \mathbb{N} \to \mathbb{N}$ is a computable function whose range is $S$. To show that $S$ is semi-decidable, use the following algorithm (noting that $f$ can be computed):

```
n:=UserInput,
m:=0,
While[n!=f(m),
   Increment[m]
],
Print[1]
```

($\Rightarrow$): Assume $P$ is some algorithm demonstrating that $S$ is semi-decidable. Since $P$ may not halt on a given input, we should only ever run the first few steps of $P$ to avoid nonhalting. The idea is that we will keep running $P$ on different inputs, for an increasing number of steps. Consider the function described by the following (rough) algorithm $Q$:

```
n:=UserInput,
NumberOfSteps:=1,
CountPrintings:=0,
While[True,
   For[Input ranging from 0 to NumberOfSteps,
      If[P[Input] prints 1 within NumberOfSteps,
         If[n>CountPrintings,
            Increment[CountPrintings],
         Else
            Print[Input] and HALT
         ]
      ]
   ],
   Increment[NumberOfSteps]
]
```

Since $S \neq \emptyset$, we may fix some $m \in S$. The algorithm $P$ must print 1 on input $m$, say after $k_m$ steps. Hence, the algorithm $Q$ will start incrementing the "CountPrintings" variable once the "NumberOfSteps" variable is as large as $\max(m, k_m)$, or $Q$ will have already halted. This means $Q$ always eventually halts. Also, $Q$ only prints a natural number when $P$ prints 1 on that input; thus $Q$ only prints elements of $S$.

Finally, we need to check that every element $m \in S$ is one of the outputs of $Q$. Let $Q'$ be $Q$, except that we always do the "else" command in the inner "if-else" routine, and we remove `and HALT`. We see that once "NumberOfSteps" is large enough (as in the previous paragraph) then $m$ will be printed by $Q'$, among other numbers. Thus, each of the elements of $S$ is printed, repeatedly, by $Q'$. Number the positions of the printed outputs of $Q'$, in order, using $\mathbb{N}$. If $m$ occurs at the $n$th position, then we see that $Q$ prints $m$ on input $n$. Thus, $Q$ describes a computable function whose range is $S$. □

Theorem 24.8 is the reason that semi-decidable sets are called recursively enumerable. These are exactly the sets of numbers that can be printed by running a program (such as $Q'$, in the previous proof) on an idealized computing machine. With some minor modifications, one can avoid repetitions in the printing.

## 24.C General decidability

Given $S \subseteq \mathbb{N}$, asking whether or not $n \in \mathbb{N}$ is a member of $S$ is just one of many questions with a yes/no answer. Generally, one might call these *decision problems*, and we are interested in concrete methods to solve them. If we can encode such problems in terms of sets of natural numbers, then we are just asking whether these sets are (semi-)decidable.

Consider the following example. Let $\mathscr{L}$ be the language of set theory, with signature $(\in)$ and variables $x_n$ (one for each $n \in \mathbb{N}$). In Section 22 we defined a bijection $\#\colon \mathrm{String}_{\mathscr{L}} \to \mathbb{N}$. This bijection is recursively defined, and it allows us to mechanically translate questions about strings of symbols in the language $\mathscr{L}$ to questions about sets of natural numbers.

One obvious decision problem in this context is membership in the subset

$$\mathrm{Form}_{\mathscr{L}} \subseteq \mathrm{String}_{\mathscr{L}}.$$

Formulas in first order logic are defined recursively using simpler formulas, and there is indeed a mechanical procedure for determining whether or not a string of symbols from $\mathscr{L}$ is a well-formed formula. (You might attempt Exercise 24.5 now.) Equivalently, the image of the set $\mathrm{Form}_{\mathscr{L}}$ under $\#$ is a decidable set of natural numbers. (This is the essence of the bonus problem in Exercise 22.9.)

It takes a bit more work to verify that the subset ZFC-Axioms $\subseteq \mathrm{Form}_{\mathscr{L}}$, consisting of the formal axioms of set theory, is also a decidable set.

Note that if we change the bijection $\#$ to a random bijection $f\colon \mathrm{String}_{\mathscr{L}} \to \mathbb{N}$, then the image of $\mathrm{Form}_{\mathscr{L}}$ under $f$ may not be a decidable set of natural numbers. This corresponds, roughly, to our inability to write a computer program that handles $f$ well.

## 24.D Exercises

**Exercise 24.1.** Given $m, n \in \mathbb{N}$ with $m < n$, the set

$$S_{m,n} = \{k \in \mathbb{N} : k \equiv m \pmod{n}\}$$

is semi-decidable. Justify this claim, and then leverage the claim to show that $S_{m,n}$ is decidable.

**Exercise 24.2.** Prove that any finite subset of $\mathbb{N}$ is decidable.

**Exercise 24.3.** Let $S$ be a (possibly empty) semi-decidable subset of $\mathbb{N}$. Describe an algorithm whose output consists of the elements of $S$, each printed exactly once.

**Exercise 24.4.** Let $S \subseteq \mathbb{N}$. Prove that there is an algorithm that prints only the elements of $S$ in order (under the usual ordering of $\mathbb{N}$) if and only if $S$ is decidable.

**Exercise 24.5.** Describe an algorithm that decides whether or not a string of symbols in the language of set theory is a well-formed formula. Next, describe a rough algorithm that decides whether or not a well-formed formula is an axiom of logic or of ZFC. (The hardest part is handling the axiom schemas.)

**Exercise 24.6.** Describe an algorithm that checks whether or not a string of symbols in the language of set theory is a concatenation of formulas. Then expand the algorithm to check if those formulas are the steps in a formal proof. (Hint: Your algorithm should check that each formula is either an axiom, an application of modus ponens to two previous formulas, or an application of universal generalization to a previous formula.)

The previous exercise shows that a computer can, in theory, check formal proofs. Putting that theory into practice, there are some communities of mathematicians and computer scientists that have devoted a lot of work into formalizing, and then verifying, numerous proofs.

**Exercise 24.7.** Explain why the valid formulas of ZFC form a semi-decidable subset of $\mathrm{Form}_{\mathscr{L}}$, where $\mathscr{L}$ is the language of set theory. (We will see later that this is not a decidable subset. Thus, there is no algorithm to decide, once and for all, whether or not an arbitrary formula has a proof. This may partly explain your own experience with the difficultly of finding proofs.)

# 25 Encoding structures into set theory

Let $\mathbb{N}$ denote the meta natural numbers and let $\omega$ be the formal set of natural numbers. Let $\mathbb{N} \to \omega$ be the usual encoding $n \mapsto \underline{n}$, defined recursively by the rules $\underline{0} = \emptyset$ and $\underline{n+1} = \underline{n} \cup \{\underline{n}\}$. We will call this function the *underline* function (or map). This meta map allows us to encode (from a metatheoretical perspective) each of the meta natural numbers as objects in formal set theory. In this section, we will study how to encode more complicated structures, such as meta sets of numbers and functions, into formal set theory. In turn, the answers to some decision questions about formal set theory are ultimately reflected in how well these encodings behave.

To simplify formulas in this section, for any set $x$ we let $x + 1$ denote the set $x \cup \{x\}$.

## 25.A Encoding meta sets of numbers

One of the simplest subsets of $\mathbb{N}$ is $\mathbb{N}$ itself. Consider what it would mean to encode the set $\mathbb{N}$ into formal set theory. The underline map $\mathbb{N} \to \omega$ is injective, assuming that ZFC is consistent. However, it is not necessarily surjective. In fact, assuming that ZFC is consistent, we can guarantee that there is some model $\mathcal{M}$ such that the map $\mathbb{N} \to \omega_{\mathcal{M}}$ is not surjective. What this means is that $\omega$ is only an *approximate* encoding of $\mathbb{N}$. It captures the image of $\mathbb{N}$, but it might also contain other, extraneous elements.

Finite meta sets of numbers do not have this problem. For instance, restricting the underline map to the first three meta numbers gives a bijection $\{0, 1, 2\} \to \{\underline{0}, \underline{1}, \underline{2}\}$. (Note that the set on the right is easily shown to be a formal set; it is just $\underline{3}$.) However, when working with infinite sets, it is difficult to guarantee that the encoding is a bijection. This leads us to carefully study what we can say about our encoding.

To start, we should ask ourselves what $\omega$ is, exactly. It is a derived constant symbol that is not part of the signature. Similarly, $\underline{0}, \underline{1}, \ldots$ are derived constant symbols. Derived symbols are obtained using formulas from the formal language that give them their meaning. Thus, we can always replace the shorthand formula $x \in \omega$ with the primitive predicate $\varphi_{\in\omega}(x)$ that tells us exactly the condition for $x$ to belong to $\omega$. The predicate we (implicitly) used in Definition 3.3 is the one saying that $x$ belongs to every inductive set. Thus, we can write $\varphi_{\in\omega}(x)$ as

$$\forall X \left( \Big( \emptyset \in X \wedge \forall y \in X\, (y + 1 \in X) \Big) \to x \in X \right), \tag{25.1}$$

and we can thus treat $\omega$ as the class $\{x : \varphi_{\in\omega}(x)\}$. (Of course, "$\emptyset$" and "$y \cup \{y\}$" are also derived symbols, so a little more unpacking is needed if we want to remove all derived notation.)

Note that, *a priori*, this only defines $\omega$ as a class. Our proof that $\omega$ is a set, as given in Section 3, relies on additional reasoning using the axioms of ZFC. In particular, we needed the axiom of infinity to guarantee that the class of inductive sets is nonempty. The theory ZFC is also strong enough to necessitate that $\omega$ contains the encoded meta numbers, leading to the following theorem schema.

**Metatheorem 25.2.** *For each meta number $n \in \mathbb{N}$, it holds that $\models_{\mathrm{ZFC}} \underline{n} \in \omega$.*

*Proof.* It is a tautology that $\varphi_{\in\omega}(\emptyset)$ holds, where $\varphi_{\in\omega}$ is given by (25.1). Therefore, we have $\models_{\mathrm{ZFC}} \underline{0} \in \omega$.

Assume, by induction (in the meta theory), that $\models_{\mathrm{ZFC}} \varphi_{\in\omega}(\underline{k})$ holds for some $k \in \mathbb{N}$. By Theorem 3.2, which was proven using only axioms from ZFC, we know that $\omega$ is an inductive set. Thus ZFC models $\underline{k+1} := \underline{k} \cup \{\underline{k}\} \in \omega$. (Alternatively, directly show that $\models_{\mathrm{ZFC}} \varphi_{\in\omega}(\underline{k}) \to \varphi_{\in\omega}(\underline{k+1})$, and use modus ponens.) □

Thinking about this metatheorem another way, we might say: In every model of ZFC, the class $\omega = \{x : \varphi_{\in\omega}(x)\}$ looks enough like $\mathbb{N}$ to contain all the encoded natural numbers. More generally, given an arbitrary predicate $\psi$, then consider the new class $\{x : \psi(x)\}$. We might hope ZFC models that this new class contains an encoding of some subset of $\mathbb{N}$. This motivates the following general definitions.

**Definition 25.3.** Let $\varphi(x)$ be a predicate in the language of set theory, let $T$ be a collection of axioms from that language, and let $S \subseteq \mathbb{N}$. We say that $\varphi$ *weakly encodes* membership in $S$ (relative to the theory $T$) when

$$\begin{cases} \models_T \varphi(\underline{n}) & \text{if } n \in S, \text{ and} \\ \models_T \neg\varphi(\underline{n}) \text{ or } \not\models_T \varphi(\underline{n}) & \text{if } n \in \mathbb{N} - S. \end{cases}$$

If the condition "or $\not\models_T \varphi(\underline{n})$" is removed, then we say that $\varphi$ *strongly encodes* membership in $S$.

**Example 25.4.** The predicate $\varphi_{\in\omega}$ strongly encodes membership in $\mathbb{N}$, relative to (just a few of) the axioms of ZFC, by Metatheorem 25.2.

Note that (weak or strong) encodability makes no claims about the predicate when a value other than an encoded natural number is plugged in for $x$. The definition is neutral about such matters. In particular, any "extraneous" elements of $\omega$ that are not encoded natural numbers are irrelevant.

Another consequence of this fact is that we can sometimes strongly encode membership with very simple formulas. For instance, the predicate $x = x$ strongly encodes membership in $\mathbb{N}$, since $\models_{\mathrm{ZFC}} \underline{n} = \underline{n}$ for each $n \in \mathbb{N}$. △

**Example 25.5.** Suppose we wish to strongly encode membership in $S := \{0\} \subseteq \mathbb{N}$ relative to ZFC. The simple predicate

$$x = \underline{0}$$

will work. Given $n \in \mathbb{N}$, it is easy to show that $\models_{\mathrm{ZFC}} \underline{n} = \underline{0}$ only when $n = 0$; more generally we will now answer the question of exactly which axioms $T \subseteq \mathrm{ZFC}$ are needed to show that this predicate strongly encodes membership in $S$.

First, note that in order to talk about the underline map, it seems that we need the axioms of empty set, pairing, and union; otherwise, the derived sets

$$\underline{0} = \emptyset,\ \underline{1} = \{\emptyset\},\ \underline{2} = \{\emptyset, \{\emptyset\}\},\ \underline{3} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \ldots$$

may not even exist. In other words, without those three axioms, there may not be a well-defined underline map. Hence, we will assume that $T$ contains these three axioms. (Alternatively, as we will do in later sections, we could adjoin to our language $\mathscr{L}$ new, primitive, constant symbols $\underline{0}, \underline{1}, \ldots$, and then we wouldn't need any axioms to talk about our underline map.)

Now, clearly $\models_T \underline{0} = \underline{0}$. Thus, if $n \in S$, then $\models_T \underline{n} = \underline{0}$. When $n \notin S$, then we want $\models_T \underline{n} \neq \underline{0}$. This will hold when $T$ contains the extensionality axiom, since $\underline{0}$ has no elements, but $\underline{0} \in \underline{n}$ when $n > 0$. △

**Example 25.6.** There is a predicate $E(x)$ that strongly encodes membership in the set of even natural numbers $S = 2\mathbb{N}$. Naively, one might want to take $E(x)$ to be

$$(x = \underline{0}) \vee (x = \underline{2}) \vee (x = \underline{4}) \vee \cdots,$$

but unfortunately this is not a well-formed formula. Instead, we can pattern our formula after the predicate $\varphi_{\in\omega}$, and use

$$\forall X \Big( \big( \emptyset \in X \wedge \forall y \in X \big( (y+1) + 1 \in X \big) \big) \to x \in X \Big).$$

It takes some effort to show that this predicate does indeed strongly encode $2\mathbb{N}$.

Alternatively, using formal recursion we can define the formal even natural numbers by the rules $e_{\underline{0}} = \underline{0}$ and $e_{x+1} = (e_x + 1) + 1$ for each $x \in \omega$. By an axiom of replacement, $\{e_x : x \in \omega\}$ is a set. The formula that describes membership in this set will yield the desired predicate.

As a third alternative, we can define ordinal addition, and then the predicate we want is simply $\exists y \in \omega,\ x = y + y$. △

**Example 25.7.** Let $\varphi(x)$ be any predicate and let $S \subseteq \mathbb{N}$ be any subset. Let $T$ be the collection of axioms $T_0 \cup \{\varphi(\underline{n})\}_{n \in S} \cup \{\neg\varphi(\underline{n})\}_{n \in \mathbb{N} - S}$, where $T_0$ is some (initial) set of axioms.

Relative to $T$, the predicate $\varphi$ strongly encodes membership in $S$. Thus, it is always possible to extend a theory $T_0$ to guarantee that any fixed predicate strongly encodes membership in any fixed subset of $\mathbb{N}$. Often (but not always), the theory $T$ constructed this way will be inconsistent. In any inconsistent theory every predicate strongly encodes membership in every subset of $\mathbb{N}$. △

## 25.B Connections to decidability

Looking again at Definition 25.3, you may notice some similarity between the definitions of (weak) encodability and (semi-)decidability. This is not an accident. When $T = \mathrm{ZFC}$, then we can encode primitive recursive and computable functions into formal set theory, using formal induction and recursion. This encoding allows us to identify (semi-)decidable sets as exactly those whose membership is (weakly) encoded.

**Metatheorem 25.8.** *Assume that ZFC is a consistent theory. Relative to ZFC, membership in a set $S \subseteq \mathbb{N}$ is weakly (or strongly) encodable if and only if $S$ is semi-decidable (respectively, decidable).*

*Proof idea.* The technical details of this theorem are many, so we only hint at the steps needed to prove it.

($\Rightarrow$): Suppose $\varphi$ is a predicate that weakly encodes membership in $S$. We can write a computer program that takes $n$ as input, and searches through all possible derivations (from ZFC) for a proof of $\varphi(\underline{n})$. If it finds such a proof, the program prints 1 and halts; otherwise it keeps looking. This algorithm shows that $S$ is semi-decidable. (Note: This algorithm exists only because the ZFC axioms are a decidable set. If ZFC is replaced by a more complicated theory, whose axioms are not decidable, then this argument will fail.)

When membership in $S$ is strongly encoded by $\varphi$, modify the algorithm to look for proofs of both $\varphi(\underline{n})$ and $\neg\varphi(\underline{n})$. Print 1 or 0 according to which of the two statements eventually is shown to have a proof.

($\Leftarrow$): First, generalize Definition 25.3 to allow encodability for membership in subsets $X \subseteq \mathbb{N}^k$, for any $k \in \mathbb{N}$. (In other words, think of $X$ as a $k$-ary relation, which we want to encode into set theory.) Thus, we can talk about encoding functions $f\colon \mathbb{N} \to \mathbb{N}$ (since such functions are just special subsets of $\mathbb{N}^2$). Next, prove that all computable functions are strongly encodable. (The successor function is strongly encodable, and the arithmetic subtraction, composition, and iteration of strongly encoded functions are strongly encodable. The same is true for first inversion of a strongly encoded surjective function.) If $S$ is decidable by a program $P$, let $g_P$ be the total function from (24.7). By Exercise 25.6, we may strongly encode this function by a class function $\varphi(x, y)$. Now $\varphi(x, \underline{1})$ strongly encodes $S$.

If $S$ is only semi-decidable, then $g_P$ is a partial function. The argument above can be modified to this case, and was first handled by Robert Ritchie and Paul Young, but given a slightly simpler proof by William Ritter. □

## 25.C Exercises

**Exercise 25.1.** Let $S \subseteq \mathbb{N}$ be a finite set. Give a simple predicate demonstrating that membership in $S$ is strongly encoded relative to ZFC. (Hint: Write $S = \{a_1, a_2, \ldots, a_n\}$.)

**Exercise 25.2.** Let $S \subseteq \mathbb{N}$ be a *cofinite* set, meaning that $\mathbb{N} - S$ is finite. Show that membership in $S$ can be strongly encoded relative to ZFC, using a simple predicate.

**Exercise 25.3.** Find a predicate $\varphi_{=\omega}(x)$, in the primitive language of set theory, that describes the shorthand formula $x = \omega$ relative to ZFC.

**Exercise 25.4.** Without worrying about primitivity, and in particular allowing the use of multiplication on ordinals, find a formula that strongly encodes membership in the set of primes.

Let $\mathscr{L}$ be a first order language. When talking about a derived constant symbol, then this means one of two (essentially equivalent) things:

(1) We expand the language of $\mathscr{L}$ to the new language $\mathscr{L}(c)$ that now includes the new constant symbol $c$, and we add axioms to our theory that describe the nature of $c$.

(2) We continue to work in the original language $\mathscr{L}$, and we only use $c$ as a shorthand in formulas.

In practice, most mathematicians follow (1) and add constant symbols quite freely. In the next exercise, we will demonstrate this idea by defining the derived constant symbol $\underline{0}$ in both ways.

**Exercise 25.5.** Find three primitive formulas in the language of set theory that are logically equivalent to the three shorthand formulas $\underline{0} = \underline{0}$, $x = \underline{0}$, and $\underline{0} = y$, where $x$ and $y$ are variables. Do the same for the three shorthand formulas $\underline{0} \in \underline{0}$, $x \in \underline{0}$, and $\underline{0} \in y$. Explain how these six formulas can be used repeatedly to remove any instance of the derived symbol $\underline{0}$ from a complicated formula, such as

$$\forall x\ ((x = \underline{0}) \vee \exists y\ ((\underline{0} \in y) \wedge (x = y))).$$

Alternatively, when working in the expanded language $\mathscr{L}(c)$, what axioms would you add to ZFC to guarantee that the constant symbol $c$ is interpreted as the empty set?

**Exercise 25.6.** Generalize Definition 25.3 to subsets $X \subseteq \mathbb{N}^k$, for any $k \in \mathbb{N}$. If the relation $X$ is a $(k-1)$-ary total function, and is strongly encoded (relative to the formal set theory axioms), then show it is strongly encoded by a class function.

**Exercise 25.7.** Find a shorthand formula that strongly encodes the (meta) successor function, relative to ZFC. (This shorthand formula might be very short.)

**Exercise 25.8.** Where is consistency used in the proof of Metatheorem 25.8?

# 26 Diagonalization and two impossible tasks

In 1891 Cantor published his proof that it is impossible to find a bijection between $\mathbb{R}$ and $\mathbb{N}$. His method, nowadays called *Cantor's diagonalization argument*, is an important tool for any mathematician to keep in mind. In this section we will review this argument, abstract from it the key ideas, and apply it to show that two other tasks are similarly impossible.

## 26.A Counting sequences

There are some technical subtleties to Cantor's proof that $\mathbb{R}$ is uncountable, arising from the fact that the definition of $\mathbb{R}$ is complicated. To avoid those issues, we will instead work with the simpler set of ($\mathbb{N}$-indexed) sequences in the symbols 0 and 1; we call this set $S$. A typical element of $S$ might look like

$$s = (0, 1, 1, 0, 0, 0, 1, 0, 1, 1, \ldots).$$

For each $n \in \mathbb{N}$, the sequence $s$ has an $n$th coordinate, which we can write as $s(n)$. Thus, for the specific sequence given above, we have $s(0) = 0$, $s(1) = 1$, $s(2) = 1$, and so forth.

