

4 Theories

In this chapter, we take a look at first-order theories: sets of sentences in a formal, first-order language that are assumed to describe a particular domain of objects. A theory might describe the behaviour of physical systems or moral norms, but we'll focus on mathematical theories – specifically, arithmetic and set theory.

4.1 Arithmetic

For most of history, people did maths in an informal manner, relying on a loose collection of techniques for solving specific types of problems. When giving proofs, assumptions that seemed obviously true – for example, that $0 \neq 1$ – were simply taken for granted.

In the 19th and early 20th century, mathematics was put on a more rigorous footing. Cauchy, Weierstrass, Dedekind, and others gave precise definitions of mathematical concepts (such as limits and continuity). They also formalized the exact assumptions that were needed to derive well-known results. These assumptions were collected into *axioms* for the relevant area of maths.

At the same time, more powerful mathematical theories were developed, such as the set theory of Cantor (formalized by Zermelo, Fraenkel, and others) or the type theory of Russell and Whitehead. All known branches of maths, it seemed, could be unified in such a theory, allowing for new results to be derived from the emerging connections between previously separate domains. Theorems from topology could be used to prove results in algebra.

Formally, a *theory* is a set of sentences that is closed under entailment, so that it contains everything that is entailed by it. In this chapter, we'll be concerned with theories in a formal, first-order language. We often write $\vdash_T A$ or $T \vdash A$ (rather than $A \in T$) to say that a sentence A is a member of the theory T .

This fits our earlier use of the turnstile: If A is in T , then $T \vdash A$ (by Mon and Id); conversely, if $T \vdash A$, then by the completeness of first-order logic, $T \models A$, and then A is in T because T is closed under entailment. We could write $T \models A$ instead of $T \vdash A$. Conceptually, however, theories belong to the “syntax” or “proof theory” side of logic.

A theory is simply a set of sentences. This set is usually specified by laying down some non-logical axioms. The theory then contains all and only the sentences that can be derived from these axioms. We say that a theory T is *axiomatized* by a set of sentences Γ if it contains exactly the sentences that are derivable from Γ .

Exercise 4.1 Let $\mathcal{L}$ be some first-order language (with identity). Let T_1 be the theory axiomatized by the set of all $\mathcal{L}$ -sentences, T_2 the theory axiomatized by the empty set of sentences, and T_3 the theory axiomatized by $\{\forall x x \neq x\}$. Which of T_1 , T_2 , and T_3 are the same?

Let's take a closer look at formal theories of arithmetic. Arithmetic is the study of the natural numbers 0, 1, 2, 3, etc. We know a lot about the natural numbers. We know, for example, that $1 + 2 = 3$, that there are infinitely many primes, or that the factorial function $x!$ grows faster than any polynomial x^n . The aim of an axiomatized formal theory of arithmetic is to capture all such truths, showing exactly which assumptions (or axioms) are needed to derive which results.

An important part of the axiomatic project is to reduce the number of primitive concepts. In section 3.4, I already mentioned that we don't need separate individual constants for each number: we can instead use a single constant '0' for the number 0 and a function symbol 's' for the successor function; the number 1 is then denoted by 's(0)', the number 2 by 's(s(0))', and so on. This is useful because it means that we don't need special axioms for each number: having defined 1, 2, and 3 as $s(0)$, $s(s(0))$, and $s(s(s(0)))$, respectively, we may hope to derive that $1 + 2 = 3$ from general assumptions about zero and the successor function. If '1', '2', and '3' were primitive symbols, it is hard to see how ' $1 + 2 = 3$ ' could be derived from more basic principles.

In section 3.4, I suggested that a first-order theory of arithmetic might use primitive symbols for 0, the successor function, addition, multiplication, and the less-than relation. In fact, the less-than relation can be defined in terms of the other concepts and logical expressions, since the following holds for all natural numbers x and y :

$$x < y \text{ iff } \exists z(x + s(z) = y).$$

We can therefore treat ' $t_1 < t_2$ ', for any terms t_1 and t_2 , as a metalinguistic abbreviation of ' $\exists x(t_1 + s(x) = t_2)$ '. (The variable x must not occur in t_1 or t_2 .)