Cantor's argument is as follows: Let $s_0, s_1, s_2, \ldots$ be any collection of elements from $S$ indexed by $\mathbb{N}$. We form the new element $t \in S$ by the rule $t(n) = \neg(s_n(n))$. (Here, "$\neg$" is just a convenient notation for the map that takes $0 \mapsto 1$ and $1 \mapsto 0$.) We see that $t$ does not appear in our list since it differs from $s_n$ on the $n$th coordinate. As the list $s_0, s_1, s_2, \ldots$ was arbitrary, we see that no enumeration by $\mathbb{N}$ captures all elements of $S$.

A typical example of this process—illustrating why it is called a *diagonalization* argument—is given in the following chart.

$$\begin{array}{rcl}
s_0 & = & (0, 1, 1, 0, 0, 0, \ldots) \\
s_1 & = & (1, 1, 1, 1, 1, 1, \ldots) \\
s_2 & = & (1, 0, 1, 0, 1, 0, \ldots) \\
s_3 & = & (1, 1, 0, 0, 0, 1, \ldots) \\
s_4 & = & (0, 0, 0, 0, 0, 0, \ldots) \\
s_5 & = & (0, 0, 0, 1, 0, 0, \ldots) \\
 & \vdots & \\
t & = & (1, 0, 0, 1, 1, 1, \ldots)
\end{array}$$

What does it take to make this argument work? First, we needed a method of flipping digits, or equivalently of turning true to false and vice versa. We needed the negation operation. Second, the set $\mathbb{N}$ plays two separate roles. We used it to enumerate a list of elements from $S$, but it also indexed coordinates of elements of $S$. This allowed us to "plug" the subscript of an enumerated object back into itself and consider the $n$th coordinate of the $n$th sequence $s_n(n)$.

The fact that there were only two available digits, 0 and 1, was not essential. A similar argument using more digits still works; just replace negation by a more complicated digit-change function. Also, using $\mathbb{N}$ as the indexing set was not strictly necessary either, as the following theorem demonstrates. (The following theorem is essentially equivalent to Theorem 21.3.)

**Theorem 26.1.** *Let $S$ be a set, and let $2^S$ denote the collection of all functions $f\colon S \to \{0,1\}$. Then $|S| < |2^S|$.*

*Proof.* Let $g\colon S \to 2^S$ be any function. We will show that $g$ is not surjective. For convenience, for each $s \in S$ write $g(s) = f_s$.

Let $h \in 2^S$ be defined by the rule

$$h(s) = \neg(f_s(s)).$$

We see that $h$ is not in the image of $g$, since for each $s \in S$ it disagrees with the function $g(s) = f_s$ at the point $s$. □

There are many other impossible tasks that arise from Cantor's argument. We will describe other important examples in this and the following sections, but the reader is encouraged to think about even more possible applications of these ideas.

## 26.B The halting problem

The collection of finite strings of 0's and 1's is well-ordered by the "shortlex" order. Strings with shorter length occur before those of longer length, whereas strings with the same length are ordered lexicographically. Computer programs are nothing more than strings of 0's and 1's. Thus, the list of all computer programs is enumerated as follows.

$$\begin{array}{rcl} P_0 & = & \qquad \text{(the empty string)} \\ P_1 & = & 0 \\ P_2 & = & 1 \\ P_3 & = & 00 \\ P_4 & = & 01 \\ P_5 & = & 10 \\ P_6 & = & 11 \\ & \vdots & \end{array}$$

(We will assume that we are running these programs on an idealized machine, as explained earlier.) One can write a program that when given input $n$ outputs $P_n$. (The reader who is familiar with real-life coding will see that such a task is not difficult. If you doubt that this is a computable function, see if you can mechanically compute this function yourself; for instance, try to find $P_{17}$.)

Recall the bijection $J\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined in Example 23.11. This function, and its inverse, are computable. We write $J^{-1}(k) = (k_1, k_2) \in \mathbb{N} \times \mathbb{N}$ for ease of notation.

Suppose, by way of contradiction, that the halting problem could be solved by a computer program $Q$. Formally, if $Q$ is given input $k$, then it outputs 0 if $P_{k_1}$ halts on input $k_2$, otherwise it outputs 1. Now consider the new program $Q'$ that takes a natural number $n$ as an input, and does the following process. It first runs $Q$ on input $k = J(n, n)$. Thus, it checks to see if $P_n$ halts on its own index $n$. If $Q$ says that $P_n$ halts on input $n$, then $Q'$ prints to the screen "I will never end" forever. On the other hand, if $Q$ says that $P_n$ does not halt on input $n$, then $Q'$ prints to the screen "I am done" and then terminates. In other words, with respect to halting, $Q'$ has the opposite behavior on $n$ that $P_n$ does.

Now comes the punchline. The new program $Q'$ is just some $P_n$, but by construction it behaves differently than any $P_n$, which yields the needed contradiction.

**Remark 26.2.** Given some fixed $n \in \mathbb{N}$ and fixed program $P$, is there a program $Q$ that correctly determines whether or not $P$ halts on input $n$? Surprisingly, yes. The program $Q$ is either "Print 0" or "Print 1"; in other words, $Q$ just accidentally solves the halting problem for $P$ on input $n$. Our solution to the general halting problem tells us that there is no computable way to decide which answer is correct. The set $\{n \in \mathbb{N} : P_n \text{ halts on input } n\}$ is semi-decidable, but not decidable. ▲

## 26.C Exercises

**Exercise 26.1.** Construct an algorithm that describes the function $n \mapsto P_n$, where $P_n$ is the $n$th string of 0's and 1's in the shortlex ordering. (Use it to find $P_{100}$.)

**Exercise 26.2.** Our solution to the halting problem shows that there is no algorithm for deciding whether or not a program $P$ will halt on an input $n \in \mathbb{N}$. Both the program $P$ and the input $n$ are allowed to range over all possibilities.

Prove the stronger claim that there is a fixed program $P^*$ such that no algorithm can decide whether or not $P^*$ will halt on an input $n \in \mathbb{N}$. (Of course, for some programs it is possible to decide if they halt or not for each $n \in \mathbb{N}$.)

**Exercise 26.3.** We will show that there is a computable function that is not primitive recursive, using diagonalization.

(1) Describe an algorithm that will list the primitive recursive functions $f_0, f_1, \ldots$, together with the code for implementing them. (As an extra problem, can you think how to make the list nonrepeating?)

(2) Show that $n \mapsto f_n(n) + 1$ is computable but not primitive recursive.

(A very important question that you should ask yourself is: What goes wrong if we try to repeat this argument with "primitive recursive" replaced by "computable"?)

# 27 Self-reference, part one

*Santa's Portrait*, ca. 1881, by Thomas Nast.

Consider the sentence, named $\sigma$, that says: If $\sigma$ is true, then Santa Claus exists.

If $\sigma$ is true, then the premise and implication are true, and so the conclusion must be true as well. On the other hand, an implication is false only if its premise is true while the conclusion fails. So if $\sigma$ is false then the premise must be true, leading to a contradiction. Thus, in all cases Santa Claus exists.

If you are unsatisfied with this proof that Santa exists, you are not alone. The sentence $\sigma$, and other similar sentences, lead to seemingly illogical conclusions. You may have also heard of the liar paradox arising from the sentence:

This sentence is false.

Historically, there have been many attempted solutions to these paradoxes. One is to deny that sentences can talk about themselves; but that would similarly rule out a large portion of normal human speech. Another is to deny that the sentence has a truth value at all, but that just begs the question why it won't have a truth value.

Another option, that seems only slightly less drastic, is to deny the ability to *talk* about truth. Indeed, we would be hypocrites to deny the ability to talk about truth, as we have done so repeatedly in this text. However, we only talked about truth using models. Truth was not defined inside the formal system itself, but relied on metamathematical interpretations. We will show that in certain contexts it is impossible to speak of truth and validity formally—that is, from inside the formal system.

## 27.A Naming strings using terms

We want a way for formal sentences to talk about themselves. This will help us communicate some (but not all) metamathematical concepts inside the formal system. One way to do this is using encodings. We saw previously that natural numbers could be encoded as sets. Further, there is a computable way to encode formulas as numbers, and so formulas can similarly be encoded as sets.

Generally, let $\mathscr{L}$ be any language of FOL; the signature will be unimportant. Let $\mathrm{String}_{\mathscr{L}}$ be the (metamathematical) collection of all strings of symbols from $\mathscr{L}$. Also let $\mathrm{ClTerm}_{\mathscr{L}}$ be the collection of closed terms; these are the terms involving no variables, such as the constant symbols. A *naming* of the strings is a (meta) function $\mathrm{String}_{\mathscr{L}} \to \mathrm{ClTerm}_{\mathscr{L}}$. Often, a naming is written as $\varphi \mapsto \ulcorner\varphi\urcorner$. The square quotes tell us that $\ulcorner\varphi\urcorner$ is the formal name being given to the string $\varphi$.

**Example 27.1.** We will construct a naming of the strings of symbols from formal set theory.

Let $\mathbb{N}$ denote the set of meta natural numbers. Let $\mathscr{L}$ be the language of set theory, whose signature is $(\in)$, and let $\mathscr{L}^*$ be the extended language of set theory where we have added constant symbols $\underline{0}, \underline{1}, \ldots$. By adding these constant symbols, we automatically have an "underline" map $\mathbb{N} \to \mathrm{ClTerm}_{\mathscr{L}^*}$ given by the rule $n \mapsto \underline{n}$. (If we work over ZFC, then we can define this underline map as in Subsection 22.D, without needing to add constant systems to the language. But that only complicates matters unnecessarily.)

Let ZFC$^*$ consist of the axioms of ZFC together with new axioms asserting that our new constant symbols satisfy the appropriate defining conditions. For instance, to properly define the constant symbol $\underline{0}$ we could adjoin the axiom

$$\forall x\ \neg(x \in \underline{0}).$$

The language $\mathscr{L}^*$ under the theory ZFC$^*$ acts essentially the same as the language $\mathscr{L}$ under ZFC. We call $\mathscr{L}^*$ a *conservative extension* of the language of set theory. Any model of ZFC$^*$ restricts to a model of ZFC, and conversely any model of ZFC extends uniquely to a model of ZFC$^*$, where the assignments of the new constant symbols are forced upon us.

In Subsection 22.D, we defined a recursive numbering of the strings from $\mathscr{L}$. One can modify that numbering to create a (slightly different) numbering of the strings from $\mathscr{L}^*$; see Exercise 27.2. Let $\#\colon \mathrm{String}_{\mathscr{L}^*} \to \mathbb{N}$ be such a recursive numbering.

Composing this $\#$ function with the underline map, we get a recursive naming $\ulcorner\cdot\urcorner\colon \mathrm{String}_{\mathscr{L}^*} \to \mathrm{ClTerm}_{\mathscr{L}^*}$.

For concreteness, let $\psi$ be the formula $\exists y\,(y \in x)$. Suppose, for sake of argument, that it is string number 103. Then its name is $\ulcorner\psi\urcorner = \underline{103}$, which is a formal constant. Every string has a number and a name. The number is a metamathematical concept, but the name is a formal constant symbol. Note that the string and its name can be quite different from one another, just as a cow is very different than the word "cow". △

## 27.B Diagonalization revisited

The key step in Cantor's diagonalization argument is using an index in two different ways. We can do something similar with namings. If $\psi(x)$ is a predicate (with one free variable $x$), then we can plug $\psi$'s name back into itself.

**Definition 27.2.** If $\ulcorner\cdot\urcorner$ is a naming, and $\psi$ is a predicate, the map $\psi \mapsto \psi(\ulcorner\psi\urcorner)$ is called the *diagonalization function*.

The diagonalization map sends predicates to formulas with no free variables.

**Example 27.3.** Let $\psi$ be the formula $\exists y\,(y \in x)$. Suppose, for sake of simplicity, that it is string number 103. The diagonalization function sends $\psi$ to the new formula $\psi(\ulcorner\psi\urcorner)$, which is just

$$\exists y\,(y \in \underline{103}).$$

This formula has no free variables, so it is a statement. It is valid in ZFC.

On the other hand, let $\theta$ be the formula $\forall y\,(y \in x)$. Suppose, for sake of simplicity, that this is string number 3937. The diagonalization function sends $\theta$ to the new formula $\theta(\ulcorner\theta\urcorner)$, which is just

$$\forall y\,(y \in \underline{3937}).$$

This formula has no free variables, so it is a statement. It is not valid in ZFC.

For both of these two predicates we *posited* a potential number, because it is difficult work to actually find their numbers. However, our numbering function is completely computable. With enough scratch paper (or by using a computer) we could find the actual numbers of these two formulas. △

The diagonalization function is metamathematical. We now want to encode it into our formal system. We do that as follows.

**Definition 27.4.** Let $\mathscr{L}$ be a FOL language, and suppose that

$$\ulcorner\cdot\urcorner\colon \mathrm{String}_{\mathscr{L}} \to \mathrm{ClTerm}_{\mathscr{L}}$$

is a naming. We say that a formula $D(x, y)$, whose free variables are $x$ and $y$, *strongly encodes* the diagonalization function (on names of predicates) relative to a theory $T$ when, for any predicate $\psi$, we have

$$\models_T \forall y\,(D(\ulcorner\psi\urcorner, y) \leftrightarrow y = \ulcorner\psi(\ulcorner\psi\urcorner)\urcorner). \tag{27.5}$$

The diagonalization function sends $\psi \mapsto \psi(\ulcorner\psi\urcorner)$. The formula $D$ in the previous definition similarly acts like a function, relating an input $x = \ulcorner\psi\urcorner$ only to the output $y = \ulcorner\psi(\ulcorner\psi\urcorner)\urcorner$. The idea of Definition 27.4 is that if the theory $T$ is powerful enough—and the naming is nice enough—then in *every* model of $T$, the interpretation of $D(x, y)$ in that model acts just like the diagonalization function.

**Example 27.6.** According to the proof of Metatheorem 25.8, assuming that ZFC is consistent, then we can strongly encode any computable function. This works just as well for the theory ZFC$^*$ in the extended language of set theory $\mathscr{L}^*$. The naming we described in Example 27.1 is computable. Also, we can computably identify the strings that are predicates (or closed terms). Thus, the diagonalization function is strongly encoded by some formula $D(x, y)$. (It takes some work, but that formula can be explicitly written down.) △

## 27.C Exercises

**Exercise 27.1.** Let $\mathscr{L}$ be the language of number theory, whose signature is $(S, \mathbf{0})$. Is there a naming $\mathrm{String}_{\mathscr{L}} \to \mathrm{ClTerm}_{\mathscr{L}}$? Is there a computable naming?

**Exercise 27.2.** Let $\mathscr{L}$ and $\mathscr{L}^*$ be as in Example 27.1. In Subsection 22.D we defined a recursive bijection $\mathrm{String}_{\mathscr{L}} \to \mathbb{N}$. Describe what must be done to modify that numbering to create a new numbering bijection $\#\colon \mathrm{String}_{\mathscr{L}^*} \to \mathbb{N}$. (As a difficult bonus problem, find the number for the formula $\forall x_2 \, (x_2 \in x_1)$, under both the old and the new numberings.)

For each of the remaining exercises of this section, let $\mathscr{L}^*$ be the extended language of set theory. Use the numbering just described and the naming from Example 27.1.

**Exercise 27.3.** Find a predicate $\psi$ that is not modeled by ZFC$^*$, but such that $\psi(\ulcorner \psi \urcorner)$ is modeled by ZFC$^*$. Do this without making unjustified assumptions about the numbering of formulas.

**Exercise 27.4.** Let $\varphi$ be the formula $x_1 \in \underline{0}$. Consider the following five concepts:

(1) $\varphi$,
(2) $\ulcorner \varphi \urcorner$,
(3) $\varphi(\ulcorner \varphi \urcorner)$,
(4) $\ulcorner \ulcorner \varphi \urcorner \urcorner$, and
(5) $\ulcorner \varphi(\ulcorner \varphi \urcorner) \urcorner$.

Which of these concepts makes sense? Which are formulas? Which are predicates? Which are closed terms? Are they all distinct (and why)?

**Exercise 27.5.** If $T$ is an inconsistent theory, is the diagonalization map strongly encoded relative to $T$? Explain.

**Exercise 27.6.** Let $D(x, y)$ be a strong encoding of the diagonalization function. Thus, if we plug in $x = \ulcorner \psi \urcorner$ for some predicate $\psi$, then there is a unique satisfying output $y = \ulcorner \psi(\ulcorner \psi \urcorner) \urcorner$. So, $D$ acts like a class function on the names of predicates.

Prove that if a formula $D(x, y)$ strongly encodes the diagonalization function relative to a theory $T$, then we can find another formula $D'(x, y)$ that also strongly encodes the diagonalization function relative to $T$, and $D'$ is a class function (on any input, not just the names of predicates).

**Exercise 27.7.** Definition 25.3 tells us what it means to weakly and strongly encode membership in a set of (meta) numbers. The diagonalization function is the (meta) relation

$$\mathscr{D} := \{(\psi, \psi(\ulcorner\psi\urcorner)) : \psi \text{ is a predicate}\} \subseteq \mathrm{String}^2_{\mathscr{L}}.$$

Since we have a numbering of strings, we can identify $\mathrm{String}^2_{\mathscr{L}}$ with $\mathbb{N}^2$. Under this identification, we identify $\mathscr{D}$ with a subset $\#\mathscr{D} \subseteq \mathbb{N}^2$. We can then say that a formula $D(x, y)$ strongly encodes membership in $\mathscr{D}$ (relative to a theory $T$) when

$$\begin{cases} \models_T D(\underline{m}, \underline{n}) & \text{if } (m, n) \in \#\mathscr{D}, \\ \models_T \neg D(\underline{m}, \underline{n}) & \text{if } (m, n) \in \mathbb{N}^2 - \#\mathscr{D}. \end{cases}$$

Is this definition any different from Definition 27.4? If so, explain any differences. If not, explain why not.

**Exercise 27.8.** Composing the numbering function $\#\colon \mathrm{String}_{\mathscr{L}^*} \to \mathbb{N}$ with the underline function $\mathbb{N} \to \mathrm{ClTerm}_{\mathscr{L}^*}$, we formed a naming $\ulcorner\cdot\urcorner\colon \mathrm{String}_{\mathscr{L}^*} \to \mathrm{ClTerm}_{\mathscr{L}^*}$. This is a meta map.

The meta map takes $\varphi \mapsto \ulcorner\varphi\urcorner$. A string of symbols $\varphi$ is not a formal set. Technically, the closed term output $\ulcorner\varphi\urcorner$ is also not a formal set, but it is always interpreted as a formal set in any model.

We can think of formally encoding the naming function $\varphi \mapsto \ulcorner\varphi\urcorner$, using the new function $\ulcorner\varphi\urcorner \mapsto \ulcorner\ulcorner\varphi\urcorner\urcorner$. Modify Definition 27.4 to define what it means to *strongly encode the naming function*, relative to a theory $T$.

# 28 Self-reference, part two

In the previous section we accomplished three important steps towards resolving the liar paradox and related conundrums.

First, we described what it means to name strings of symbols. This allows us to encode formulas as formal objects. In the extended language of set theory, this means that every formula is (computably) named by a formal natural number.

Second, we examined a construction mirroring an important step in Cantor's diagonalization argument. By focusing on the predicates, one may plug the name of a predicate back into itself. This creates the diagonalization function.

Third, we described what it would mean to (strongly) encode this meta diagonalization function formally, by using a formula. For set theory, if we use a computable naming of strings, then such an encoding of the diagonalization function exists under a sufficiently strong theory like ZFC$^*$.

In this section there are three more steps to take. We first describe an important consequence of being able to encode the diagonalization function. Next, we explain how to formally talk about the meta concepts of truth and validity. Finally, we put this all together, via work of Tarski, to show that formal systems are naturally limited in what they can express.

## 28.A A sentence asserting a property of itself

Fix a predicate $\psi$. When we plug the name of $\psi$ into $\psi$, we are applying the diagonalization map. More generally, for *any* string $\varphi$, we can plug the name of $\varphi$ into $\psi$ and create the new statement $\psi(\ulcorner\varphi\urcorner)$.

**Example 28.1.** For this example, we work in the extended language of set theory $\mathscr{L}^*$, with its usual numbering and naming.

Let $\psi(x)$ be the predicate $\exists y\ (y \in x)$. There are many possible strings of symbols in $\mathscr{L}^*$, such as

$$\varphi_1 := \ \text{(the empty string)},\ \ \varphi_2 := \underline{3} \in \underline{4} \in \underline{4},\ \ \text{and}\ \ \varphi_3 := (\underline{0} = \underline{0}).$$

Suppose, for sake of simplicity, that we have

$$\#\varphi_1 = 0,\ \ \#\varphi_2 = 39764,\ \ \text{and}\ \ \#\varphi_3 = 500.$$

We then find that $\psi(\ulcorner\varphi_1\urcorner)$ is the statement $\exists y\ (y \in \underline{0})$, that $\psi(\ulcorner\varphi_2\urcorner)$ is the statement $\exists y\ (y \in \underline{39764})$, and that $\psi(\ulcorner\varphi_3\urcorner)$ is the statement $\exists y\ (y \in \underline{500})$.

Notice that $\psi(\ulcorner\varphi_1\urcorner)$ is not valid in ZFC$^*$, but the other two statements are valid in ZFC$^*$. Both $\varphi_1$ and $\varphi_2$ are not formulas, but $\varphi_3$ is a formula. In fact, $\varphi_3$ is a valid statement in ZFC$^*$. Thus, we have

$$\models_{\mathrm{ZFC}^*} \varphi_3 \leftrightarrow \psi(\ulcorner\varphi_3\urcorner).$$

In other words, from our metatheoretical perspective, $\varphi_3$ asserts "$\psi$ is true on my name". We can say that $\varphi_3$ is *fixed* under $\psi$. △

In the previous example, it was purely by accident that we found a statement fixed under $\psi$. Surprisingly, there are general situations where fixed statements must always exist. This is a consequence of the following result called Gödel's fixed point lemma. It is so named because for any predicate $\psi$ it claims the existence of a new formula $\sigma$ that is fixed under $\psi$.

**Metalemma 28.2.** *Let $T$ be a theory in some language $\mathscr{L}$ of first order logic, and let $\ulcorner\cdot\urcorner\colon \mathrm{String}_{\mathscr{L}} \to \mathrm{ClTerm}_{\mathscr{L}}$ be a naming. Assuming that the diagonalization function is strongly encoded by $D(x, y)$ relative to $T$, then for each predicate $\psi$ there exists a statement $\sigma$ such that*

$$\models_T \sigma \leftrightarrow \psi(\ulcorner\sigma\urcorner).$$

We think of $\sigma$ as (being equivalent to) the sentence "$\psi$ is true when applied to this sentence's name". Thus, the fixed point lemma gives us a way to view sentences as making assertions about themselves.

*Proof.* Let $\psi(x)$ be any fixed predicate. (For instance, $\psi(x)$ might be the assertion "$x$ is odd".) Consider the new predicate $\theta(x)$ given by the formula $\exists y\,(D(x, y) \wedge \psi(y))$. (Thinking, from a metatheoretical perspective, of $D(x, y)$ as asserting "if $x$ is the name of a predicate, then $y$ is the name of its diagonalization", then $\theta(x)$ is asserting "if $x$ is the name of a predicate, then the name of its diagonalization satisfies $\psi$". Thus, when $\psi(x)$ is the predicate "$x$ is odd", then $\theta(x)$ is the predicate "if $x$ is the name of a predicate, the name of its diagonalization is odd".)

Take $\sigma := \theta(\ulcorner\theta\urcorner)$, the diagonalization of $\theta$. (What is $\sigma$ asserting?)

From how $\theta$ was defined, $\sigma$ is the sentence

$$\exists y\,(D(\ulcorner\theta\urcorner, y) \wedge \psi(y)).$$

From (27.5), we have that $\sigma$ is equivalent (in the theory $T$) to the formula

$$\exists y\,((y = \ulcorner\theta(\ulcorner\theta\urcorner)\urcorner) \wedge \psi(y)),$$

which in turn is equivalent to

$$\psi(\ulcorner\sigma\urcorner).$$

Thus $\models_T \sigma \leftrightarrow \psi(\ulcorner\sigma\urcorner)$. □

## 28.B Formalizing validity

Recall that a statement $\sigma$ is *valid* in a theory $T$ when it evaluates as true in every model of $T$. The statement $\underline{0} \in \underline{17}$ is valid in ZFC$^*$, while the (derived) statement $\omega = \omega + 1$ is not valid.

There is a metamathematical predicate $\mathrm{Valid}_T(x)$ that asserts "$x$ is valid relative to $T$". Thus, $\mathrm{Valid}_{\mathrm{ZFC}^*}(\underline{0} \in \underline{17})$ holds. What would it mean to encode the meta predicate $\mathrm{Valid}_{\mathrm{ZFC}^*}$ as a formal predicate? The following definition does that.

**Definition 28.3.** Let $\mathscr{L}$ be a FOL language, and suppose that

$$\ulcorner\cdot\urcorner\colon \mathrm{String}_{\mathscr{L}} \to \mathrm{ClTerm}_{\mathscr{L}}$$

is a naming. A predicate $V(x)$ of the language $\mathscr{L}$ *strongly encodes validity* on names of statements, relative to the theory $T$, if for any statement $\sigma$ of the language $\mathscr{L}$ it happens that

$$\begin{cases} \models_T V(\ulcorner\sigma\urcorner) & \text{if } \sigma \text{ is a valid statement under } T, \\ \models_T \neg V(\ulcorner\sigma\urcorner) & \text{otherwise.} \end{cases} \tag{28.4}$$

Thus, in any model of $T$, the predicate $V$ should accurately pick out the names of the valid statements and reject the names of the statements that are not valid. (We do not particularly care what $V$ says about objects that are not names of statements.) Note that we are modifying Definitions 25.3 and 27.4 to make this definition.

**Example 28.5.** Consider the extended language of set theory $\mathscr{L}^*$ under the theory ZFC$^*$. Instead of using the usual naming, consider the new (potentially noncomputable) naming given by the rule

$$\varphi \mapsto \begin{cases} \underline{0} & \text{if } \varphi \text{ is a valid statement,} \\ \underline{1} & \text{otherwise.} \end{cases}$$

The predicate $V(x)$ given by the formula $x = \underline{0}$ picks out exactly the names of the valid statements, so validity is strongly encoded in ZFC$^*$ using this new naming. △

Is validity strongly encoded using the usual naming of $\mathscr{L}^*$? The following very strong metatheorem of Tarski answers that question, among others.

**Metatheorem 28.6.** *Let $T$ be a theory in some language $\mathscr{L}$ of first order logic, and let $\ulcorner\cdot\urcorner\colon \mathrm{String}_{\mathscr{L}} \to \mathrm{ClTerm}_{\mathscr{L}}$ be any naming. At least one of the following three conditions fails:*

(1) *The diagonalization function is strongly encoded, relative to $T$.*
(2) *Validity is strongly encoded, relative to $T$.*
(3) *There is some model of $T$.*

*Proof.* Assume that validity is strongly encoded (relative to $T$) by the predicate $V(x)$. Also assume that diagonalization is strongly encoded. Applying Gödel's fixed point lemma to the predicate $\neg V(x)$, there exists a statement $\sigma$ such that

$$\models_T \sigma \leftrightarrow \neg V(\ulcorner\sigma\urcorner). \tag{28.7}$$

**Case 1**: Suppose $\sigma$ is valid. Since $V$ strongly encodes validity, we then have $\models_T V(\ulcorner\sigma\urcorner)$. Thus, by (28.7), we have $\models_T \neg\sigma$. But $\sigma$ is valid, so $\models_T \sigma$. There can be no model where both $\sigma$ and $\neg\sigma$ hold.

**Case 2**: Suppose $\sigma$ is not valid. A similar argument as in Case 1 works here. □

The formula $\sigma \leftrightarrow \neg V(\ulcorner\sigma\urcorner)$ in (28.7) is a formalized version of the self-referential statement "This sentence is not valid". Thus $\sigma$ encodes a liar paradox.

Tarski's theorem implies that if formal set theory has at least one model, and if we use the naming given in Example 27.1, then there is no formula that is interpreted as the validity predicate in every model. Tarski's theorem says nothing about the possibility that there is a formula that is interpreted as the validity predicate in some of the models.

Another important consequence of Tarski's theorem is that by Metatheorem 25.8 the collection of valid formulas is not decidable. The valid formulas are semi-decidable, so this is our second example of a semi-decidable set that is not decidable.

**Remark 28.8.** Assume that ZFC$^*$ has at least one model. From what we said above, then there is no predicate that strongly encodes validity relative to ZFC$^*$, but we could do the following. First, extend the language even further by adding a new predicate symbol $V$. Call that new language $\mathscr{L}^{**}$. Next, adjoin to ZFC$^*$ the new axiom schema $V(\ulcorner\varphi\urcorner)$ where $\varphi$ runs over the valid statements of the theory ZFC$^*$, and then also adjoin the schema $\neg V(\ulcorner\varphi\urcorner)$ where $\varphi$ runs over the invalid statements. Call the extended collection of axioms ZFC$^{**}$. Note that there is no algorithm that can decide whether or not a statement belongs to ZFC$^{**}$.

Now, relative to the stronger (undecidable) theory ZFC$^{**}$, the predicate $V$ strongly encodes the valid formulas of in the weaker theory ZFC$^*$ in the smaller language $\mathscr{L}^*$. To create a validity predicate in the larger language $\mathscr{L}^{**}$, under the stronger theory ZFC$^{**}$, it appears that we must once again extend both the language and the theory. ▲

## 28.C Exercises

**Exercise 28.1.** In the proof of Metalemma 28.2, what does the constructed sentence $\sigma$ assert, if $\psi(x)$ is the predicate "$x$ is odd"?

**Exercise 28.2.** Modify the naming in Example 28.5 so that validity is strongly encoded and the naming is injective.

**Exercise 28.3.** Give the details for Case 2 in the proof of Metatheorem 28.6.

**Exercise 28.4.** Modify Definition 28.3 to give a definition of what it means to *weakly encode validity* (on names of statements).

**Exercise 28.5.** In this exercise we will prove a metatheorem, called Gödel's first incompleteness theorem, that applies to a wide array of theories.

Let $\mathscr{L}$ be any first order language, let $\ulcorner\cdot\urcorner\colon \mathrm{String}_{\mathscr{L}} \to \mathrm{ClTerm}_{\mathscr{L}}$ be any naming, and let $T$ be a theory in the language $\mathscr{L}$. Show that at least one of the following four conditions must fail.

(1) $T$ is consistent.
(2) Diagonalization is strongly encoded by a class function relative to $T$.
(3) Validity is weakly encoded relative to $T$.
(4) $T$ is syntactically complete.

(Hint: Show that (3) and (4) together imply that validity is strongly encoded.)

Note: The language of set theory, or even the extension where we add the new constant symbols $\underline{n}$ for each $n \in \mathbb{N}$, is a very simple language. With the axioms we have chosen for formal set theory, one can algorithmically search for proofs of statements. Further, any computable algorithm can be encoded into set theory. Thus, validity is weakly encodable. As we mentioned earlier, the diagonalization function is strongly encoded (being primitive recursive). Thus either our formal set theory is inconsistent or syntactically incomplete.

**Exercise 28.6.** Let $T$ be any first order theory and let $\sigma$ be an arbitrary statement. Prove that $T \cup \{\sigma\}$ has no models if and only if $\models_T \neg\sigma$.

**Exercise 28.7.** Let $\mathscr{L}^*$ be the extended language of set theory. We can list the statements as $\sigma_0, \sigma_1, \ldots$, and we can even make this enumeration computable if we so choose.

Set $T_0 = \text{ZFC}^*$, and for $n > 0$ recursively take

$$T_n = \begin{cases} T_{n-1} \cup \{\sigma_{n-1}\} & \text{if this theory has a model,} \\ T_{n-1} \cup \{\neg\sigma_{n-1}\} & \text{otherwise.} \end{cases}$$

(1) Prove that $T_\omega = \bigcup_{n \in \mathbb{N}} T_n$ has a model, assuming ZFC$^*$ does.
(2) Prove that $T_\omega$ is syntactically complete; each statement is either provable or refutable.
(3) Explain how this doesn't contradict Gödel's first incompleteness theorem. In other words, which of the four conditions in Exercise 28.5 fails for the theory $T_\omega$ (assuming consistency), and why?

**Exercise 28.8.** Let $T \supseteq \text{ZFC}^*$ be a theory in the extended language of set theory. Prove that either $T$ is inconsistent, syntactically incomplete, or undecidable (i.e., there is no algorithm for deciding whether or not a statement belongs to $T$).

Assuming that ZFC$^*$ is consistent, give examples demonstrating that any one of those three conditions can hold while the other two conditions fail.

**Exercise 28.9.** In this exercise we will give a definition of truth and also show, paradoxically, that truth may be undefinable. Let $T$ be any first order theory in some language $\mathscr{L}$, and suppose we have a naming $\ulcorner\cdot\urcorner$ of formulas. Also let $\mathcal{M}$ be a model of $\mathscr{L}$ satisfying $T$.

A predicate $\text{Tr}(x)$ in the language $\mathscr{L}$ is called a *truth predicate* for $\mathcal{M}$ if for each statement $\sigma$ we have

$$\mathcal{M} \models \sigma \leftrightarrow \text{Tr}(\ulcorner\sigma\urcorner).$$

Prove that if $T$ is a consistent theory and the diagonalization function is strongly encoded, relative to $T$, then there is no truth predicate for $\mathcal{M}$. (Was the consistency assumption necessary?)

# Chapter V

# Number Theory

*Life is not measured by the number of breaths we take, but by the moments that take our breath away.* Anonymous

What is a number? There are many possible responses. Here are some options:

- They are the things used to count objects, such as marks on a page.
- They are the elements of the smallest infinite well-ordered set, up to order isomorphism.
- Do you mean a formal number, or a meta number?

In this chapter we will study many different formalizations of the natural numbers, from many different viewpoints.

On the one hand, we would want to give as simple a description of the natural numbers as possible. On the other hand, we want the description to uniquely describe the natural numbers. We will see that these two principles conflict in fundamental ways. Even so, using very little information it is often possible to derive a large amount of information about the natural numbers. Induction plays an essential role in the study of natural numbers, but we will see that there are different versions (and hence consequences) of this seemingly fundamental principle, depending on the formalization one uses.

# 29 Robinson arithmetic

## 29.A Identifying the natural numbers

Consider the following imaginary scenario. A young child is taught by her older brother, at a very young age, about the natural numbers, and she quickly picks up the idea. However, her brother writes the number 4 as (what looks to her) P. She proudly writes out the first few numbers as

$$0, 1, 2, 3, \mathrm{P}, 5, \ldots, 12, 13, 1\mathrm{P}, 15, \ldots, 39, \mathrm{P}0$$

and shows her mother. Her mother realizes what has happened, congratulates her on the nice list of numbers, but teaches her to replace each P by the new symbol 4.

Less fictitiously, numbers have been written using many different notations down through the ages. Base 10 notation has enjoyed recent popularity (in geological terms), but other bases were used in the past, and in the present. You may have been taught Roman numeral notation, which does not use modern base notation. To talk about some large numbers we sometimes find it convenient to avoid all of these notations. For instance, try writing $3^{1938230283}$ in Roman numerals, or even in decimal notation.

The symbols we use for numbers are immaterial. Moreover, it is immaterial if there is a physical symbol for the number. If this universe happens to be finite, and we simply cannot write down all numbers up to $10^{100}$ (for sake of argument), then we will still take the philosophical position that such a number as $10^{100}$ exists (abstractly).

So, can we think of the set $\{p, t, i, \ldots\}$ as the natural numbers? Before we answer this question, we start with an easier question. Can the set $\{p, t, i\}$ be thought of as the natural numbers? The answer here is obviously no. The set is simply too small. Similarly, we do not want $\mathbb{R}$ to be reinterpreted as the natural numbers, because it is too big. Only countable sets should have any hope of acting as a version of the natural numbers.

Conversely, given any set $M$ and a bijection $f\colon M \to \mathbb{N}$, the function $f$ acts as a translation device. It tells us how to rename each element of $M$ as a natural number. Hence, as an abstract, completed collection, the natural numbers are merely *any countable collection.*

However, this point of view is not only ahistorical, but some might call it misleading. Honestly answer, for yourself, the question of whether or not the picture in your head of the natural numbers is some completed, abstract object, possessing a certain infinite "size" and no further structure. Such a picture is quite foreign to most humans. Rather, children learn numbers not by measuring infinite sizes, but by moving from one number to the next, using small examples.

The numbers come equipped with a successor operation. After that, children are taught about addition, multiplication, and other functions on the numbers. They are also taught to distinguish finite sizes and compare numbers with each other. We will now investigate consequences of assuming more and more structure on $\mathbb{N}$.

## 29.B Successor operation

Let $\mathscr{L}$ be a first order language whose signature consists of a single unary function symbol $S$, which we will call the "successor" function.

Without any axioms, there are models of this structure that look nothing like the natural numbers. Even the cardinality can be wrong. For instance, consider the model $\mathcal{M}$ with domain $M = \{a\}$, where we take $S_{\mathcal{M}}$ to be the identity function $M \to M$. However, one model is so important it has a special name.

**Definition 29.1.** The *standard model* of the signature $(S)$ is the model $\mathcal{M}$ whose domain is $M = \mathbb{N}$, and where we let $S_{\mathcal{M}}$ be the usual successor function on the domain $\mathbb{N}$.

So, what sorts of rules must the standard successor operation satisfy? First, and most importantly, there is a special number, 0, that acts differently than all of the other numbers with respect to successors. It is the point where the numbers start.

Thus, in the language $\mathscr{L}$ we might posit an axiom asserting that zero exists, in shorthand form, as follows:

$$\exists! z\, \forall x\, S(x) \neq z.$$

Thus, we can think of 0 as a derived concept. However, most formalizations of arithmetic instead assume that the signature for $\mathscr{L}$ contains a constant symbol $\mathbf{0}$. (In the standard model, we should have $\mathbf{0}_{\mathcal{M}} = 0 \in \mathbb{N}$. If we add other symbols to the signature, we will similarly interpret them in the "standard" way in the standard model.) Furthermore, such theories of arithmetic usually contain the two axioms

$$S(x) \neq \mathbf{0}$$

and

$$(x = \mathbf{0}) \vee \exists y\, (x = S(y)).$$

Note that in both axioms one can think of the variable $x$ as implicitly bound by a universal quantifier, via universal generalization. If you prefer that the free variables in your axioms are universally quantified, feel free to make that change.

There is another fundamental property that the successor operation in the standard model satisfies; it is injective. Thus we assume

$$(S(x) = S(y)) \to (x = y).$$

These axioms seem to capture the idea of natural numbers, as taught to children. Start with the special number 0 and using the successor function move from one number to a new number. However, before we pat ourselves on the back, we should consider whether or not these axioms uniquely determine the natural numbers (up to a renaming that respects the signature). The answer is no.

**Example 29.2.** We construct a (nonstandard) model $\mathcal{M}$ for these axioms. Consider the totally ordered set $\mathbb{N} + \mathbb{Z}$, whose underlying set is $M = (\{1\} \times \mathbb{N}) \cup (\{2\} \times \mathbb{Z})$.

We may define a unary function $S_{\mathcal{M}}\colon M \to M$ by the rule that $(i,n) \mapsto (i,n+1)$. This is an injective function, and $(1,0)$ is the unique element of $M$ not in the image of $S_{\mathcal{M}}$. In other words, $\mathbf{0}_{\mathcal{M}} = (1,0)$.

Next, we need to show that there is no renaming of $M$ that makes it look like the natural numbers with the usual successor function. It should be obvious that $\mathbb{N}$ and $\mathbb{N}+\mathbb{Z}$ are not order isomorphic. However, our signature has no order symbol, so we must generalize the notion of an order isomorphism.

Assume, by way of contradiction, that there exists some bijection $f\colon M \to \mathbb{N}$ such that $f(1,0) = 0$ and $f(S_{\mathcal{M}}(i,n)) = S_{\mathbb{N}}(f(i,n))$. In other words, the map $f$ respects the signature. This is an example of what is called a *model isomorphism*.

Now, since $\emptyset \neq f(\{2\}\times\mathbb{Z}) \subseteq \mathbb{N}$, and $\mathbb{N}$ is well-ordered under its standard order, there is a unique minimal element in $f(\{2\}\times\mathbb{Z})$, call it $m = f(2,n)$ for some $n \in \mathbb{Z}$. As $f$ is injective, we know $m \neq 0 = f(1,0)$. Thus

$$S_{\mathbb{N}}(m-1) = m = f(2,n) = f(S_{\mathcal{M}}(2,n-1)) = S_{\mathbb{N}}(f(2,n-1)).$$

Injectivity of the successor function implies that $m-1 = f(2,n-1)$, contradicting the minimality of $m$. △

Example 29.2 reveals that we are missing a fundamental fact about the successor relation. We need to guarantee that every element is, after a finite number of iterations of the successor operation, a successor of 0. In symbols,

$$(x = \mathbf{0}) \vee (x = S(\mathbf{0})) \vee (x = S(S(\mathbf{0}))) \vee (x = S(S(S(\mathbf{0})))) \vee \cdots . \tag{29.3}$$

However, this is not a well-formed formula of first order logic. More disastrously, even if we add as axioms all of statements in the signature $(S, \mathbf{0})$ that are satisfied by the standard model, we cannot rule out nonstandard models; see Exercise 29.5.

There are two standard solutions to this problem. First, one can pass from a first order language to a stronger language, capable of expressing concepts like (29.3). Second, one can work in a more complicated first order theory, such as ZFC, capable of talking about how its formal version of the natural numbers is well-ordered. These two solutions are, in some sense, not too different.

**Advice 29.4.** When needed, work in a stronger setting.

There are downsides to these solutions. One problem is that languages stronger than first order ones are difficult to pin down, formally describe, or prove meta results for. Also, complicated theories, like ZFC, may have nonisomorphic models, even if the formalized natural numbers can be proved (inside the theory) to be unique; so we have just traded one type of nonuniqueness for another.

Hereafter, we will not overly focus on the nonuniqueness of our formal natural numbers. We will, instead, attempt to describe just what one can (and cannot) prove given a modest axiomatization. We will see that, just as for set theory, a large portion of arithmetic can be developed using only a few axioms and general principles. Towards that end, you might begin by trying to answer the following questions:

**Question 29.5.** Are there any first order statements about the successor operation that are true in the standard model, but they are not consequences of the axioms described previously in this section? ▲

**Question 29.6.** Does every model of the axioms of arithmetic (that we described previously) contain an isomorphic copy of the standard natural numbers? ▲

## 29.C Addition and multiplication

Addition is just repeated succession, and so we can define it by the recursive rules $m + 0 = m$ and $m + S(n) = S(m + n)$, for meta numbers $m, n \in \mathbb{N}$. Similarly, multiplication is repeated addition, and can be defined recursively.

Throughout this subsection, let $\mathscr{L}$ be the *language of arithmetic*, whose signature is $(S, \mathbf{0}, +, \cdot)$. The theory consisting of the seven axioms

(1) $\neg(S(x) = \mathbf{0})$,
(2) $(x = \mathbf{0}) \vee \exists y\, (x = S(y))$,
(3) $(S(x) = S(y)) \to (x = y)$,
(4) $x + \mathbf{0} = x$,
(5) $x + S(y) = S(x + y)$,
(6) $x \cdot \mathbf{0} = \mathbf{0}$,
(7) $x \cdot S(y) = (x \cdot y) + x$

is called *Robinson arithmetic*, and is usually denoted Q. As weak as this theory appears, it was developed by Raphael Robinson for the express purpose of encoding *all* computable functions.

To begin, we give names to each of the meta natural numbers, in the language of arithmetic, using the recursion

$$0 \mapsto \underline{0} = \mathbf{0} \text{ and } n + 1 \mapsto \underline{n+1} := \underbrace{S \circ S \circ \cdots \circ S}_{n+1 \text{ times}}(\mathbf{0}) = S(\underline{n}).$$

This underline map is injective to the closed terms. By axioms (1) and (3) of Q, the images of the meta numbers in any model of Q are distinct.