Less obviously, we can define the concept of a prime number. Remember that a number is prime if it is greater than 1 and divisible only by 1 and itself. A number x is divisible by a number y if there is a number z such that $z \times y = x$. Thus we can express

‘ t is prime’ as:

$$s(0) < t \wedge \forall y(\exists z(z \times y = t) \rightarrow (y = s(0) \vee y = t)).$$

Exercise 4.2 Define the concepts of (a) an even number and (b) a square number.

Other concepts are harder to define. It is not obvious how one could define exponentiation x^y or the factorial $x!$ in terms of 0, s , $+$, and $\times$. We’ll see in chapter 8 how it can be done. Indeed, we’ll see that all computable functions and relations on the natural numbers can be defined in terms of our four primitives. That is, whenever there is an algorithm for computing a function, or for determining whether a relation holds between some numbers, then the function or relation can be defined in terms of 0, s , $+$, and $\times$.

Exercise 4.3 Can you find another primitive that we could use instead of ‘ s ’? (That is, can you find a primitive symbol φ so that $s(t)$ can be defined from 0, φ , $+$, and $\times$?)

Let’s turn to the second part of the axiomatic project. Having reduced the set of primitive concepts, we need to lay down axioms that describe how the remaining concepts behave. The aim is to reduce all truths about the natural numbers to a small number of basic principles.

The first axioms we’ll consider are just about 0 and s . Later, we’ll add axioms for $+$ and $\times$. What do we know about 0 and s ? We know, for example, that every number has a successor. But we don’t need to postulate this as an axiom: all function symbols in first-order logic denote total functions. What isn’t guaranteed is that ‘ s ’ denotes an injective function: we need to postulate that no two numbers have the same successor.

$$Q1 \quad \forall x \forall y (s(x) = s(y) \rightarrow x = y)$$

We also know that 0 is not the successor of any number:

$$Q2 \quad \forall x 0 \neq s(x)$$

These two axioms are already quite powerful. Let’s think about what a model of them must look like. There must be at least one object, denoted by 0. There must also be an object $s(0)$. Can this be the same as 0? No: otherwise 0 would be the successor of itself,

which contradicts Q2. So $s(0)$ is another object. What about $s(s(0))$? This can't be 0, by Q2. And so it can't be $s(0)$ either, by Q1: if $s(s(0)) = s(0)$, then $s(0)$ and 0 would have the same successor. So $s(s(0))$ is a third object. By iterating this reasoning, we can see that any model of Q1 and Q2 must have a chain of infinitely many objects

$$0, s(0), s(s(0)), s(s(s(0))), \dots,$$

connected by the successor function.

Exercise 4.4 Can you find a model in which Q1 and Q2 are true, but $\forall x(s(x) \neq x)$ is false?

Exercise 4.4 shows that Q1 and Q2 don't suffice to capture all truths about 0 and s . The problem is that the two axioms don't rule out the existence of other objects, outside the chain $0, s(0), s(s(0)), \dots$. On these other objects, the successor relation must still be injective, but it can go in a loop, or it can form a second infinite chain $a, s(a), s(s(a)), \dots$. The following axiom rules out such additional chains, by stipulating that there is no object other than 0 that is not a successor.

$$\text{Q3} \quad \forall x (x \neq 0 \rightarrow \exists y x = s(y))$$

This doesn't help with the looping case, however. We'd like to have an axiom saying that every number can eventually be reached from 0 by repeated application of s . But there's no way to express this in first-order logic (as we proved in section 3.4). Still, we can get close by adding the following axiom schema, called the *induction schema*:

$$\text{Ind} \quad (A(0) \wedge \forall x (A(x) \rightarrow A(s(x)))) \rightarrow \forall x A(x)$$

Here, $A(x)$ is any formula with one free variable. Think of every such formula as expressing a property. Ind then says that if some (expressible) property holds of 0, and if it is inherited from any number to its successor, then it holds of all numbers. The schema is obviously related to the method of inductive proof, where we show that all numbers have a property by showing that 0 has it and that it is inherited from any number to its successor.

Ind rules out the looping case. Consider the simplest version, where there's an object a outside the chain $0, s(0), s(s(0)), \dots$ that is its own successor. In this model, $\forall x(s(x) \neq x)$ is false. But $\forall x(s(x) \neq x)$ follows from Q1, Q2, and Ind, as follows.