Now we define what it means to encode functions.

**Definition 29.7.** A (possibly partial) meta function $f\colon \mathbb{N}^k \to \mathbb{N}$ is *weakly encoded* by a formula $\varphi(x_1, \ldots, x_k, y)$, relative to some theory $T$ in the language of arithmetic, when given any $(m_1, \ldots, m_k, n) \in \mathbb{N}^{k+1}$ we have

$$\begin{cases} \models_T \varphi(\underline{m_1}, \ldots, \underline{m_k}, \underline{n}) & \text{if } f(m_1, \ldots, m_k) = n, \text{ and} \\ \models_T \neg\varphi(\underline{m_1}, \ldots, \underline{m_k}, \underline{n}) \text{ or } \not\models_T \varphi(\underline{m_1}, \ldots, \underline{m_k}, \underline{n}) & \text{if } f(m_1, \ldots, m_k) \neq n. \end{cases}$$

It is *strongly encoded* when the condition "or $\not\models_T \varphi(\underline{m_1}, \ldots, \underline{m_k}, \underline{n})$" can be removed from the second case. Moreover, if

$$\models_T \exists! y\, \varphi(x_1, \ldots, x_k, y),$$

then we say that $\varphi$ is a *class function* weakly (or strongly) encoding $f$.

Note that if $T$ is a consistent theory, then the definition of a weak encoding can be simplified; the first "if" can be improved to an "if and only if" statement, and the second case can be removed entirely. Many authors use that simplified definition, implicitly assuming that the theory under study is consistent.

**Warning 29.8.** We have used the word "encoding" in this book, but other authors use "define" or "represent" or a host of other words. They may also use "strong" and "weak" in slightly different ways.

**Example 29.9.** In this example we prove that the *meta* successor function is strongly encoded by the formula $S(x) = y$, relative to Q.

First, we show that the formula $S(x) = y$ is a class function. Given any $x$, the existence of a satisfying output $y$ is obvious, since we may take $y$ to be the term $S(x)$. Uniqueness is also clear; there is no choice for the value of $y$ other than $S(x)$.

Second, our definition of the underline naming map guarantees that

$$\underline{m+1} = S(\underline{m}),$$

and thus the first case in the definition of a strong encoding holds. On the other hand, if $(m, n) \in \mathbb{N}^2$ with $n \neq m + 1$, then axioms (1) and (3) together with a meta induction yield $\models_{\mathrm{Q}} \underline{n} \neq \underline{m+1}$; see Exercise 29.4. Hence, $\models_{\mathrm{Q}} \neg(S(\underline{m}) = \underline{n})$. △

Raphael Robinson proved in 1950 that all computable functions are strongly encoded relative to Q. Later, in 1969, Robert Ritchie and Paul Young proved that all computable *partial* functions are weakly encoded (by class functions) in Q. We direct the interested reader to more specialized textbooks, or those original papers, for proofs of these claims.

By Gödel's incompleteness theorem, Exercise 28.5, we have the following consequence. Any theory of arithmetic containing Q that is consistent and decidable (i.e., there is an algorithm telling us which statements belong to the theory) must be syntactically incomplete. In particular, there is no way to decide which statements are true or false in the standard model.

## 29.D Order relation

The order relation $<$ on the natural numbers is a fundamental component of its structure. It can be developed as a derived concept in the language of arithmetic, by defining $x < y$ to mean $\exists z\, (x + S(z) = y)$.

Robinson arithmetic is so weak that it does not prove much about this relation, including the irreflexive condition. (Can you find a model of Q where the derived order relation is not irreflexive?) It also is incapable of proving seemingly simple facts about addition and multiplication. For instance, we cannot prove that addition is commutative, or even the easier claim that $\mathbf{0} + x = x$. For these we will need induction, which we discuss in the next section.

It is possible to replace the previously defined derived order with a better derived order, by using a more complicated formula. However, the purpose of Q is to be as minimal as possible, and so there is little point in pursuing such improvements here. In the next section we will work in a theory where this derived order relation acts very nicely, satisfying all of the usual properties.

## 29.E Exercises

**Exercise 29.1.** Write an unabbreviated formula that asserts zero exists, in the signature with one unary function symbol $S$.

**Exercise 29.2.** Give a formal derivation that $\underline{2} + \underline{2} = \underline{4}$, using only the axioms of Q and the definitions $\underline{2} := S(S(\underline{0}))$ and $\underline{4} := S(S(S(S(\underline{0}))))$.

**Exercise 29.3.** Find a model of Robinson arithmetic that is nonstandard. Try to make it as simple as possible. If possible, give explicit, recursive definitions for the model's addition and multiplication functions.

**Exercise 29.4.** Show $\vdash_{\mathrm{Q}} \underline{m} \neq \underline{n}$ for any elements $m, n \in \mathbb{N}$ with $m > n$. (Hint: Use meta induction.)

**Exercise 29.5.** Let $\mathscr{L}$ be a first order language whose signature, $(S, \mathbf{0})$, consists of a unary function symbol and a constant symbol. Let $T$ be the collection of all statements in this language that are satisfied in the standard (natural number) model. Let $\mathscr{L}'$ be the language we obtain from $\mathscr{L}$ by adjoining to the signature another constant symbol $c$. For each meta number $n$, let $\varphi_n$ be the statement $\underbrace{S \circ S \circ \cdots \circ S}_{n \text{ times}}(\mathbf{0}) \neq c$.

(1) Prove that any finite subcollection of $T' := T \cup \{\varphi_n : n \in \mathbb{N}\}$ is satisfiable.
(2) Use the compactness theorem to conclude that $T'$ is satisfiable, hence showing that there exists a nonstandard model for $T$.

**Exercise 29.6.** Let $\varphi(x, y)$ and $\psi(x, y)$ be class functions strongly encoding (total, meta) functions $f, g \colon \mathbb{N} \to \mathbb{N}$. Find class functions strongly encoding $f \dot{-} g$ and $f \circ g$, relative to $Q$. (Which axioms of $Q$ were needed?) If $f$ is surjective, find a class function strongly encoding $f^{[-1]}$. (We will handle $f^{\square}$ in the next section.)

**Exercise 29.7.** Prove that if a class function weakly encodes a total function (relative to some theory $T$), then it strongly encodes that function.

**Exercise 29.8.** Let $\varphi(x, y)$ be the formula $\mathbf{0} + x = y$. Note that $\vdash_{\mathrm{FOL}} \forall x\, \exists! y\, \varphi(x, y)$; the unique value of $y$ is $\mathbf{0} + x$. Prove the metatheorem that if $n \in \mathbb{N}$ then $\vdash_Q \varphi(\underline{n}, \underline{n})$. (Hint: You may use induction over $n \in \mathbb{N}$, in the metatheory.) In other words, $\varphi$ is a strong encoding of the meta identity function, relative to the theory $Q$. How does this reconcile with the fact that $\mathbf{0} + x = x$ is not provable in $Q$?

# 30 Peano arithmetic

Throughout this section, we let $\mathscr{L}$ be the language of arithmetic, whose signature is $(S, \mathbf{0}, +, \cdot)$. We also let $\mathscr{L}_0$ be the sublanguage with signature $(S, \mathbf{0})$. In the previous section, we listed seven formulas in $\mathscr{L}$ (three of which come from $\mathscr{L}_0$) that are true when interpreted in the natural numbers, giving us Robinson arithmetic. Robinson arithmetic is very weak, and is missing a lot of statements that are true when interpreted in $\mathbb{N}$. In this section we add new axioms to our list, creating Peano arithmetic. We also (re)prove Gödel's incompleteness theorem, which tells us that any decidable list of axioms will be missing formulas satisfied in $\mathbb{N}$.

## 30.A Differences in languages

Recall that the first three axioms of Robinson arithmetic are in the language $\mathscr{L}_0$. Are there any other formulas in this language that are true when interpreted in $\mathbb{N}$, but are not consequences of the three axioms? In other words, we might ask whether those three axioms are enough to imply all other true statements about the successor operation on the meta numbers. The answer is no. Indeed, Robinson arithmetic is missing some very basic facts about the meta numbers.

**Metatheorem 30.1.** *The formula $\forall x\, \neg(S(x) = x)$ is true when interpreted in the natural numbers, but is not implied by Robinson arithmetic.*

*Proof.* The first part is clear. To show the second part, we will construct a model, $\mathcal{M}$, of Robinson arithmetic that does not satisfy the new formula. The domain of our model is $M = \mathbb{N} \cup \{\infty\}$, where $\infty$ is just a new symbol. We take $\mathbf{0}_{\mathcal{M}} = 0 \in \mathbb{N}$. Our successor function $S_{\mathcal{M}}$ when restricted to $\mathbb{N}$ is the standard successor operation, and we also take $S_{\mathcal{M}}(\infty) = \infty$. Addition and multiplication will also be defined in the usual way on $\mathbb{N}$, and we posit the additional rules

$$\infty + m = m + \infty = \infty \cdot m = m \cdot \infty = \infty$$

for each $m \in M$, except that $\infty \cdot 0 = 0$. One may check that the seven axioms of Q hold, as well as check that the new formula is not satisfied in this model. □

It seems reasonable that a "good" theory of arithmetic should guarantee that $S(x) \neq x$. More generally, for each meta number $n \geq 1$, let $\varphi_n$ be the formula

$$\forall x\, \neg(\underbrace{S \circ S \circ \cdots \circ S}_{n \text{ times}}(x) = x). \tag{30.2}$$

This is a schema of statements that are true when interpreted in the standard model. Surprisingly, we now have a *complete* theory of the (meta) successor operation; any statement in the language $\mathcal{L}_0$ that is satisfied in the standard model is a consequence of (30.2) together with the first three axioms of Robinson arithmetic. (We will not prove this in this book, but it follows from standard techniques in model theory.)

We might hope that, similarly, after adding enough axioms in the stronger language $\mathscr{L}$, we would obtain the complete theory of arithmetic. However, as strange as it may seem, interpretations of the languages $\mathscr{L}$ and $\mathscr{L}_0$ behave very differently. The complete theory of the successor operation can easily be written down by a computer, but not the complete theory of arithmetic.

**Metatheorem 30.3.** *There is no way to recursively enumerate a list of axioms, from the language of arithmetic, whose consequences consist of the formulas satisfied in the standard model.*

*Proof sketch.* In Subsection 22.D, we constructed a recursive bijection from the strings of symbols in the language of set theory to the meta natural numbers $\mathbb{N}$. The same can easily be done for the language of arithmetic. Using similar notation, we will write such a bijection as $\#\colon \mathrm{String}_{\mathscr{L}} \to \mathbb{N}$. In the previous section we defined a recursive (underline) function $\mathbb{N} \to \mathrm{ClTerm}_{\mathscr{L}}$, by the rules $0 \mapsto \underline{0} = \mathbf{0}$ and $n+1 \mapsto S(\underline{n})$. The composition of these two maps is a naming, $\ulcorner\cdot\urcorner\colon \mathrm{String}_{\mathscr{L}} \to \mathrm{ClTerm}_{\mathscr{L}}$, that can be implemented on a computer. Thus, the diagonalization function is computable.

Relative to these encodings, let $T$ be any recursively enumerable list of axioms satisfied in $\mathbb{N}$. We can add the seven axioms of $Q$ to the beginning of this list and it is still recursively enumerable. So, without loss of generality, we will assume $T$ contains $Q$. Since we take it as granted that the standard model exists, $T$ is consistent.

Diagonalization is strongly encoded relative to $Q$, and hence in the stronger theory $T$, since $Q$ is powerful enough to strongly encode all computable functions on the natural numbers and we may computably encode formulas as natural numbers. Validity is weakly encoded since $T$ is recursively enumerable. By Exercise 28.5 there exists a statement that is neither provable nor disprovable from $T$. □

Metatheorem 30.3 is commonly referred to as Gödel's first incompleteness theorem (for arithmetic). A large part of the nonsketched proof consists of carefully showing that all computable functions are strongly encodable, or at least those computable functions necessary for the proof. Gödel's original proof took place in Peano arithmetic, which is a much stronger theory than $Q$.

There are multiple ways to attempt to generalize this theorem. For instance, one can try to work in a more complicated language $\mathscr{L}' \supseteq \mathscr{L}$, perhaps involving more functions on the natural numbers such as exponentiation. As long as the new language has a model where $Q$ is satisfied, and there is a computable way to list the axioms being used, then the same proof shows that the theory will be syntactically incomplete. (Some languages have too many formulas for computers to be able to list them all. For instance, add uncountably many constants to $\mathscr{L}$.)

There is a simple way to *noncomputably* describe the full theory of arithmetic; just take $T$ to be the full collection of statements that are satisfied in the standard model. Metatheorem 30.3 implies that no computer algorithm can print the elements of $T$ (and only the elements of $T$), but $T$ does exist (metatheoretically).

## 30.B Many truths from one concept

Before Gödel proved his first incompleteness theorem in 1931, it was hoped that there might be an easy way to formalize all of arithmetic using only a few axioms. Towards that goal, Presburger showed in 1929 that the complete theory of arithmetic *without multiplication* can be computably described; it can be done using the first five axioms of $Q$ and one new axiom schema. He also gave an algorithm for correctly deciding whether a formula in the signature $(S, \mathbf{0}, +)$ is satisfied in the standard model.

The axioms of Presburger arithmetic are the first five axioms of $Q$ together with an axiom schema that is a first order version of induction. Formally, for each formula $\varphi(x)$ in the signature $(S, \mathbf{0}, +)$, then

$$\Big(\varphi(0) \wedge \forall x \left(\varphi(x) \to \varphi(S(x))\right)\Big) \to \forall x\, \varphi(x) \tag{30.4}$$

is an axiom.

In a similar way, one can define Peano arithmetic, or PA for short. This is the theory consisting of the seven axioms of $Q$ together with the induction schema (30.4), except that the formulas $\varphi(x)$ range over the full language of arithmetic $\mathscr{L}$, and can include multiplication. Peano arithmetic is more expressive than Presburger arithmetic, as it allows inductive reasoning about multiplication. It is also much stronger than Robinson arithmetic. For instance, one can prove in PA, by induction, that addition is commutative and associative. Gödel's incompleteness theorem implies that there are statements in the language of arithmetic whose truth (as interpreted in the standard model) cannot be decided using only PA; however, for many "natural" statements you are likely to encounter, PA is enough to prove or disprove them. For an example of a natural statement that cannot be proved in PA, see Exercise 30.6.

Restricting the types of formulas $\varphi(x)$ allowed in the induction schema weakens the schema. When one studies the axioms needed to prove theorems, this is called *reverse mathematics*, as it goes in the reverse direction from theorems to axioms. Weakened versions of induction, recursion, and the comprehension schemas are a common sight in reverse mathematics, and moreover these weakened axiom schemas are sometimes equivalent (inside some very weak, usually second order, general theory) to well known theorems.

In the other direction, if one extends the language $\mathscr{L}$, and one lets $\varphi(x)$ run over formulas of the extended language in the induction schema, then one can potentially go beyond PA. Even more generally, consider the second order version of induction, which says that for any subset $A \subseteq \mathbb{N}$, then

$$\Big(0 \in A \wedge \forall n \in \mathbb{N} \left(n \in A \to n + 1 \in A\right)\Big) \to \forall n \in \mathbb{N}\, (n \in A). \tag{30.5}$$

It defines $\mathbb{N}$ as the (unique) smallest inductive set. If $\varphi(x)$ is any first order formula, then taking $A = \{n \in \mathbb{N} : \varphi(n)\}$ in (30.5) will give (30.4). However, second order induction does not presuppose that the set $A$ must be defined by a first order formula—it allows $A$ to be an arbitrary subset of $\mathbb{N}$. We can interpret second order statements like (30.5) in ZFC, and indeed ZFC proves more statements about arithmetic than PA can.

## 30.C The upward Löwenheim-Skolem theorem

There is a very surprising fact about first order theories. They are notoriously bad at limiting cardinalities. In particular, PA will have models with very big cardinality. This is a consequence of the upward Löwenheim-Skolem theorem.

**Theorem 30.6.** *Let $T$ be any theory, in some first order language, with an infinite model. Then $T$ has models of arbitrarily large cardinality.*

*Proof.* Let $\mathcal{M}$ be some infinite model of $T$, with domain $M$. Fix some cardinal $\kappa > |M|$, and adjoin to the language new constant symbols $c_\alpha$ for ordinals $\alpha < \kappa$.

For each pair $\alpha < \beta < \kappa$, let $\varphi_{\alpha,\beta}$ be the statement $c_\alpha \neq c_\beta$, and let $T'$ be the collection of all these statements. Note that $\mathcal{M}$ can model any finite subset $X \subseteq T'$, since $M$ is infinite; just interpret the finitely many new constants that appear in $X$ as distinct elements of $M$, and all of the other new constants as any fixed element of $M$. Thus $T \cup T'$ is consistent, by the compactness theorem. By the completeness theorem, there is a model of $T \cup T'$. Any model must have a domain whose cardinality is at least $\kappa$. □

This reveals one more limitation in the expressive power of first order languages.

## 30.D Exercises

**Exercise 30.1.** Can the successor function be defined, as a derived concept, in the language with signature $(\mathbf{0}, +)$? (Hint: Can you define $\mathbf{1}$?)

**Exercise 30.2.** Justify the claim that the meta numbers satisfy $\forall x\, \neg(S(x) = x)$. (Hint: Prove it in PA.)

**Exercise 30.3.** For each meta number $n \geq 1$, let $\varphi_n$ be as in (30.2). Show that for each $m \geq 1$, the formula $\neg\varphi_m$ is consistent with the first three axioms of Robinson arithmetic together with $\{\varphi_n\}_{m \nmid n}$. Thus, these new axioms are (mostly) independent of one another.

**Exercise 30.4.** Prove that the second axiom of $Q$ is a consequence of the induction schema (30.4). Similarly, for each meta number $n \geq 1$, show that $\varphi_n$ as in (30.2) is a consequence of the induction schema and the first and third axioms of $Q$.

**Exercise 30.5.** Recall the derived relation, $<$, defined by the condition that $x < y$ holds exactly when there exists some $z \neq \mathbf{0}$ with $x + z = y$. Determine whether or not Presburger arithmetic can prove that $<$ satisfies the conditions of a strict total order relation.

**Exercise 30.6.** Let $n \in \mathbb{N}$. For each integer base $b \geq 2$, the number $n$ has a base $b$ representation. This means that

$$n = a_k \cdot b^k + a_{k-1} \cdot b^{k-1} + \cdots + a_1 \cdot b^1 + a_0 \cdot b^0$$

for some $a_0, a_1, \ldots, a_{k-1}, a_k \in \{0, 1, 2, \ldots, b-1\}$, with $k$ large enough. For instance, the base 3 representation for 107 is $1 \cdot 3^4 + 0 \cdot 3^3 + 2 \cdot 3^2 + 2 \cdot 3^1 + 2 \cdot 3^0$. The convention is to drop all summands whose coefficients are 0, to write $1 \cdot b^i$ just as $b^i$, and to write $b^1$ as $b$. Also, the summand $a_0 \cdot b^0$ is usually written as $a_0$. So we write the base 3 representation for 107 as $3^4 + 2 \cdot 3^2 + 2 \cdot 3 + 2$.

The *hereditary* base $b$ expansion of a number $n$ is given by writing all exponents using base $b$ expansions. For instance, the hereditary base 3 representation for 107 is

$$3^{3+1} + 2 \cdot 3^2 + 2 \cdot 3 + 2.$$

The base 2 expansion of 107 is $2^6 + 2^5 + 2^3 + 2 + 1$. The hereditary base 2 expansion of 107 is

$$2^{2^2+2} + 2^{2^2+1} + 2^{2+1} + 2 + 1.$$

In general, the hereditary base $b$ expansion involves only addition, multiplication, exponentiation, and symbols for numbers from 1 to $b$.

Let $S_b \colon \mathbb{N} \to \mathbb{N}$ be the map that takes the hereditary base $b$ representation for $n$, and replaces each $b$ by $b+1$. For instance, $S_3(107) = 4^{4+1} + 2 \cdot 4^2 + 2 \cdot 4 + 2$.

For each number $n \in \mathbb{N}$, the *Goodstein sequence* for $n$ is

$$n, S_2(n) - 1, S_3(S_2(n) - 1) - 1, S_4(S_3(S_2(n) - 1) - 1) - 1, \ldots,$$

stopping if it ever reaches 0.

Describe the Goodstein sequences for $n = 2, 3, 4$. (The sequence starting with 4 should increase for about $3 \cdot 2^{402653209}$ steps, then stay constant for a while, then start decreasing.) Prove, in general, that every Goodstein sequence terminates. (This fact can be stated in the language of arithmetic, but Kirby and Paris showed in 1982 that it cannot be proven in PA. Thus, your proof will require thinking either about sets or about higher order induction.)

**Exercise 30.7.** (Only do this exercise if you know the Chinese Remainder Theorem.) Let $\mathrm{Mod}(x, y) = z$ be shorthand for the formula $(z < y) \wedge \exists w\, (x = wy + z)$. Thus, this represents the open sentence asserting that $z$ is the remainder upon division of $x$ by $y$. PA proves that for any $x$ and $y$, then there is a unique such $z$ whenever $y \neq 0$, and no such $z$ otherwise. (The proof given in "Transition" is, essentially, such a proof.)

Gödel, when encoding computable functions, used a clever trick called *sequence encoding*. His idea was to use the fact that arithmetic sequences are complex enough to encode any finite sequence. To get an idea of the method, consider the 3-term sequence $0, 3, 15$. Now $\mathrm{Mod}(53, 1) = 0$, $\mathrm{Mod}(53, 1{+}9) = 3$, and $\mathrm{Mod}(53, 1{+}2{\cdot}9) = 15$. This is no accident. Prove that for any finite sequence $n_0, n_1, \ldots, n_{k-1}$ there exist some $a, b, c \in \mathbb{N}$ with $\mathrm{Mod}(c, a + ib) = n_i$ (for each $i < k$).

Now, if a function $f$ can be encoded, then we can encode $f^{\square}(x) = y$ by the formula

$$\exists a, b, c \left[ (\mathrm{Mod}(c, a) = 0) \wedge (\mathrm{Mod}(c, a + xb) = y) \wedge \right.$$
$$\left. \forall z \Big( (z < x) \to f(\mathrm{Mod}(c, a + zb)) = \mathrm{Mod}(c, a + S(z)b) \Big) \right].$$

# 31 Building models

## 31.A Models of PA

Let $\mathcal{M}$ be any model of PA (or even just Q), and let $M$ be the domain. We know that $M$ contains an element $m_0 = \mathbf{0}_{\mathcal{M}}$, as well as other elements such as

$$m_1 = S(\mathbf{0})_{\mathcal{M}} = S_{\mathcal{M}}(\mathbf{0}_{\mathcal{M}}) = S_{\mathcal{M}}(m_0).$$

One of the axioms of PA asserts that $\mathbf{0}$ is not a successor, and so since $\mathcal{M}$ satisfies PA, we know $m_0 \neq m_1$. For each $k \in \mathbb{N}$, we can take $m_{k+1} = S_{\mathcal{M}}(m_k)$. These elements are all distinct, and collecting them together we have a subset $M' = \{m_i : i \in \mathbb{N}\} \subseteq M$.

We might ask how addition behaves on $M'$? One of the axioms of PA is that $x + \mathbf{0} = x$. Thus, we must have

$$m_i +_{\mathcal{M}} m_0 = m_i,$$

for each $i \in \mathbb{N}$. This is the base case of an induction, over the meta mathematical numbers $j \in \mathbb{N}$, proving that

$$m_i +_{\mathcal{M}} m_j = m_{i+j}.$$

In other words, addition on $M'$ is performed by adding the subscripts as usual in the standard model $\mathbb{N}$.

One can similarly show that $m_i \cdot_{\mathcal{M}} m_j = m_{i \cdot j}$, for each $i, j \in \mathbb{N}$. Thus, from the model $\mathcal{M} = (M, S_{\mathcal{M}}, \mathbf{0}_{\mathcal{M}}, +_{\mathcal{M}}, \cdot_{\mathcal{M}})$, we can create a new structure

$$\mathcal{M}' = (M', (S_{\mathcal{M}})|_{M'}, \mathbf{0}_{\mathcal{M}}, (+_{\mathcal{M}})|_{M' \times M'}, (\cdot_{\mathcal{M}})|_{M' \times M'}).$$

Up to renaming the elements of $M'$, this new structure is just a copy of the standard model. In particular, $\mathcal{M}'$ still models PA.

The elements of $M'$ are often called the *standard numbers* in $M$, and the elements of $M - M'$ (if any) are called the *nonstandard numbers*. As we saw at the end of Section 30, the cardinality of $M - M'$ can be arbitrarily large.

In this section, we will study to what extent similar considerations hold for other theories, in languages other than arithmetic.

## 31.B Model existence

Let $\mathscr{L}$ be any language of FOL, and let $T$ be a theory in that language (i.e., a collection of axioms). It is quite possible that there are no models of $T$, such as in the following situation.

**Metatheorem 31.1.** *If $T$ is an inconsistent theory, then it has no model.*

*Proof.* By way of contradiction, assume the existence of a model $\mathcal{M}$. Let $\varphi$ be any statement (such as $\forall x\,(x = x)$). Now, as $T$ is inconsistent, we have both $\vdash_T \varphi$ and $\vdash_T \neg\varphi$. By the soundness of FOL, Metatheorem 10.3, we have $\models_{\mathcal{M}} \varphi$ and $\models_{\mathcal{M}} \neg\varphi$. This is impossible, since if a statement is interpreted as true, then its negation is interpreted as false. □

The converse is related to Gödel's completeness theorem, Metatheorem 10.13.

**Metatheorem 31.2.** *Assuming Gödel's completeness theorem, if $T$ is a consistent theory, then it has a model. Vice versa, if every consistent theory has a model, then Gödel's completeness theorem is true.*

*Proof.* Assume Gödel's completeness theorem and assume that $T$ is consistent. Thus, we know $\nvdash_T \exists x\,(x \neq x)$. Hence, by the contrapositive of Metatheorem 10.13, there must be at least one model where $\exists x\,(x \neq x)$ fails to be satisfied.

Now we work vice versa, and assume every consistent theory has a model. Further assuming $\models_T \varphi$ (for some statement $\varphi$), we need to show that $\vdash_T \varphi$. Notice that $T \cup \{\neg\varphi\}$ has no models, since $\varphi$ holds in all models where $T$ is satisfied. Thus, by the contrapositive of our assumption, $T \cup \{\neg\varphi\}$ is inconsistent, meaning $\neg\varphi \vdash_T \bot$. By the deduction theorem, $\vdash_T \neg\varphi \to \bot$. By basic logic, we derive $\vdash_T \varphi$, as desired. □

Previously, we assumed Gödel's completeness theorem without proof. We will now sketch a proof, by showing that every consistent theory has a model—the so called "model existence theorem"—first proved by Henkin in his Ph.D. thesis, without reference to Gödel's proof. Showing some of the details is left to the exercises.

**Step 1**: Extend the theory and the language.

Let $T$ be a consistent theory. After extending $T$, we may as well assume that $T$ is syntactically complete; see Exercise 31.4. Thus, for any statement $\varphi$, either $\varphi \in T$ or $\neg\varphi \in T$. (From work done in previous sections, we know that this cannot generally be done in a decidable manner.)

**Step 2**: Create a potential domain for a model of $T$.

Consider the collection of closed terms $\mathrm{ClTerm}_{\mathscr{L}}$. In the case of PA, these closed terms are just enough to create a model. However, in general a few things could go wrong with attempting to make $\mathrm{ClTerm}_{\mathscr{L}}$ the domain of a model for $T$.

First, there may not be any closed terms, and models cannot be empty. An easy fix to this problem is to add a constant symbol to the language.

Second, the theory $T$ might contain a statement of the form $\exists x\,\psi(x)$, and yet $T$ also contains $\neg\psi(t)$ for each $t \in \mathrm{ClTerm}_{\mathscr{L}}$. We can fix this problem by once again adding new constant symbols $c_\psi$ to the language (one for each such existential formula in $T$), but also adding $\psi(c_\psi)$ to $T$.

Of course, adding new constants to a language means that the theory $T$ may no longer be syntactically complete, thus ruining Step 1. One could imagine a scenario where after adding new constant symbols then the theory needs to be expanded, but after expanding the theory more new constants must be added. Fortunately, this back and forth expansion stabilizes quickly (after no more than $\omega$ steps); this is worked out in Exercise 31.5.

Third, the theory $T$ might contain a formula of the form $t_1 = t_2$ for distinct elements $t_1, t_2 \in \mathrm{ClTerm}_{\mathscr{L}}$. To handle this situation, define an equivalence relation $\sim$ on the closed terms, by saying that $t_1 \sim t_2$ if and only if $(t_1 = t_2) \in T$. (One should prove that this is, indeed, an equivalence relation.) Now, the set of equivalence classes

$$M = \mathrm{ClTerm}_{\mathscr{L}} / \sim$$

can act as a domain for a model of $T$. We will write equivalence classes using "square-bracket" notation.

**Step 3**: Finish constructing a potential model.

We still need to construct interpretations of the symbols from the signature of $\mathscr{L}$. The interpretations of the constant symbols is obvious; take $c_{\mathcal{M}} = [c]$.

For unary function symbols, we would like to define $f_{\mathcal{M}} \colon M \to M$ by the rule $[t] \mapsto [f(t)]$. We need to check that this gives us a (well-defined) function. Thus, suppose that $t, t'$ are two closed terms with $[t] = [t']$. Hence $\vdash_T t = t'$. Applying the function $f$, we have $\vdash_T f(t) = f(t')$. (This step takes a little more work, if one uses Mendelson's axiomatic system.) Therefore $[f(t)] = [f(t')]$, as desired. A similar argument works just as well for $n$-ary function symbols.

Finally, suppose that $R$ is some $n$-ary relation symbol of the language. We take

$$R_{\mathcal{M}} = \{([t_1], \dots, [t_n]) \, : \vdash_T R(t_1, \dots, t_n)\}.$$

It is helpful to prove, conversely, that if $([t_1], \dots, [t_n]) \in R_{\mathcal{M}}$, then $\vdash_T R(t_1, \dots, t_n)$.

**Step 4**: Show that $\mathcal{M}$ models $T$.

Induct over formula complexity. First consider $t_1 = t_2$. Such an equality is provable in $T$ if and only if $[t_1] = [t_2]$ (by the definition of the equivalence relation). Similar considerations apply to the other type of atomic formula (by the work in the previous paragraph). Thus an atomic formula, or its negation, belongs to $T$ if and only if it is modeled in $\mathcal{M}$. Note that any formula is logically equivalent to one where all negations are applied only to atomic formulas.

Next, we handle the other logical connectives. Suppose that $(\varphi \vee \psi) \in T$, for some formulas $\varphi, \psi$ of complexity already handled. Since $T$ is syntactically complete, either $\varphi$ or $\neg\varphi$ belongs to $T$, and similarly either $\psi$ or $\neg\psi$ belongs to $T$. We cannot have both $\neg\varphi$ and $\neg\psi$ belong to $T$, since $T$ is consistent. Thus, without loss of generality, we may assume $\varphi \in T$. By our inductive hypothesis, we have $\models_{\mathcal{M}} \varphi$. Hence $\models_{\mathcal{M}} \varphi \vee \psi$. The case for conjunction is even easier.

Finally, we handle quantifiers. If $\exists x \, \varphi(x)$ belongs to $T$, then there exists some constant $c$ such that $\varphi(c)$ is in $T$. This formula has smaller complexity, and so our inductive hypothesis yields $\models_{\mathcal{M}} \varphi(c)$. Thus $\models_{\mathcal{M}} \exists x\, \varphi(x)$ also holds, from how we defined satisfaction of an existential statement in a model.

Last of all, suppose that $\forall x\, \varphi(x)$ belongs to $T$. Thus $\varphi(t) \in T$ for every closed term $t$. These statements all have smaller complexity, and so $\models_{\mathcal{M}} \varphi(t)$. In other words, $(\varphi(t))_{\mathcal{M}}$ is true, for each closed term $t$. But $(\varphi(t))_{\mathcal{M}} = \varphi_{\mathcal{M}}(t_{\mathcal{M}}) = \varphi_{\mathcal{M}}([t])$, and $[t]$ runs over all the elements of the domain $M$. Thus $\models_{\mathcal{M}} \forall x\, \varphi(x)$.

## 31.C The downward Löwenheim-Skolem theorem

With the model existence theorem in hand, we can now prove the existence of small models. We note that the set of closed terms, from the language used in the previous subsection, had cardinality at most $|\mathscr{L}| + \aleph_0$. We can use this fact to great effect. The following result is called the *downward Löwenheim-Skolem theorem*, because it often allows one to create a model of smaller cardinality.

**Theorem 31.3.** *Let $\mathscr{L}$ be a language, let $T$ be a theory in that language, and let $\mathcal{M}$ be an infinite model of $T$. Then there exists an infinite model of $T$ with cardinality at most $|\mathscr{L}| + \aleph_0$.*

*Proof.* Without loss of generality, we may extend the language to include countably many new constant symbols $c_0, c_1, \ldots$, and add to $T$ the new formulas $c_i \neq c_j$ (when $i \neq j$). The model $\mathcal{M}$ still models this extended theory. Thus $T$ is consistent with these new formulas.

By the work from the previous subsection, there is a model of $T$ of cardinality at most $|\mathscr{L}| + \aleph_0$. It is an infinite model, because $T$ requires at least countably many distinct constant symbols. □

Consider the theory ZFC in the language of set theory. Assuming ZFC is consistent, the downward Löwenheim-Skolem theorem says that there is a countable model. After applying a bijection with $\mathbb{N}$, we might as well assume that the domain of the model is $\mathbb{N}$. Yet ZFC proves that there are uncountable sets. This apparently contradictory situation is known as Skolem's paradox.

The solution is to note that inside the model there are indeed no bijections between certain sets. The countability of the model is only seen from the "outside", where we have access to many more functions that are not definable in the theory.

This still raises the question of whether there are really any uncountable sets, or if everything becomes countable inside some bigger system. Of course, if one does not accept uncountable sets, then there is no completed power set of $\mathbb{N}$. Rather, $\mathscr{P}(\mathbb{N})$ is "class sized" and we can only talk about small collections of subsets of $\mathbb{N}$.

Most mathematicians accept the possibility of uncountable sets. They view this paradox as just emphasizing one of the many weaknesses of first order logic. FOL is notoriously bad at controlling infinite cardinalities, or capturing certain concepts. We have seen, previously, that there are other second order concepts that first order logic is simply incapable of talking about. Moreover, some meta concepts cannot be described formally (inside a system), on pain of contradiction.

## 31.D Exercises

**Exercise 31.1.** Let $\mathcal{M}$ be a model of Q. Take $m_0 = \mathbf{0}_{\mathcal{M}}$ and for each $k \in \mathbb{N}$ put $m_{k+1} = S_{\mathcal{M}}(m_k)$. Prove that these elements are distinct. Also prove the equality $m_i +_{\mathcal{M}} m_j = m_{i+j}$ for each $i, j \in \mathbb{N}$.

**Exercise 31.2.** Let $t_1$ and $t_2$ be terms, and let $f$ be a function symbol. Using Mendelson's axiomatic system, prove that if $\vdash_T t_1 = t_2$ then $\vdash_T f(t_1) = f(t_2)$.

**Exercise 31.3.** Let $T$ be a consistent theory, in some language $\mathscr{L}$, containing an existential statement $\exists x\, \psi(x)$. Show that $T \cup \{\psi(c)\}$ is a consistent theory in the language $\mathscr{L}(c)$, where $c$ is a constant symbol not in $\mathscr{L}$. (Hint: Use the result about adjoining new constant symbols, Metatheorem 17.15.)

More generally, let

$$S = \{\psi : \vdash_T \exists v\, \psi(v), \text{ for some variable } v \text{ not occurring in } \psi, \text{ a predicate}\}.$$

Show that if $\{c_\psi\}_{\psi \in S}$ is a collection of distinct constant symbols not in $\mathscr{L}$, then

$$T \cup \{\psi(c_\psi)\}_{\psi \in S}$$

is a consistent theory in the language $\mathscr{L}(\{c_\psi\}_{\psi \in S})$.

**Exercise 31.4.** Let $T$ be a consistent theory, in some language $\mathscr{L}$. Prove that there exists some consistent theory $T' \supseteq T$ that is syntactically complete. (Hint: Use the axiom of choice, in your meta theory. The Zorn's lemma formulation works particularly well. Also, a previous exercise might help here.)

**Exercise 31.5.** Assume $\mathscr{L}_0$ is a language (with at least one constant symbol). Let $T_0$ be a consistent theory in that language.

We will recursively define languages and theories, as follows. For any $i \in \mathbb{N}$, given a consistent theory $T_i$ in a language $\mathscr{L}_i$, then take $T_i'$ to be any syntactically complete, consistent extension of $T_i$. Let $S_i$ be the set of predicates $\psi$ such that $\exists v\, \psi(v)$ is a statement in $T_i'$, for some variable $v$ (free in $\psi$), but $\psi(c)$ does not belong to $T$ for any constant symbol $c$. Take $\mathscr{L}_{i+1} = \mathscr{L}_i \cup \{c_\psi\}_{\psi \in S_i}$ and take $T_{i+1} = T_i' \cup \{\psi(c_\psi)\}_{\psi \in S_i}$.

Finally, take $\mathscr{L}_\infty = \bigcup_{i \in \mathbb{N}} \mathscr{L}_i$ and $T_\infty = \bigcup_{i \in \mathbb{N}} T_i'$. Prove that $T_\infty$ is a consistent, syntactically complete theory in the language $\mathscr{L}_\infty$. Also prove that if $\exists v\, \psi(v)$ is a statement in $T_\infty$, then there exists some constant symbol $c$ such that $\psi(c) \in T_\infty$.

# Chapter VI

# Stability

*Therefore, I would that ye should be steadfast and immovable, always abounding in good works, that Christ, the Lord God Omnipotent, may seal you his...* King Benjamin

Stability is not only a valuable property in people, but also in mathematical objects. We've already seen that transfinite processes often stabilize. There are many other different concepts that exemplify stability; closure and fixed points are just two of them. In this chapter we will investigate these very basic ideas, since they find expression in disparate areas of mathematics.

In this chapter we also include a section on the process of generalization, as a mathematical tool. Entire fields of mathematical research—such as topology, (abstract) algebra, nonstandard geometry, and so forth—are the result of generalized definitions that have reached a place of stability. Not only can definitions be generalized, but so too can arguments and proofs. We saw this previously for Cantor's diagonalization argument. Another natural example in this chapter will be the Knaster-Tarski fixed point theorem.

# 32 Closure operators

The word "closure" suggests a process of

- adjoining objects,
- but only those objects that are needed,
- reaching a state of perfection.

We used this very meaning when defining the reflexive closure of a relation; we add just the right ordered pairs to make the relation reflexive.

Abstractly, let $S$ be an arbitrary set, and let $\mathrm{cl}\colon \mathscr{P}(S) \to \mathscr{P}(S)$ be a function. First, the idea of "adjoining objects" is encapsulated by the condition

$$X \subseteq \mathrm{cl}(X) \text{ for each } X \subseteq S.$$

When this property holds we say that cl is *extensive*. Second, the idea of "reaching a state of perfection" is reflected by the condition

$$\mathrm{cl}(\mathrm{cl}(X)) = \mathrm{cl}(X) \text{ for each } X \subseteq S,$$

in which case cl is said to be *idempotent*; we might say that the closed subsets cannot be made any better. Third, and finally, to reflect the idea that the added objects "are needed" we posit that

$$X \subseteq Y \text{ implies } \mathrm{cl}(X) \subseteq \mathrm{cl}(Y) \text{ for any } X, Y \subseteq S,$$

in which case cl is *monotone*.

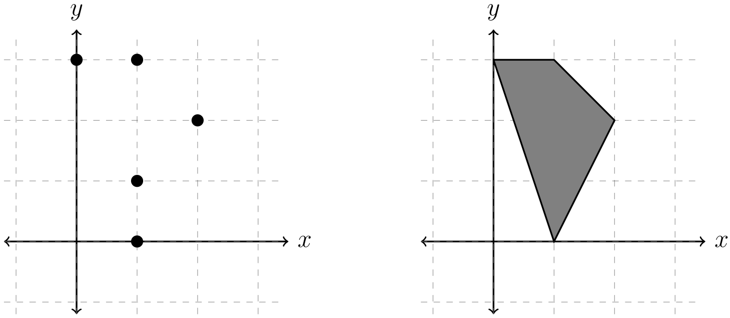
**Definition 32.1.** Let $S$ be a set, and let $\mathrm{cl}\colon \mathscr{P}(S) \to \mathscr{P}(S)$ be some function. If cl is extensive, idempotent, and monotone, then it is a *closure operator* on $S$. For $X \subseteq S$, the set $\mathrm{cl}(X)$ is called the *closure* of $X$. The sets that are fixed by cl (or, equivalently, those in the image of cl) are called the cl-*closed* sets.

The remainder of this section will be devoted to exploring important examples of closure operators throughout mathematics.

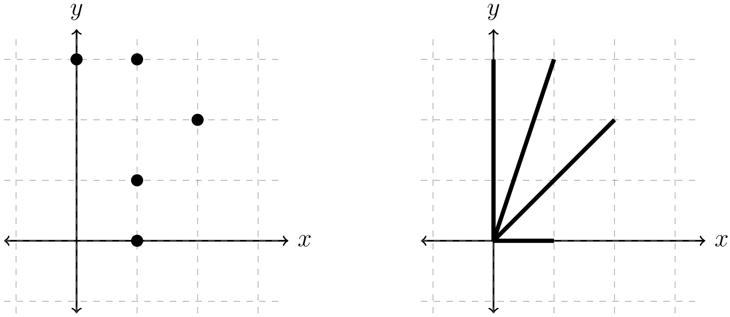
## 32.A Closure operators and topology

Many closure operators arise in the context of shapes and subsets of Euclidean space. A natural example is the "convex hull" operation on $\mathbb{R}^n$. Recall that a subset of $\mathbb{R}^n$ is convex if it contains each of the line segments between any two of its points. The convex hull of $X \subseteq \mathbb{R}^n$ is the smallest convex subset of $\mathbb{R}^n$ containing $X$. (What does "smallest" mean here?) An example from $\mathbb{R}^2$ is given below.

One can vary this operation to make any number of other closure operators. For instance, one could define the "star closure" of $X \subseteq \mathbb{R}^n$ to be the collection of all line segments between the points in $X$ and the origin. A simple example from $\mathbb{R}^2$, using the same points as in the previous example, is graphed below.



Some points in $\mathbb{R}^2$ and their convex hull.



Some points in $\mathbb{R}^2$ and their star closure.

You have likely seen the word "closed" appear in another context with respect to $\mathbb{R}^1$, as follows. The closed intervals are of the following four forms, for arbitrary $a, b \in \mathbb{R}$.

$$\begin{aligned}
[a, b] &= \{x \in \mathbb{R} : a \leq x \leq b\} \\
[a, \infty) &= \{x \in \mathbb{R} : a \leq x\} \\
(-\infty, b] &= \{x \in \mathbb{R} : x \leq b\} \\
(-\infty, \infty) &= \mathbb{R}
\end{aligned}$$

Note that $\emptyset = [0, -1]$ is a closed interval. Also, any singleton $\{a\} = [a, a]$ is a closed interval.

There is a closure operator whose closed sets are exactly the closed intervals. It is just the convex hull operator on $\mathbb{R}^1$. However, this closure operator is too "coarse". It doesn't capture many subsets of $\mathbb{R}^1$ that we would like think of as closed, such as $[0, 1] \cup [3, 4]$.