Let $A(x)$ be the formula $s(x) \neq x$. Then $A(0)$ is $s(0) \neq 0$. This is entailed by Q2. $\forall x(A(x) \rightarrow A(s(x)))$ is $\forall x(s(x) \neq x \rightarrow s(s(x)) \neq s(x))$. This is entailed by Q1. By Ind, we can derive $\forall x(s(x) \neq x)$.

Exercise 4.5 How does Ind rule out loops with two elements? That is, why isn't there a model of Q1, Q2, and Ind with two objects a and b outside the chain $0, s(0), s(s(0)), \dots$ that are successors of each other?

Ind also rules out models with a second chain $a, s(a), s(s(a)), \dots$. We can see this from the fact that it entails Q3:

Proposition 4.1

Ind entails Q3.

Proof. let $A(x)$ be the formula $x \neq 0 \rightarrow \exists y x = s(y)$. Q3 is $\forall x A(x)$. To derive this via Ind, we need to derive

- (i) $A(0)$, and
- (ii) $\forall x(A(x) \rightarrow A(s(x)))$.

Both of these are valid (and therefore provable) in pure first-order logic. (i) holds because $\models 0 = 0$; so the antecedent of $A(0)$ is false and $A(0)$ is true. For (ii), note that the consequent of $A(s(x))$ is $\exists y(s(x) = s(y))$, which is trivial; so $A(s(x))$ can never be false; so $A(x) \rightarrow A(s(x))$ is always true. $\square$

Let's turn to addition and multiplication. A common way to define a function on the natural numbers is to describe how it applies to 0 and then define its value for any successor number in terms of its value for the previous number. For example, the factorial function $n!$ that maps every number n to the product $1 \times 2 \times \dots \times n$ can be defined by the following two clauses:

- (i) $0! = 1$
- (ii) $s(n)! = n! \times s(n)$

This is called a definition by (*primitive*) *recursion*. It may at first look circular, but it is not. Take, for example, the input 2 to the factorial function. By clause (ii) of the

definition, $2!$ is $1! \times 2$. To evaluate this, we need to know $1!$. By clause (ii) again, $1!$ is $0! \times 1$. By the first clause, $0!$ is 1. Putting all this together, we have

$$2! = (1 \times 1) \times 2 = 2.$$

We can similarly define the addition function by primitive recursion on its second argument:

- (i) $x + 0 = x$
- (ii) $x + s(y) = s(x + y)$

These two claims are easily translated into the language of arithmetic, which gives us our next two axioms:

- Q4 $\forall x(x + 0 = x)$
- Q5 $\forall x \forall y(x + s(y) = s(x + y))$

The same trick works for multiplication, which we can define as repeated addition:

- Q6 $\forall x(x \times 0 = 0)$
- Q7 $\forall x \forall y(x \times s(y) = (x \times y) + x)$

Exercise 4.6 Explain how the primitive recursive definition of addition determines the value of $3 + 2$.

The theory axiomatized by Q1–Q7 is called *Robinson Arithmetic*, or Q. It will play an important role in chapter ???. The standard first-order theory of arithmetic, called *Peano Arithmetic*, or PA, replaces Q3 by Ind: its axioms are Q1, Q2, Ind, and Q4–Q7. (The theory is named after Giuseppe Peano, although Peano points out that essentially the same theory was proposed earlier by Dedekind).

Are all truths in the language of arithmetic entailed by the axioms of PA? For a while, this seemed plausible. Gödel’s first *incompleteness* theorem revealed that the answer is no: there are arithmetical truths that aren’t provable in PA. So $PA \neq Th(\mathfrak{N})$. We’ll prove this in ch. ??. As we’ll see, the problem can’t be fixed by adding a few more axioms or axiom schemas. PA isn’t just incomplete; there’s a good sense in which it is *incompletable*.

Exercise 4.7 Show that the following are in PA:

- (a) $\forall x x < s(x)$; (b) $\forall x \forall y (x < y \rightarrow 0 < y)$; (c) $\forall x \forall y (x + y = y + x)$.

Exercise 4.8 We've seen that Q1–Q3 don't rule out structures in which the successor function goes in a loop for some objects outside $0, s(0), s(s(0)), \dots$

- (a) Show that adding Q4–Q7 doesn't help: define a model $\mathfrak{M}$ of Q1–Q7 with two objects a and b that are successors of each other.
- (b) Using the definition of ' $<$ ' from earlier in this section, determine whether $a < b$, $a < a$, and $0 < a$ are true in your model $\mathfrak{M}$.
- (c) Is $\forall x \forall y (x + y = y + x)$ true in your model $\mathfrak{M}$? If yes, change the interpretation of $+$ to make it false while keeping Q1–Q7 true.