Thus, we might want to allow finite unions of closed sets to be closed. In other words, we might want our closure operation to satisfy an additional axiom. For each

$n \in \mathbb{N}$, and for any subsets $A_1, A_2, \ldots, A_n \subseteq S$, we want

$$\mathrm{cl}(A_1 \cup A_2 \cup \cdots \cup A_n) = \mathrm{cl}(A_1) \cup \mathrm{cl}(A_2) \cup \cdots \cup \mathrm{cl}(A_n).$$

Any closure operator satisfying this additional "finite union" axiom is called a *topology*. Rather than work with closed sets, topologists instead usually work with the complements of the closed sets, which are called *open*. Note that the $n = 0$ case of this new axiom forces $\mathrm{cl}(\emptyset) = \emptyset$. Thus, the empty set is always closed in a topology. The $n = 1$ case is trivial. The cases of the axiom when $n \geq 2$ all follow from the $n = 2$ case, by induction. It is thus standard to take the $n = 0$ and $n = 2$ cases as separate axioms when defining a topology.

## 32.B Closure operators in algebra and logic

For those familiar with linear algebra, there is a very useful closure operator on sets of vectors called the *span*. Given a (real) vector space $S$, then for $X \subseteq S$ we define

$$\mathrm{Span}(X) = \left\{ s \in S \,:\, s = \sum_{i=1}^{n} \alpha_i x_i \text{ for some } n \in \mathbb{N}, \alpha_i \in \mathbb{R}, \text{ and } x_i \in X \right\}.$$

In other words, $\mathrm{Span}(X)$ is the set of all linear combinations of elements from $X$. It is left as an exercise to prove that this is a closure operator. The closed sets are exactly the subspaces of $S$.

Vector spaces are just one of many different types of algebraic structures that possess a similar closure operator. In general, for an algebraic structure $S$, there is an operator on subsets $X \subseteq S$ called the "subalgebra generated by $X$" operator. If you have had a course in abstract algebra, you may have seen this operator for groups, rings, ideals, or modules. It exists for basically any type of algebra you are likely to encounter. The subalgebra closure operators satisfy an additional axiom, making them *finitary*, in the following sense:

$$\mathrm{cl}(X) = \bigcup \{\mathrm{cl}(Y) \,:\, Y \subseteq X \text{ is a finite subset}\}.$$

In other words, whether or not an element belongs to $\mathrm{cl}(X)$ is determined by a finite amount of information in $X$.

There is another natural example of a finitary closure operator. Let $\mathscr{L}$ be a first order language, and let $S$ be the set of all statements in the language $\mathscr{L}$. For any subset $X \subseteq S$ define the *consequence operator* to be the map that sends $X$ to the set

$$\overline{X} = \{\sigma \in S \,:\, \vdash_X \sigma\}, \tag{32.2}$$

which we might call the set of provable consequences of $X$.

**Proposition 32.3.** *The consequence operator is a finitary closure operation.*

*Proof.* Let $X, Y \subseteq S$ be arbitrary, with $S$ as above. The consequence closure operation is extensive, since each $\sigma \in X$ when treated as an axiom gives a one-line proof of itself. The finitary condition is also easy to see, since any proof involves only finitely many lines, hence only finitely many elements of $X$ are needed to prove any given consequence. Monotonicity is even easier to observe, since if $X \subseteq Y$, then any proof from $X$ is also a proof from $Y$.

To show idempotence, consider $\sigma \in \overline{\overline{X}}$. Then $\sigma$ can be proved from a finite collection $\tau_1, \ldots, \tau_n \in \overline{X}$, and each $\tau_i$ can be proved from a finite collection of statements in $X$. Thus, $\sigma$ can, in turn, be proved from those finitely many statements in $X$, by simply concatenating each of the proofs. □

## 32.C Closure operators on relations

Let $M$ be an arbitrary set. If we take $S = M \times M$, then $\mathscr{P}(S)$ is the collection of all binary relations on $M$. We have already seen many closure operators on $S$. The map $\mathscr{P}(S) \to \mathscr{P}(S)$ given by the rule $R \mapsto R^{\text{ref}}$ is a closure operator; this follows from Proposition 11.4. Similarly, Exercise 11.10 consists of showing that the symmetric and transitive closures are closure operators (hence the names).

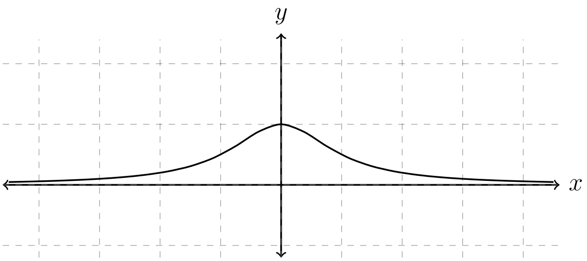
## 32.D Exercises

**Exercise 32.1.** Explain why the convex hull of any subset of $\mathbb{R}^2$ exists, and prove that this yields a closure operator. (As a bonus question, determine whether or not this is a finitary closure operator.)

**Exercise 32.2.** Prove that the span of vectors is a closure operator.

**Exercise 32.3.** Prove that there is a "reflexive and symmetric" closure operator.

**Exercise 32.4.** Let $X$ be the subset of $\mathbb{R}^2$ consisting of the points on the curve $y = \frac{1}{x^2+1}$, as well as the points above this curve. The curve is called the "witch of Agnesi". Show that $X$ is a closed set, but its convex hull is an open set that is not closed (in the usual topological sense).



Graph of the witch of Agnesi.

**Exercise 32.5.** Which of the three conditions of a closure operator does the "irreflexive reduction" satisfy? Fully justify your answer.

**Exercise 32.6.** Show that the consequence closure operation of logic is not a topology, by finding three statements $\sigma_1, \sigma_2, \psi$ in some first order language $\mathscr{L}$ such that $\psi$ is a consequence of $\sigma_1$ and $\sigma_2$ together, but not separately. (Justify all your claims.)

**Exercise 32.7** (Galois connections)**.** Let $S$ and $T$ be sets, and let $R \subseteq S \times T$ be any relation. For $A \subseteq S$ and $B \subseteq T$ put

$$\begin{aligned} A^* &= \{t \in T \,:\, aRt \text{ for every } a \in A\}, \text{ and} \\ B^* &= \{s \in S \,:\, sRb \text{ for every } b \in B\}. \end{aligned}$$

We have thus defined two operations, both written using the same asterisk symbol, one from $\mathscr{P}(S) \to \mathscr{P}(T)$ and the other from $\mathscr{P}(T) \to \mathscr{P}(S)$.

For $X, Y \subseteq S$ and $U, V \subseteq T$, prove the following two properties:

(1) $X \subseteq Y$ implies $X^* \supseteq Y^*$, and similarly $U \subseteq V$ implies $U^* \supseteq V^*$, so these operations are inclusion reversing, and

(2) $X \subseteq X^{**}$ and $U \subseteq U^{**}$, so the double asterisk operators are extensive.

Using only properties (1) and (2), prove that $(\cdot)^{**} \colon \mathscr{P}(S) \to \mathscr{P}(S)$ is a closure operator on $S$, and that the closed subsets of $S$ are exactly those of the form $U^*$. (Of course, dually, the map $(\cdot)^{**} \colon \mathscr{P}(T) \to \mathscr{P}(T)$ will also be a closure operator. It may help to show this at the same time.)

A pair of maps satisfying conditions (1) and (2) are called a *Galois connection*. They show up in many situations, but perhaps the most well known examples comes from studying finite extensions of fields.

Here is another example: Let $S$ be the set of $1 \times n$ matrices (thought of as row matrices) and let $T$ be the set of $n \times 1$ matrices (thought of as column matrices). Define a relation by saying that $sRt$ when the matrix product of $s$ and $t$ is zero. (In a sense, this Galois connection passes between vector spaces and their dual spaces.)

# 33 Another way to think about closure operators

In this section, we show how multiple closure operators can be combined into a single, *joint* operator. For instance, Corollary 11.12 suggests that there should be an "equivalence relation closure" operator. It will be obtained by joining together the reflexive, symmetric, and transitive closure operators.

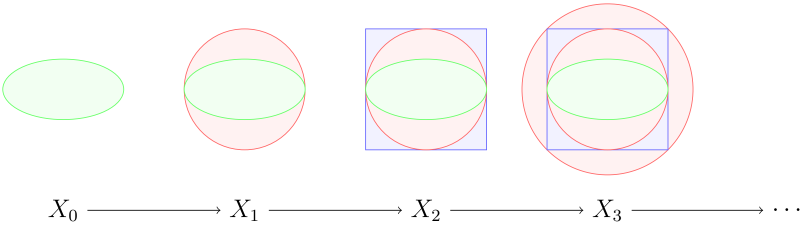
We also give an important alternative characterization of closure operators. This characterization often makes proving statements about closure operators much easier.

## 33.A Bottom-up construction

Suppose we have two closure operators $\mathrm{cl}_1$ and $\mathrm{cl}_2$ on a set $S$. If $X_0 \subseteq S$, what would the joint closure of $\mathrm{cl}_1$ and $\mathrm{cl}_2$ of $X_0$ look like?

The joint closure must contain $X_1 = \mathrm{cl}_1(X_0)$, but the joint closure might not equal $X_1$, since $X_1$ may not be closed under $\mathrm{cl}_2$. Thus, the joint closure must also contain the (potentially) larger set $X_2 = \mathrm{cl}_2(X_1)$. Now, the new set $X_2$ may not be closed under $\mathrm{cl}_1$, so the joint closure must contain the even larger set $X_3 = \mathrm{cl}_1(X_2)$. This process may not end after a finite number of steps, and thus the joint closure must then contain $X_\omega = \bigcup_{n\in\omega} X_n$. This set, again, may not be closed under $\mathrm{cl}_1$, and so we might need to consider $X_{\omega+1} = \mathrm{cl}_1(X_\omega)$, and so forth.

Here is a pictorial representation of this process, where we start with a (green) ellipse, and the two closure operators are represented by enclosing the shape with a (red) circle, or a (blue) square, respectively.



Passing from one closure to the next.

It turns out that this transfinite recursion must eventually stabilize. We have created a nondecreasing sequence $(X_\alpha)_{\alpha\in\mathrm{Ord}}$ of subsets of $S$, indexed by a proper class. However, $\mathscr{P}(S)$ is only a set, so it cannot contain a class-sized strictly increasing chain.

With just a little more work, one can show that any number of closure operators can be combined into a joint closure. One benefit of this transfinite recursion is that it gives us an intuitive idea of how we would build, starting from the *bottom* $X_0$, and working upwards (through bigger and bigger sets), to the finished product. There is another—slicker but less intuitive—way.

## 33.B Top-down construction

There is an important fact about how closure operators interact with intersections.

**Lemma 33.1.** *Let* $\mathrm{cl}\colon \mathscr{P}(S) \to \mathscr{P}(S)$ *be some fixed closure operator on some set* $S$*. If* $C \subseteq \mathscr{P}(S)$ *is a collection of* cl*-closed subsets, then* $\bigcap C$ *is* cl*-closed.*

*Proof.* Set $X = \bigcap C$. Given $s \in S - X$, we can fix some $Y \in C$ such that $s \notin Y$. Since $X \subseteq Y$, the monotone property implies that $\mathrm{cl}(X) \subseteq \mathrm{cl}(Y)$ and the idempotent property implies $\mathrm{cl}(Y) = Y$. Thus $s \notin \mathrm{cl}(X)$. As $s \in S - X$ is arbitrary, this means $\mathrm{cl}(X) \subseteq X$. The reverse containment holds from the extensive property, and thus $X = \mathrm{cl}(X)$, or in other words $X$ is cl-closed. □

A sort of converse is true. Any collection of subsets of $S$ that is closed under intersections gives rise to a closure operator; see Exercise 33.3.

With Lemma 33.1 in hand we can now prove the following.

**Theorem 33.2.** *Let* $\{\mathrm{cl}_i\}_{i \in I}$ *be a collection of closure operators on a set* $S$*. There exists a unique closure operator on* $S$*, called the joint closure* $\mathrm{cl} = \bigvee_{i \in I} \mathrm{cl}_i$*, with the property that any set* $X \subseteq S$ *is* cl*-closed if and only if it is* $\mathrm{cl}_i$*-closed for each* $i \in I$*.*

*Proof.* For each $X \subseteq S$, let

$$P_X = \{Y \in \mathscr{P}(S) \,:\, Y \supseteq X \text{ and } Y \text{ is } \mathrm{cl}_i\text{-closed for each } i \in I\}.$$

These are the joint closed sets containing $X$. Note that $S$ is $\mathrm{cl}_i$-closed for each $i \in I$ (by the extensive property, since $S$ cannot extend to a larger set), and so $S \in P_X$.

Define a map $\mathrm{cl}\colon \mathscr{P}(S) \to \mathscr{P}(S)$ by the rule $\mathrm{cl}(X) = \bigcap P_X$. Since $S \in P_X$, this intersection is contained in $S$, so the rule makes sense. Every set in $P_X$ contains $X$, and so $X \subseteq \mathrm{cl}(X)$. In other words, cl is extensive.

Next, if $Y \in P_X$, then $\mathrm{cl}(X) \subseteq Y$, as $Y$ is one of the terms in the intersection defining $\mathrm{cl}(X)$. Thus $Y \in P_{\mathrm{cl}(X)}$. The converse is obvious, and so $P_{\mathrm{cl}(X)} = P_X$. Thus,

$$\mathrm{cl}(\mathrm{cl}(X)) = \bigcap P_{\mathrm{cl}(X)} = \bigcap P_X = \mathrm{cl}(X).$$

This proves that cl is idempotent.

If $X \subseteq Y \subseteq S$, then $P_X \supseteq P_Y$. Hence

$$\mathrm{cl}(X) = \bigcap P_X \subseteq \bigcap P_Y = \mathrm{cl}(Y),$$

so cl is monotone.

We have now shown that cl is a closure operator. For each $i \in I$, since $\mathrm{cl}(X)$ is an intersection of $\mathrm{cl}_i$-closed sets, it is $\mathrm{cl}_i$-closed by Lemma 33.1. Conversely, if $X \subseteq S$ is $\mathrm{cl}_i$-closed for every $i \in I$, then $X \in P_X$ and so $\mathrm{cl}(X) = \bigcap P_X = X$. Thus cl has the property posited in the statement of the theorem. Uniqueness is left as an exercise for the reader. □

## 33.C Exercises

**Exercise 33.1.** Given two closure operators $\mathrm{cl}_1$ and $\mathrm{cl}_2$ on a set $S$, we described a nondecreasing sequence of sets $(X_\alpha)_{\alpha\in\mathrm{Ord}}$ that stabilizes to the joint closure of $X_0$. Technically, we only described this sequence up to $\alpha = \omega + 1$. Give the general recursive definition of $X_\alpha$, for every ordinal $\alpha$. (Hint: At successor ordinals, instead of alternating between $\mathrm{cl}_1$ and $\mathrm{cl}_2$, there is an easier way.)

**Exercise 33.2.** Let $S$ be a set. Prove that a map $\mathrm{cl}\colon \mathscr{P}(S) \to \mathscr{P}(S)$ is a closure operator if and only if the following condition holds:

$$\text{For any } A, B \subseteq S, \text{ then } A \subseteq \mathrm{cl}(B) \text{ if and only if } \mathrm{cl}(A) \subseteq \mathrm{cl}(B).$$

This gives a single (complicated) criterion for checking that a map is a closure operator.

**Exercise 33.3.** Let $S$ be a set and let $\mathscr{X} \subseteq \mathscr{P}(S)$ with $S \in \mathscr{X}$. Assume that if $C \subseteq \mathscr{X}$ is nonempty, then $\bigcap C \in \mathscr{X}$. Prove that there exists a unique closure operator on $S$ with the property that $\mathscr{X}$ is the collection of closed sets.

**Exercise 33.4.** Finish the proof of Theorem 33.2 by proving the uniqueness of the joint closure.

**Exercise 33.5** (Open problem, proposed by Péter Frankl)**.** Let $S$ be a finite set and let $\mathrm{cl}\colon \mathscr{P}(S) \to \mathscr{P}(S)$ be a closure operator such that $\mathrm{cl}(\emptyset) \neq S$. (Thus, the closure operator is not completely trivial.) Does there exist some $s \in S$ such that $s$ belongs to no more than half of the closed sets?

# 34 Generalization

An important—but often neglected—skill for the professional mathematician is the ability to generalize. When handed a research article with new definitions, especially as a referee, one should not just take those definitions for granted. Instead, one should ask questions like: Why did the authors choose those definitions? Do they include unnecessary hypotheses? Is there any way to generalize the concept being studied, and if so does it simplify the paper?

This section will serve as an example of how this process unfolds. We will take the concept of "closure operators" as studied in the previous section and generalize it. If you took that and other concepts in this book for granted, now would be a good time to start probing all definitions, thus practicing the skill of generalization.

## 34.A Break down the structure

There seem to be three basic components in the definition of a closure operator. First, a closure operator is a function. We could try to generalize this property in a number of ways. One of the common generalizations of "function" is "partial function"; a partial function is a map only defined on part of the intended domain. A partial closure operation may be useful if one were in a situation where the closure of any given set may be undefined. On the other hand, in situations where you want every set to have a closure, this generalization only complicates matters. A good definition walks the balance between generality and usefulness.

The second structural aspect consists of the three defining properties of a closure operator, the extensive, idempotent, and monotone properties. One should naturally ask: Are these properties independent of one another? You will show, while solving Exercise 34.1, that they are indeed independent.

The next question to answer is whether each of those properties is necessary to capture the important consequences of the definition. For instance, did we need all three properties to prove Lemma 33.1? We did, and Exercise 33.3 further shows that maps satisfying those three properties are in bijective correspondence with intersection closed collections of subsets of $S$ containing $S$. Thus, the properties seem to correctly capture the nature of what one means by "closure". This is not to say that it would never be interesting to study idempotent maps, just that such maps are not necessarily closure maps (without the other two properties).

The only remaining structural components of a closure operator are the domain and codomain of the operator. Given a set $S$, a closure operator is a function from $\mathscr{P}(S)$ to itself. Is this necessary? For instance, is it necessary that the domain and codomain are equal? The idempotent property requires that the image of the map should be included in the domain. Thus, taking the domain and codomain equal is quite natural.

Next we might ask: What would prevent us from replacing $\mathscr{P}(S)$ with an arbitrary set $P$? Shockingly, apparently nothing. While the monotone and extensive properties seem to require that some of the elements of $P$ be subsets of each other, these properties still make perfect sense in general. Even more generally, we could replace the subset relation with an arbitrary binary relation $R$ on $P$. This leads us to a preliminary possible generalization of closure operators:

**Definition 34.1** (Only preliminary)**.** Given a set $P$ with a binary relation $R$, a *closure operator* (relative to $R$) is a function $\mathrm{cl}\colon P \to P$ satisfying, for any $a, b \in P$, the following properties:

| | |
|---|---|
| (Extensive) | $aR\mathrm{cl}(a)$ or $a = \mathrm{cl}(a)$, |
| (Idempotent) | $\mathrm{cl}(\mathrm{cl}(a)) = \mathrm{cl}(a)$, and |
| (Monotone) | if $aRb$, then $\mathrm{cl}(a)R\mathrm{cl}(b)$ or $\mathrm{cl}(a) = \mathrm{cl}(b)$. |

To assess whether a new definition does a good job of generalizing the old one, you can use a hindsight test. Did the original definition point towards this generalization? In our case, the original wording when defining closure operations was: "adjoining the needed objects to reach a state of perfection". It is not a big leap to go from "adjoining" to "increasing" and so our generalization is quite natural, at least in the case when $R$ is a poset relation. For the remainder of the section, we will study what conditions we need to place on $R$ to remain close to our original intent.

## 34.B Look for simplifications

It is entirely possible to generalize a concept to the point of inanity, for instance by working in a broader context that requires adding a huge number of additional conditions. While our definition may not yet be inane, it is definitely more complicated. For instance, we needed to slightly alter the extensive and monotone properties, as we are not necessarily assuming $R$ is reflexive. Whenever generalizing a concept you should ask yourself whether or not extra generality, and its associated complexity, really broadens the scope of the concept. With respect to the reflexive property, the answer in our situation is no.

**Lemma 34.2.** *Let $P$ be a set with a binary relation $R$. A function $\mathrm{cl}\colon P \to P$ is a closure operator relative to $R$, if and only if it is a closure operator relative to $R^{\mathrm{irref}}$, if and only if it is a closure operator relative to $R^{\mathrm{ref}}$.*

*Proof.* The idempotent property does not depend on $R$ at all. We will handle the extensive property, leaving the case of the monotone property as an exercise.

If $R \subseteq S$ are two binary relations, and $aR\mathrm{cl}(a)$, then $aS\mathrm{cl}(a)$. So, once the extensive property holds for a relation, it holds for all larger relations. In particular, if cl is extensive under $R^{\mathrm{irref}}$, then it is extensive under $R$, which in turn implies it is extensive under $R^{\mathrm{ref}}$. Thus, it suffices to consider going from $R^{\mathrm{ref}}$ to $R^{\mathrm{irref}}$. If $aR^{\mathrm{ref}}\mathrm{cl}(a)$, then $aR^{\mathrm{irref}}\mathrm{cl}(a)$ unless $a = \mathrm{cl}(a)$. Hence, at least one of the two clauses (in the definition of the extensive property using $R^{\mathrm{irref}}$) still holds. □

In light of Lemma 34.2, it seems appropriate to limit ourselves to reflexive relations when handling closure operators. We lose no generality, but we save some complexity. Thus, our second attempt at a generalized definition could be:

**Definition 34.3** (Second attempt)**.** Given a set $P$ with a reflexive (binary) relation $R$, a *closure operator* (relative to $R$) is a function $\mathrm{cl}\colon P \to P$ satisfying, for any $a, b \in P$, the following properties:

(Extensive) $a R \mathrm{cl}(a)$,
(Idempotent) $\mathrm{cl}(\mathrm{cl}(a)) = \mathrm{cl}(a)$, and
(Monotone) if $aRb$, then $\mathrm{cl}(a) R \mathrm{cl}(b)$.

## 34.C Test generalized theorems

Another way to test whether a broadened definition does a good job at generalization is to check whether basic propositions about that definition similarly generalize. We begin by checking whether or not Exercise 33.2 generalizes. We will study both directions of that exercise separately. Throughout, let $P$ be a set with a reflexive relation $R$, and let $\mathrm{cl}\colon P \to P$ be some function.

**($\Leftarrow$)**: Assume that for any $a, b \in P$,

$$a R \mathrm{cl}(b) \text{ if and only if } \mathrm{cl}(a) R \mathrm{cl}(b). \tag{34.4}$$

Under what conditions can we guarantee that cl is a closure operator?

Taking $a = b$, then (34.4) entails that cl is extensive (since $R$ is reflexive). Next, taking $a = \mathrm{cl}(b)$, then in this case (34.4) entails that $\mathrm{cl}(\mathrm{cl}(b)) R \mathrm{cl}(b)$. The reversal $\mathrm{cl}(b) R \mathrm{cl}(\mathrm{cl}(b))$ holds by the extensive property. Without any further information, we quickly see that we cannot conclude equality. (For instance, consider taking $R = P \times P$. The extensive and monotone properties will hold for free, as will (34.4), but it isn't hard to come up with an example where cl is not idempotent.) However, if we assume that $R$ is antisymmetric, then we can conclude that idempotence holds.

Finally, consider the monotone property. Take as given some $a, b \in P$ with $aRb$. By the extensive property we have $b R \mathrm{cl}(b)$. If $R$ is transitive, then we can conclude that $a R \mathrm{cl}(b)$, and then by (34.4) we obtain $\mathrm{cl}(a) R \mathrm{cl}(b)$, as desired. (Is transitivity necessary here?)

**($\Rightarrow$)**: Now assume $\mathrm{cl}\colon P \to P$ is a closure operator. We want to see under what conditions (34.4) holds. The forward direction is easy, since if $a R \mathrm{cl}(b)$, then by the monotone property we have $\mathrm{cl}(a) R \mathrm{cl}(\mathrm{cl}(b))$, and then the idempotent property tells us that this simplifies to $\mathrm{cl}(a) R \mathrm{cl}(b)$, as desired.

Conversely, assume that $\mathrm{cl}(a) R \mathrm{cl}(b)$ for some $a, b \in P$. By the extensive property $a R \mathrm{cl}(a)$. Again, if $R$ were transitive, we could conclude $a R \mathrm{cl}(b)$, which is what we want. Summing up we have:

**Lemma 34.5.** *Let $(P, R)$ be a nonstrict poset, and let* $\mathrm{cl}\colon P \to P$ *be a map. This map is a closure operator if and only if* (34.4) *holds.*

Technically, antisymmetry and transitivity were applied only in very special cases. Thus, this lemma has some possible room for improvement and further generalization.

The next result we will attempt to generalize is Lemma 33.1. This requires us to define a concept corresponding to $\bigcap C$ for $C \subseteq \mathscr{P}(S)$. The intersection $\bigcap C$, in the poset $(\mathscr{P}(S), \subseteq)$, is exactly the greatest lower bound of $C$. Thus, it would be natural to assume that all subsets of $P$ have greatest lower bounds, in terms of the relation $R$.

Technically, we previously defined greatest lower bounds (and least upper bounds) only with respect to a total order. Try to define greatest lower bounds for an arbitrary reflexive relation on your own. Afterwards, compare your definition with the definition given below.

**Warning 34.6.** If you jump ahead and read the definition below, you will rob yourself of an important learning opportunity.

**Definition 34.7.** Let $P$ be a set with a reflexive relation $R$, and let $C \subseteq P$.

We say $x \in P$ is the *greatest lower bound* for $C$ if

(1) $xRy$ for each $y \in C$,
(2) for each $w \in P$, if $wRy$ for each $y \in C$, then $wRx$, and
(3) $x$ is the unique element satisfying the previous two properties.

We will write $x = \bigwedge C$ in this case.

Uniqueness of greatest lower bounds is connected to an important property on the relation $R$.

**Lemma 34.8.** *Let $P$ be a set with a reflexive relation $R$. If all two element subsets have greatest lower bounds, then $R$ is antisymmetric, which in turn implies that any element satisfying the first two conditions of a greatest lower bound also satisfies the third (uniqueness) condition.*

*Proof.* Let $a, b \in P$ with $aRb$ and $bRa$. Consider the set $C = \{a, b\}$. From the reflexive property, $aRa$. Thus $a$ satisfies the first property of a greatest lower bound for $C$.

Next, suppose $w \in P$ satisfies $wRy$ for each $y \in C$. In other words, $wRa$ and $wRb$. Then, tautologically, $wRa$. This shows that $a$ satisfies the second property of being a greatest lower bound for $C$.

By symmetry, $b$ also satisfies the first two conditions of being a greatest lower bound for $C$. Thus, if $C$ has a (unique) greatest lower bound, then $a = b$, which verifies that $R$ is antisymmetric.

Now just assume that $R$ is antisymmetric, and assume that $x, x' \in P$ are elements that satisfy conditions (1) and (2) of a greatest lower bound, for some subset $C \subseteq P$. As $x$ satisfies condition (1), and as $x'$ satisfies condition (2), we obtain $xRx'$. Similarly $x'Rx$. Thus $x = x'$. □

*Least upper bounds* are defined in a similar way to greatest lower bounds, and the previous lemma remains true after replacing greatest lower bounds with least upper bounds (by a symmetrical argument). Interestingly, existence of least upper bounds is connected to existence of greatest lower bounds.

**Proposition 34.9.** *Let $P$ be a set with a reflexive relation $R$. All subsets of $P$ have greatest lower bounds if and only if all subsets have least upper bounds.*

*Proof.* Symmetry considerations tell us that we only need to prove one direction, as the other will be similar. So assume all subsets of $P$ have greatest lower bounds. Letting $C \subseteq P$, we need to find a least upper bound for $C$.

Let $D = \{y \in P : wRy \text{ for each } w \in C\} \subseteq P$. The set $D$ consists of the "upper bounds" on $C$. Let $x = \bigwedge D$; we will show that $x$ is a least upper bound for $C$. First, fix $w \in C$. For any $y \in D$ we have $wRy$. So by condition (2) on greatest lower bounds, and since $x = \bigwedge D$, we must have $wRx$. As $w \in C$ is arbitrary, we see that $x$ satisfies the first condition of being a least upper bound for $C$.

Next, assume that $y \in P$ satisfies $wRy$ for each $w \in C$. Then $y \in D$, and so $xRy$. This proves that $x$ satisfies the second condition of being a least upper bound for $C$.

Finally, the third (uniqueness) condition follows from Lemma 34.8, since $R$ is antisymmetric. □

We are now ready to prove the corresponding generalization of Lemma 33.1.

**Lemma 34.10.** *Let $P$ be a set with a reflexive relation $R$ such that all subsets of $P$ have greatest lower bounds. Let $\mathrm{cl}\colon P \to P$ be closure operator. If $C \subseteq P$ is a collection of* cl*-closed elements of $P$, then $\bigwedge C$ is* cl*-closed.*

*Proof.* Set $x = \bigwedge C$ and let $z = \mathrm{cl}(x)$. For any $y \in C$, then $xRy$ by condition (1) of greatest lower bounds. Then the monotone property implies that $zR\mathrm{cl}(y)$. Since $C$ is a collection of closed elements, we have $y = \mathrm{cl}(y)$ by idempotence, and hence $zRy$. As $y \in C$ is arbitrary, condition (2) of greatest lower bounds yields $zRx$. On the other hand $xRz$ by the extensive property. Antisymmetry must hold because of Lemma 34.8, and so $x = z = \mathrm{cl}(x)$. □

## 34.D Advice

The best mathematical writers craft their definitions carefully. All variable names are chosen intentionally. Properties are simplified and streamlined. Unnecessary conditions are dropped. This not only helps readers understand the material being presented, it makes it easier to present. Results flow naturally from the definitions.

## 34.E Exercises

**Exercise 34.1.** Find a set $S$ and eight examples of functions $\mathscr{P}(S) \to \mathscr{P}(S)$, one for each of the eight possible combinations of being, or not being, extensive, idempotent, or monotone. (For added complexity one might add "finitary" to the mix and try to make the cardinality of $S$ as small as possible.)

**Exercise 34.2.** Finish the proof of Lemma 34.2.

**Exercise 34.3.** Give an example of a set $P$ with a reflexive, antisymmetric relation $R$, such that every subset of $P$ has a greatest lower bound, but $R$ is not transitive. (Hint: Let $P = \{0, a, b, c, 1\}$, let $\leq$ be the total order that yields $0 < a < b < c < 1$, and let $R = \leq - \{(a, c)\}$.)

**Exercise 34.4.** Give an example of a set $P$, together with a reflexive relation $R$, a subset $C \subseteq P$, and distinct elements $x, x' \in P$ satisfying properties (1) and (2) in Definition 34.7.

**Exercise 34.5.** Is there a set $P$ with a reflexive, antisymmetric relation $R$, together with a closure operator $\mathrm{cl}\colon P \to P$, such that (34.4) holds, but $R$ is not transitive?

**Exercise 34.6.** Generalize Theorem 33.2.

**Exercise 34.7.** Definition 34.7 was in terms of a reflexive relation. Modify the definition to an arbitrary relation $R$ (by replacing each instance of "$aRb$" with "$aRb$ or $a = b$"), and show that a subset $C \subseteq P$ has a greatest lower bound with respect to $R$ if and only if it has a greatest lower bound with respect to $R^{\mathrm{ref}}$. (Hint: It might help to show, more generally, that the lower bounds on $C$ are exactly the same in both cases.)

**Exercise 34.8.** Our original definition of well-foundedness for a relation $R$ required us to think of the left side of $R$ as *strictly* smaller. One could modify the definition to allow for more generality, as follows. Let us say that a relation $R$ is well-founded if every nonempty set $S$ has an $R$-minimal element $y \in S$, in the (more general) sense that

$$\forall x\, ((x \in S \wedge x \neq y) \to \neg(xRy)).$$

Show the following.

(1) A relation $R$ is well-founded if and only if $R^{\mathrm{ref}}$ is well-founded.

(2) A relation $R$ is well-founded if and only if $R^{\mathrm{irref}}$ is well-founded (under the original definition).

The motivated reader might check what modifications would be needed, if any, to redo Section 17 in this greater generality.

**Exercise 34.9.** Let $M$ be a set, let $S = M \times M$, and let $\mathscr{P}(S)$ be the set of all binary relations on $M$. Is irreflexive reduction $X \mapsto X^{\mathrm{irref}}$ a closure operation $\mathscr{P}(S) \to \mathscr{P}(S)$? (Warning: Use the generalized definition.)

# 35 Lattices

In mathematics, there are many different concepts that go by the name "lattice". Here, we will not be studying those lattices related to recurring patterns. Instead, we study an abstract algebraic structure related to greatest lower bounds.

In the previous section, we noted that for any set $S$, its power set $\mathscr{P}(S)$ is a poset under the inclusion relation $\subseteq$. Moreover, every nonempty subset $C \subseteq \mathscr{P}(S)$ has a greatest lower bound in this poset, given by $\bigcap C$. The study of lattices is in one sense the study of greatest lower bounds, but surprisingly lattices can be defined without direct reference to any poset (or more general) structure.

Much of the material in this section is very abstract. If you are struggling to think of examples satisfying (or failing) the conditions of the results, take $M$ to be a subset of $\mathscr{P}(S)$ (for some set $S$) and take $R$ to be the inclusion relation on $M$. The picture in Example 11.15 may also be useful.

## 35.A Defining properties—part one

Let $M$ be a set and let $R$ be a reflexive relation on $M$. We are interested in situations where $M$ has greatest lower bounds (with respect to $R$) for subsets $C \subseteq M$. We continue to use the notation $\bigwedge C$, when the greatest lower bound of $C$ exists, and call this the *meet of* $C$. Similarly, a least upper bound for $C$ will be written as $\bigvee C$, and will be read as the *join of* $C$. We will focus on the case when $C$ is finite. To simplify notation, when $C = \{x, y\}$ we will write $x \wedge y$ for $\bigwedge C$, and similarly for joins.

We begin by studying what happens when pairs in $M$ have greatest lower bounds. The *commutative property*,

$$\forall x, y \in M,\ x \wedge y = y \wedge x,$$

must hold since $x \wedge y$ and $y \wedge x$ are simply different ways of writing $\bigwedge\{x, y\}$, which does not depend on the order of the two elements. For the pair $\{x, x\}$, the *idempotence property*,

$$\forall x \in M,\ x \wedge x = x.$$

must also hold. (The idempotence property for meets is not to be confused with the idempotence property for closure operators). In fact, the idempotence property does not require the existence of meets for all pairs, as the following lemma makes clear.

**Lemma 35.1.** *Let $M$ be a set with a reflexive relation $R$, and also let $x \in M$. If $\bigwedge\{x\}$ exists, then $x \wedge x = x$.*

*Proof.* Assume $\bigwedge\{x\}$ exists. Note that $x$ satisfies condition (1) of Definition 34.7, taking $C = \{x\}$. It also satisfies condition (2), since for each $w \in M$ satisfying $wRy$ for each $y \in \{x\}$, we have (automatically) $wRx$. So the uniqueness condition (3) forces $x = \bigwedge\{x\} = x \wedge x$. □

We still haven't mentioned what happens for the smallest subset of $M$, the empty set. If $\bigwedge \emptyset$ exists, it is the unique greatest element of $M$. It is common to write $1 = \bigwedge \emptyset$, and similarly to write $0 = \bigvee \emptyset$. When it exists, 1 is called the *top* element, and similarly 0 is called the *bottom* element.

**Warning 35.2.** This is another example of mathematicians overusing notation. These 0 and 1 have very little to do with $0, 1 \in \mathbb{N}$.

**Advice 35.3.** Recall that "meets" were meant to generalize "intersections". For any nonempty subset $C \subseteq \mathscr{P}(S)$, then $\bigcap C$ is defined. While $\bigcap \emptyset$ is not defined, it is sometimes convenient (following this analogy) to take $\bigcap \emptyset = S$, which is the top element in the poset $(\mathscr{P}(S), \subseteq)$. If you do this, be kind to the reader and say so. (Note: There is no corresponding problem with "joins" and "unions".)

Now that we have discussed meets for the subsets of $M$ with cardinality 2 or less, consider a triplet set $\{x, y, z\} \subseteq M$. There are three natural quantities to consider:

(1) $\bigwedge\{x, y, z\}$,
(2) $(x \wedge y) \wedge z$, and
(3) $x \wedge (y \wedge z)$.

We would like all three quantities to be the same. The equality of the last two objects, universally quantified, is the *associativity property*,

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

This property again implies a familiar condition on the relation $R$.

**Lemma 35.4.** *Let $M$ be a set with a reflexive relation $R$ such that all pairs have meets. The associative property holds for meets if and only if $R$ is transitive.*

*Proof.* First let $x, y, z \in M$ with $xRy$ and $yRz$. It is not hard to show that $x \wedge y = x$ and $y \wedge z = y$. (Try to justify one of these for yourself.) Using associativity,

$$x \wedge z = (x \wedge y) \wedge z = x \wedge (y \wedge z) = x \wedge y = x.$$

Hence $xRz$, which shows the transitivity of $R$.

We now prove the converse. Let $x, y, z \in M$. Let $s = x \wedge (y \wedge z)$ and $t = (x \wedge y) \wedge z$. We will show that $sRt$. Note that $sRx$ and $sR(y \wedge z)$. Hence, by transitivity, $sRy$ and $sRz$. From the second property of meets (applied to the meet $x \wedge y$) we get $sR(x \wedge y)$, and then similarly $sR((x \wedge y) \wedge z)$, hence $sRt$. By a symmetrical argument $tRs$, so by antisymmetry (from Lemma 34.8) we get $s = t$. □

If we want all nonempty finite subsets of $M$ to have greatest lower bounds, and that those meets are expressible in terms of the pairwise meet, then we should assume that $R$ is a poset relation on $M$. Thus, in such situations we will write $(M, \leq)$ instead of $(M, R)$.

It is time to explain where the names "meet" and "join" come from. The meet of two elements is the place where they "first meet together" when going downwards in the poset, and the join is where they "first join together" when going upwards in the poset. Pictorially:

$$\begin{tikzcd}
 & x \vee y & \\
x \arrow[ur, dash] \arrow[dr, dash] & & y \arrow[ul, dash] \arrow[dl, dash] \\
 & x \wedge y &
\end{tikzcd}$$

**Advice 35.5.** It can be disconcerting that the symbol "$\wedge$" is used to denote where two elements meet, when going downward in a poset, since one's eyes must move upwards when passing from the edges of the symbol to the middle. Remember that the symbol is generalizing "$\cap$".

We now address the equality $\bigwedge\{x, y, z\} = (x \wedge y) \wedge z$.

**Proposition 35.6.** *Let $(M, \leq)$ be a poset such that all pairs have meets. Given $x, y, z \in M$, then $\bigwedge\{x, y, z\}$ exists and equals $(x \wedge y) \wedge z$.*

*Proof.* Set $v = (x \wedge y) \wedge z$. We will show that $v$ satisfies the conditions making it a greatest lower bound for the set $C = \{x, y, z\}$. From Lemma 34.8, it suffices to show the first two conditions of a greatest lower bound.

Clearly $v \leq z$. Also by associativity, $v = x \wedge (y \wedge z)$, so $v \leq x$. From commutativity, and another application of associativity, we obtain

$$v = (y \wedge x) \wedge z = y \wedge (x \wedge z)$$

and hence $v \leq y$. Therefore $v$ is a lower bound on $C$ (i.e., it satisfies condition (1) in Definition 34.7).

Next, suppose that $w \in M$ is a lower bound on $C$. In particular, $w \leq x \wedge y$. Hence $w \leq (x \wedge y) \wedge z = v$. This show that condition (2) holds for $v$. □

Similar arguments can now be used to handle finite sets of larger sizes, as described in Exercise 35.3. Putting this all together, we have the following nice result.

**Proposition 35.7.** *Let $(M, \leq)$ be a poset such that all pairs have meets. The pairwise meet function $\wedge \colon M \times M \to M$ satisfies the idempotence, commutative, and associative properties. All nonempty finite subsets of $M$ have meets, and these can be expressed using the pairwise meet function (in any of the obvious ways—they are all equivalent).*

## 35.B Defining properties—part two

The idempotence, commutativity, and associativity properties are just three properties among infinitely many that greatest lower bounds (on pairs) satisfy. We chose to focus on them for historical reasons, as well as for aesthetic appeal. A natural question to ask is: What properties are we missing?

The answer is that we are missing nothing; any other property that $\wedge$ satisfies is a consequence of those three basic properties. This may seem surprising, or appear difficult to prove. We prove it by showing that for any set $M$ and any function $\wedge\colon M \times M \to M$ satisfying those three properties, we can reconstruct (uniquely) a partial order on $M$ that makes $\wedge$ the pairwise greatest lower bound operation.

**Theorem 35.8.** *Let $M$ be a set and let $\wedge\colon M\times M \to M$ be a function satisfying the idempotence, commutativity, and associativity properties. The relation*

$$R = \{(x, y) \in M \times M \ : \ x = x \wedge y\}$$

*is a partial ordering of $M$. Under this partial ordering all pairs have meets and its pairwise meet function is $\wedge$. Moreover, $R$ is the unique partial ordering with these properties, called the* induced *partial order.*

*Proof.* **(Reflexive)**: Let $x \in M$. By idempotence $x = x \wedge x$, and so $(x, x) \in R$.

**(Antisymmetric)**: Let $x, y \in M$. Assume $(x, y), (y, x) \in R$, so $x = x \wedge y$ and $y = y \wedge x$. By commutativity $x \wedge y = y \wedge x$, so $x = y$.

**(Transitive)**: Let $x, y, z \in M$. Assume $(x, y), (y, z) \in R$. So $x = x \wedge y$ and $y = y \wedge z$. Then, using associativity for the third equality below, we find

$$x = x \wedge y = x \wedge (y \wedge z) = (x \wedge y) \wedge z = x \wedge z.$$

Hence $(x, z) \in R$.

**(Meets are unchanged)**: Let $y_1, y_2 \in M$ and put $x = y_1 \wedge y_2$. We need to show that $x$ is the greatest lower bound for $\{y_1, y_2\}$ in the poset $(M, R)$. Using the three properties of $\wedge$, we compute

$$x \wedge y_1 = (y_1 \wedge y_2) \wedge y_1 = (y_2 \wedge y_1) \wedge y_1 = y_2 \wedge (y_1 \wedge y_1) = y_2 \wedge y_1 = y_1 \wedge y_2 = x. \tag{35.9}$$

Thus $(x, y_1) \in R$. Similarly $(x, y_2) \in R$, and hence $x$ is a lower bound on $\{y_1, y_2\}$.

If $w$ is any other lower bound, then $wRy_1$ and $wRy_2$. In other words, $w = w \wedge y_1$ and $w = w \wedge y_2$. We find

$$w \wedge x = w \wedge (y_1 \wedge y_2) = (w \wedge y_1) \wedge y_2 = w \wedge y_2 = w.$$

Thus $wRx$, which shows that $x$ is the greatest lower bound.

**(Uniqueness of the relation)**: This is left as an exercise. □

## 35.C Defining properties—part three

Let $(M, \leq)$ be a poset such that all pairs have **joins** (rather than meets). Symmetry considerations tells us that after making the necessary modifications, the work in the previous two sections can be redone with this new information. A poset with pairwise joins is equivalent to being given a function $\vee \colon M \times M \to M$ that satisfies the idempotence, commutativity, and associativity properties.

Now suppose that $(M, \leq)$ is a poset such that all pairs have both meets and joins. We then have two functions, $\wedge$ and $\vee$, both satisfying the idempotence, commutativity, and associativity properties. Theorem 35.8 says that from either of these two functions we can reconstruct the partial order on $M$. Conversely, given such functions it is a natural question whether we can express—solely in terms of properties on $\wedge$ and $\vee$—that the partial order reconstructed from $\wedge$ is the same as that from $\vee$.

Note that the idempotence, commutativity, and associativity conditions are not sufficient. Let $M$ be any set with two different poset relations $\leq$ and $\leq'$, such that all pairs have both meets and joins under those relations. If we take $\wedge$ to be the meet operation under the relation $\leq$, and $\vee$ to be the join relation under $\leq'$, then idempotence, commutativity, and associativity hold for both functions, but the partial orders induced by these functions are not the same.

**Example 35.10.** Taking $M = \mathscr{P}(\{\emptyset\}) = \{0, 1\}$, then take $\leq$ to be the usual non-strict order relation $\subseteq$, and also take $\leq'$ to be the reversed relation $\supseteq$. The pairwise meet function with respect to $\leq$ is simply pairwise intersection on $M$. The pairwise join function with respect to $\leq'$ is also pairwise intersection on $M$. (Since these are the same functions, why do they induce different relations $\subseteq$ and $\supseteq$?) △

Similar reasoning shows that no additional conditions that are placed solely on $\wedge$ or $\vee$ will be enough to guarantee that the partial orders they induce are the same. Instead, we should look for some joint condition involving both $\wedge$ and $\vee$ at the same time.

Thinking of joins as "going up" in a poset, and meets as "coming down", then two natural conditions are the following *compatibility* (or *absorption*) *properties*,

$$\forall x, y \in M,\ x \wedge (x \vee y) = x \text{ and } x \vee (x \wedge y) = x.$$

These simple conditions are both necessary and sufficient to guarantee that the induced partial orders are the same.

**Lemma 35.11.** *Let $M$ be a set and let $\wedge, \vee \colon M \times M \to M$ be any two functions satisfying the idempotence, commutativity, and associativity properties. Let $R_1$ be the partial order relation for which $\wedge$ is the pairwise meet function. Let $R_2$ be the partial order relation for which $\vee$ is a pairwise join function. Then $R_1 = R_2$ if and only if the compatibility conditions hold.*

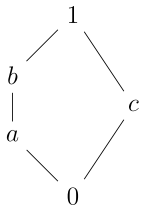
We can now define algebraic lattices.

**Definition 35.12.** A *lattice* is a set $M$ together with two functions, denoted $\wedge, \vee\colon M \times M \to M$, that satisfy the idempotence, commutativity, associativity, and compatibility properties. It is a poset under a unique relation $\leq$ such that $\wedge$ is the pairwise meet function (or, equivalently, $\vee$ is the pairwise join function), which we call the *induced* partial order.

The next section will be devoted to giving examples of lattices and some applications. They arise in many natural situations.

## 35.D Exercises

**Exercise 35.1.** Consider the poset, called $N_5$, given by the following diagram.



Prove that this is a lattice. Further show that there exist elements $x, y, z \in N_5$ with $x \leq z$ but $x \vee (y \wedge z) \neq (x \vee y) \wedge z$. (A sort of converse is also true. Such an inequality will hold somewhere in a lattice if and only if the lattice "contains" a copy of $N_5$. Can you sketch a proof of this converse?)

**Exercise 35.2.** Draw the diagram of a poset that is not a lattice. Try to make the cardinality as small as possible.

**Exercise 35.3.** Prove the following generalization of Proposition 35.6: Let $(M, \leq)$ be a poset with meets for all pairs. Given any $n \in \mathbb{Z}_{\geq 2}$, and any elements $x_1, \ldots, x_n \in M$, then

$$x_1 \wedge (x_2 \wedge \cdots (x_{n-1} \wedge x_n) \cdots ) = \bigwedge \{x_1, x_2, \ldots, x_{n-1}, x_n\}.$$

**Exercise 35.4.** For each of the equalities in (35.9), explain which property is being used to obtain that specific equality. Similarly, fully justify that $x \wedge y_2 = x$.

**Exercise 35.5.** Prove the uniqueness claim in Theorem 35.8.

**Exercise 35.6.** Prove that the relation $R$ in Theorem 35.8 can alternatively be described as

$$R = \{(x \wedge y, y) \,:\, x, y \in M\}.$$

**Exercise 35.7.** Prove that the collection of reflexive relations on a set $S$ is a lattice, with "meet" given by "intersection" and "join" given by "union". Do the same for the symmetric and transitive relations, except that in the latter case the join is the transitive closure of the union.

**Exercise 35.8.** Prove Lemma 35.11.

**Exercise 35.9.** Let $M$ be a set with two functions $\wedge, \vee \colon M \times M \to M$. Show that if both of the compatibility conditions hold, then the idempotence conditions hold.

**Exercise 35.10.** Let $M$ be a set with a reflexive relation $R$ such that all pairs have meets. Letting $\wedge$ denote the pairwise meet function, as usual, prove that it satisfies the following conditions:
(C1) $\forall x, y \in M$, $(x \wedge y) \wedge y = x \wedge y$.
(C2) $\forall x, y, z \in M$, if $x \wedge y = x$ and $x \wedge z = x$, then $x \wedge (y \wedge z) = x$.

Conversely, let $M$ be a set and let $\wedge \colon M \times M \to M$ be a function satisfying the idempotence and commutative properties in addition to (C1) and (C2). Putting

$$R = \{(x, y) \in M \times M \ : \ x = x \wedge y\}$$

show that $R$ is a reflexive relation on $M$. Show that under this relation all pairs have meets and its pairwise meet function is $\wedge$. Moreover, prove that $R$ is the unique reflexive relation with these properties.

(The motivated student might try showing that (C1) and (C2) are independent of one another, as well as independent of the idempotent and commutative properties, even jointly. Note that if $R$ is transitive and $\wedge$ is idempotent, then (C1) holds, while (C2) follows just from transitivity of $R$. One might search for an example satisfying all of the conditions in the previous paragraph, but with $R$ not transitive. Other questions to answer might include the following: Can condition (C2) be replaced by a universally quantified equality? Can Lemma 35.11 be proved without associativity or transitivity? Does the concept of lattice generalize to sets with only reflexive relations?)

# 36 Examples and uses of lattices

## 36.A Natural examples

The prototypical example of a lattice is the poset $(\mathscr{P}(S), \subseteq)$, for any set $S$. However, this lattice has a special property not enjoyed by every lattice; not only do meets and joins of finite subsets exist, the meets and joins of arbitrary subsets exist. This leads to the following definition.

**Definition 36.1.** We say that a poset $(P, \leq)$ is a *complete lattice* if all subsets of $P$ have greatest lower bounds.

**Example 36.2.** All finite lattices are complete, because all subsets are finite. Is the lattice of finite subsets of $\mathbb{N}$ complete? △

Note that by Proposition 34.9, such a structure has both meets and joins, so it really is a lattice. Our use of the word "complete" here differs slightly from our previous usage for totally ordered sets (since we don't allow unbounded subsets of $P$). All complete lattices must have top and bottom elements.

Thus, to construct more examples of lattices, one can take any subset of $\mathscr{P}(S)$ that is closed under arbitrary intersections. (This should include the empty intersection, which we consider is equal to $S$.)

Other natural examples of (complete) lattices come from algebra. Given a vector space $V$, then the collection of subspaces forms a complete lattice. The meets are given by intersection, but joins are not unions. Instead, the join of two subspaces is the span of their union.

If you are familiar with groups (or rings), you may know that the subgroups (or subrings) form a complete lattice. Meets are again formed by intersections, but joins are formed by taking the subgroup (or subring) generated by the union.

These are all just special cases of the following general construction.

**Theorem 36.3.** *Let $(P, \leq)$ be a complete lattice, let $\mathrm{cl}\colon P \to P$ be a closure operator, and let $Q$ be the set of $\mathrm{cl}$-closed elements of $P$. The poset $(Q, \leq|_Q)$ is a complete lattice, where meets are formed by taking the same meet from $P$, and joins are formed by taking the closure of the join from $P$.*

*Proof.* Except for the very last clause, this follows from Lemma 34.10. Given any subset $C \subseteq Q$, then $\bigvee_Q C$ is an upper bound for $C$ in $Q$, hence it is an upper bound in $P$. Thus, $\bigvee_P C \leq \bigvee_Q C$. Using all three properties of a closure operator, we have

$$\bigvee_P C \leq \mathrm{cl}(\bigvee_P C) \leq \mathrm{cl}(\bigvee_Q C) = \bigvee_Q C.$$

Since $\bigvee_Q C$ is the *least* upper bound in $Q$ for $C$, there is equality in the middle. □

**Example 36.4.** Let $\mathscr{L}$ be a first order language, and let $S$ be the set of all statements in the language $\mathscr{L}$. Define the consequence operator as in (32.2). The closed sets form a complete lattice, with meets given by intersections. The join of collections of statements is the consequence closure of their union.

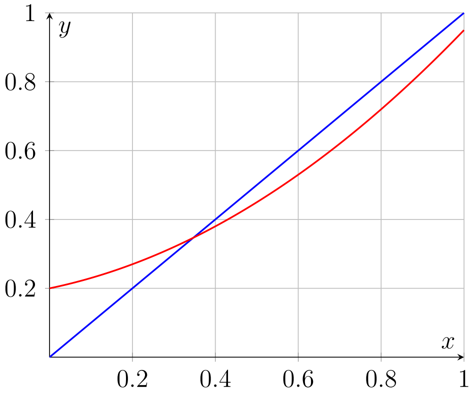
The bottom element of this lattice is the collection of tautological statements. The top element is $S$, the set of all statements, which is an inconsistent theory. (In fact, the consequence closure of any inconsistent subset of $S$ is $S$. This principle is called *ex falso quodlibet*, which means "from falsehood anything [follows]".)

To see explicitly that the join of two subsets of $S$ is not always just their union, let $\mathscr{L}$ be the language whose signature consists of a single binary relation $<$. Let $S_1$ be the consequence closure of the statement asserting that $<$ is irreflexive. Similarly, let $S_2$ be the consequence closure of the statement asserting that $<$ is transitive. Now, $S_1 \cup S_2$ does not contain the statement that $<$ is asymmetric, because we cannot prove that $<$ is asymmetric just using the irreflexive or the transitive property alone. However, together they do imply asymmetry, and so their join is $\overline{S_1 \cup S_2} = \overline{\mathrm{POSET}}$. $\triangle$

These examples illustrate an interesting phenomenon. Given a complete lattice $(P, \leq)$ there are many subsets $Q \subseteq P$ that remain complete lattices when restricting the poset relation; i.e., $(Q, \leq|_Q)$ is a complete lattice. Moreover, the meet function in $Q$ can be the restriction of the meet function in $P$; or in symbols $\bigwedge_Q = (\bigwedge_P)|_{Q\times Q}$. All of this can happen and yet $\bigvee_Q \neq (\bigvee_P)|_{Q\times Q}$; that is, the join function in $Q$ is not the restriction of the join function from $P$. This reveals that the relationship between the order relation and its corresponding meet and join functions can be complicated.

## 36.B Fixed points

The results in this subsection are based on one of many possible generalizations of the following example. Consider the closed interval $[0, 1] \subseteq \mathbb{R}$. Let $f\colon [0, 1] \to [0, 1]$ be any continuous function. We claim that there exists some point $c \in [0, 1]$ such that $f(c) = c$, a so-called *fixed point*. The idea is pictured below.



Here is the quick proof of the claim.

*Proof.* If $f(0) = 0$ or $f(1) = 1$, then we are done. So we may assume $f(0) > 0$ and $f(1) < 1$. The function $g\colon [0,1] \to \mathbb{R}$, given by the rule $g(x) = f(x) - x$, is also continuous. It is positive at 0, but negative at 1. Thus, by the intermediate value theorem, there is some point $c \in (0,1)$ such that $g(c) = 0$. Equivalently, $f(c) = c$. □

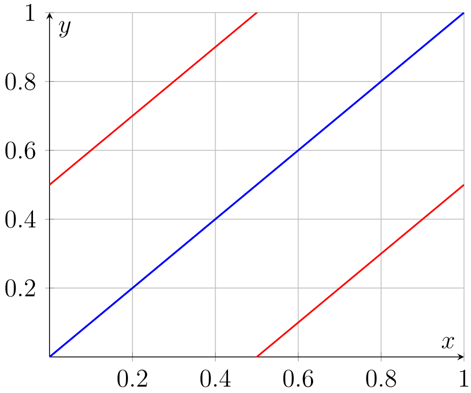
To begin generalizing this result, let's find the parts that seem necessary. If we consider arbitrary functions $f\colon S \to T$, there is no hope of fixed points if there exists some $t \in T - S$, since we could just take $f$ to be the constant function $f(x) = t$. Whatever hypotheses we eventually settle on, we do not want to rule out constant functions. So, the codomain should at least be included in the domain; and thus we might as well take $T = S$.

Next, if we replace the closed interval $[0,1]$ with the open interval $(0,1)$, then the proof fails. (Can you find a function $f\colon (0,1) \to (0,1)$ without a fixed point?) The endpoints are adding something necessary.

Finally, let's consider the continuity hypothesis. It too adds something to the example; consider the piecewise increasing function $f\colon [0,1] \to [0,1]$ given by the rule

$$f(x) = \begin{cases} x + \frac{1}{2} & \text{for } x < \frac{1}{2}, \\ x - \frac{1}{2} & \text{for } x \geq \frac{1}{2}. \end{cases}$$

The graph of $f$ is given below, in red.



In this example, the problem is not that the function has a jump, but that the jump is in the opposite direction to the growth of the function. Thus we might ask if being monotone is sufficient. It is.

**Theorem 36.5** (Knaster-Tarski theorem). *Let $(P, \leq)$ be a complete lattice and let $f\colon P \to P$ be a monotone function. Then $f$ has a fixed point.*

*Proof.* Let $C = \{x \in P : f(x) \leq x\}$. Put $y = \bigwedge C$. Given $x \in C$ then $y \leq x$. Hence $f(y) \leq f(x) \leq x$. Since $x \in C$ is arbitrary, $f(y) \leq \bigwedge C = y$. Thus $y \in C$, and so it must be the least element of $C$.

Now, since $f$ is monotone, we have $f(f(y)) \leq f(y)$. So $f(y) \in C$ as well. By minimality, $f(y) = y$. □

Anne Davis has shown that, conversely, if $(P, \leq)$ is a lattice such that every monotone function $f\colon P \to P$ has a fixed point, then $P$ is a complete lattice. Thus, fixing points using monotone lattice maps generally requires completeness.

## 36.C Modeling set operations with lattices

Meets generalize intersections, and joins generalize unions. Other set conditions can similarly be reflected by conditions in lattices. For instance, the empty set and the full set are reflected in the following type of lattice.

**Definition 36.6.** Let $(P, \leq)$ be a lattice. We say $P$ is *bounded* if it has a top element 1 and bottom element 0.

In any bounded lattice, we have the following *identity properties*:

$$\forall x \in P,\ x \vee 0 = x \text{ and } x \wedge 1 = x.$$

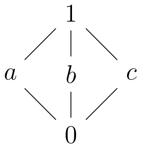
Conversely, any element 0 satisfying the universally quantified identity $x \vee 0 = x$ must be the bottom element (and symmetrically for 1 and the other equality).

What about generalizing complementation?

**Definition 36.7.** Let $(P, \leq, \wedge, \vee, 0, 1)$ be a bounded lattice. A *complement* for an element $x \in P$ is any $y \in P$ such that $x \vee y = 1$ and $x \wedge y = 0$. We say that $P$ is *complemented* if every element has a complement.

Complements in lattices do not need to be unique. For example, in the lattice $N_5$ described in Exercise 35.1, both $a$ and $b$ are complements to $c$. Moreover, even in bounded lattices, complements do not necessarily exist. (Can you construct an example?)

**Example 36.8.** Consider the lattice $M_3$ given by the following diagram.



One can check that every element has a complement, so this is a complemented lattice. Only the top and bottom elements have unique complements. △

When first learning about intersections, unions, and complements, students are taught some basic properties. These include the commutative rules, double negation, associativity, and De Morgan's laws. Often neglected are the distributive rules, but these are actually quite important properties. Generalizing, we say that a lattice $(P, \leq, \wedge, \vee)$ is *distributive* if the following *distributive properties* hold:

$$\begin{aligned} &\forall x, y, z \in P, \quad x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z), \\ &\forall x, y, z \in P, \quad x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z). \end{aligned}$$

One can check that the lattices $N_5$ and $M_3$ are not distributive. In fact, the distributive laws fail if and only if either $N_5$ or $M_3$ is a "sublattice". (We will neither define "sublattice" nor prove this claim. However, the motivated reader is welcome to try to do so.) Distributive lattices possess an important uniqueness condition.

**Lemma 36.9.** *If $(P, \leq, \wedge, \vee, 0, 1)$ is a bounded distributive lattice, then any complement of $x \in P$ is unique.*

*Proof.* Let $y_1, y_2 \in P$ be any two complements for $x$. Then we find

$$y_1 = 1 \wedge y_1 = (y_2 \vee x) \wedge y_1 = (y_2 \wedge y_1) \vee (x \wedge y_1) = (y_2 \wedge y_1) \vee 0 = y_2 \wedge y_1.$$

By a symmetric argument, we also have $y_2 = y_2 \wedge y_1$. □

Thus, in a complemented distributive lattice, there is a unary function that sends each element to its unique complement. In that case, authors may write the complement of $x$ either using the notation $\overline{x}$ or $\neg x$, although other notations, such as $x^{\perp}$, may be used. Of course, the *complement properties*

$$\forall x \in P, \; x \wedge \neg x = 0 \text{ and } x \vee \neg x = 1$$

must hold. Conversely, assuming these properties hold, then they force $\neg x$ to be some complement of $x$. So, if the lattice is bounded and distributive, then $\neg$ must be the unique complement function.

We can now characterize complemented distributive lattices using the properties we have previously listed.

**Theorem 36.10.** *Let $M$ be an arbitrary set, equipped with two binary functions $\wedge, \vee \colon M \times M \to M$, one unary function $\neg \colon M \to M$, and two zeroary functions $0, 1 \colon M^0 \to M$ (i.e., distinguished elements of $M$). Then $(M, \wedge, \vee, \neg, 0, 1)$ is a complemented distributive lattice if and only if the idempotent, commutative, associative, compatibility, identity, distributive, and complement properties hold.*

*Proof.* Combine Definition 35.12 with the work done in this subsection. □

Complemented distributive lattices are sometimes called *Boolean algebras*, due to Boole's early work in describing an algebraic system that represented first order logic.

Of course, $(\mathscr{P}(S), \subseteq, \cap, \cup, ^{-}, \emptyset, S)$ is the prototypical (complete) Boolean algebra, as is any subset $M \subseteq \mathscr{P}(S)$ closed under those functions. Theorem 36.10 gives a large number of properties that intersections, unions, and complements must satisfy; but there are many others that hold in $\mathscr{P}(S)$. For instance, we have double negation and the De Morgan laws. Surprisingly, these properties, and any others, will follow automatically from the properties we have already listed. Can you see a path to proving this result?

(Hint: If you've seen the proof of Cayley's theorem for groups, try to generalize. Alternatively, try the case for finite Boolean algebras first.)

## 36.D Exercises

**Exercise 36.1.** Prove the following fact that was used implicitly in the proof of Theorem 36.3: Let $(P, \leq)$ be a complete lattice, let $Q \subseteq P$, and assume that for any subset $C \subseteq Q$ we have $\bigwedge_P C \in Q$. Then $\bigwedge_Q C$ exists in the poset $(Q, \leq|_Q)$ and equals $\bigwedge_P C$.

**Exercise 36.2.** Let $(P, \leq)$ and $(Q, \leq)$ be posets, and let $f\colon P \to Q$ be a function. If $f$ respects (pairwise) meets, is it monotone? How about the converse? Justify your answers.

**Exercise 36.3.** Let $(P, \leq)$ be a complete lattice and let $f\colon P \to P$ be a monotone function. Prove that the set of fixed points is a complete lattice under the relation $\leq$ (when restricted to that set of fixed points).

**Exercise 36.4.** Prove that the map $\alpha \mapsto \aleph_\alpha$, on ordinals, is monotone. Thus by the Knaster-Tarski theorem (modified to work for classes) it has a fixed point. The smallest fixed point is $\kappa = \mathrm{lub}_{\mathrm{Ord}}\{0, \aleph_0, \aleph_{\aleph_0}, \ldots\}$, so $\kappa$ is the $\kappa$th infinite cardinal.

**Exercise 36.5.** Prove that the map $\alpha \mapsto \omega^\alpha$, on ordinals, is monotone. The smallest ordinal fixed by this map is

$$\epsilon_0 = \mathrm{lub}_{\mathrm{Ord}}\{0, 1, \omega, \omega^\omega, \omega^{\omega^\omega}, \ldots\}.$$

Prove that $\epsilon_0$ is countable. (Can you find the Cantor normal form for $\epsilon_0$?)

**Exercise 36.6.** Let $(P, \leq, \wedge, \vee, 0, 1)$ be a bounded lattice. Prove that 1 is the unique complement of 0, and vice versa.

**Exercise 36.7.** Justify each of the equalities written in Lemma 36.9. Also, explicitly write out the complete "symmetric argument".

**Exercise 36.8.** Huntington, in 1933, proved that a complemented distributive lattice can be characterized, much more simply, by giving a set $M$ together with one binary function $\wedge\colon M \times M \to M$ and one unary function $\neg\colon M \to M$, satisfying the commutative property (just for meets), the associative property (just for meets), and the Huntington equality

$$\forall x, y \in M,\ \neg(\neg x \wedge y) \wedge \neg(\neg x \wedge \neg y) = x.$$

Prove that Huntington's equality holds in complemented distributive lattices.

(Showing, conversely, that the other properties come from these three, is more difficult. The function $\vee \colon M \times M \to M$ could be defined by the rule $(x, y) \mapsto \neg(\neg x \wedge \neg y)$. One would need to show that $x \wedge \neg x = y \wedge \neg y$ for any $x, y \in M$. Then we could call this common element 0. One could then define $1 = \neg 0$. Finally, one would show that all of the other properties for complemented distributive lattices hold.

It is also interesting to note that McCune, Veroff, Fitelson, Harris, Feist, and Wos jointly showed that Boolean algebras can be axiomatized by the single axiom:

$$\forall x, y, z, u \in M,\ \neg(\neg(\neg(x \vee y) \vee z) \vee \neg(x \vee \neg(\neg z \vee \neg(z \vee u)))) = z.$$

Their proof used significant help from computer searches.)

# Index