4.2 Set theory

In the 19th century, set-theoretic concepts were increasingly used by mathematicians to make their theories and definitions more precise. For example, Dedekind defined the real numbers in terms of sets of rational numbers, which allowed for new, more rigorous proofs of many results in real analysis.

The concept of a set was initially not seen as belonging to a separate mathematical theory (*set theory*). Rather, it was treated as a logical concept. To speak of the set of such-and-suchs, it was assumed, is just to speak of the such-and-suchs taken together. As Georg Cantor put it in 1895: a set is 'a collection of definite, well-differentiated objects [...] into a whole'. It was assumed that, as a matter of logic, whenever there are some (definite, well-differentiated) objects, there is also a set of these objects.

Dedekind had defined the real numbers in terms of sets of rational numbers. The rational numbers can, in turn be defined in terms of sets and integers, and the integers in terms of sets and natural numbers. Frege realized that one can define the natural numbers entirely in terms of sets. (See section 4.3 below for one way to do this.) Familiar properties of the natural numbers – and, by extension, of the integers, rationals, and reals – can then be derived from apparently logical properties of sets. Hence there emerged the philosophical project of *logicism*: the idea that all of maths could be reduced to logic and definitions.

This was the life project of Frege, who invented the calculus of predicate logic in order to show that all of arithmetic could be derived from purely logical axioms by simple

logical rules like MP and Gen. Frege’s “logical axioms” included one assumption about sets – his “axiom V”. This is a second-order axiom involving the term-forming operator $\{x : A(x)\}$. We can express it as a first-order schema:

$$\forall \{x : A(x)\} = \{x : B(x)\} \leftrightarrow \forall x(A(x) \leftrightarrow B(x)).$$

$A(x)$ and $B(x)$ are arbitrary formulas with one free variable. Axiom V says that different sets never have the very same members. This makes sense if a set of things is just those things “considered as a whole”. But the use of the set operator $\{x : A(x)\}$ in an otherwise standard first-order language also implies that for any formula $A(x)$ there is a corresponding set $\{x : A(x)\}$. This is known as the *naive comprehension principle*.

Unfortunately for Frege, the naive comprehension principle is inconsistent, as Bertrand Russell pointed out to him in a letter in 1902. Consider the formula $x \notin x$, saying that x is not a member of itself. Assume that there is a set of all things to which this formula applies. Call this set R . Is R a member of itself? If it is, then by the definition of R , it is not a member of itself. If it isn’t, then by the definition of R , it is a member of itself. This is a contradiction. So $x \notin x$ is a formula for which there is no corresponding set $\{x : x \notin x\}$.

There is something odd about the idea that a set might contain itself. One imagines sets as abstract “containers”, and a container can hardly contain itself. Ernst Zermelo, who had independently noticed Russell’s paradox, developed this intuition into a paradox-free formal theory.

According to Zermelo, we should think of the sets as built in layers or stages. We start with things that are not sets, called *individuals* or *urelements*. At the next stage, we form all sets of these individuals. We may now have sets of rocks and cities, like $\{\text{Athens, Berlin}\}$, but we don’t have any sets containing other sets. At the next stage, we form all sets whose elements are either individuals or sets of individuals. This includes all sets from the first stage, but it also includes sets like $\{\{\text{Athens, Berlin}\}, \text{Athens}\}$, with sets from the previous stage as elements. We continue in this manner. Whenever a set occurs at some stage, it can be used as an element of sets at later stages. But not otherwise: a set can only appear at a stage after all its elements have appeared. So we never get a set that contains itself. Nor do we get a set of all sets that don’t contain themselves: this would be the set of all sets; such a set would contain itself, which is impossible.

Oddly, this hierarchical construction works even if there are no individuals. Starting with no individuals, we can construct one set of individuals: the empty set $\emptyset$. From this, we can form another set: $\{\emptyset\}$. And once we have $\emptyset$ and $\{\emptyset\}$, we can form $\{\emptyset, \{\emptyset\}\}$ and $\{\{\emptyset\}\}$. And off we go. For purely mathematical applications, it turns out that this *pure* hierarchy is often enough.

Let's make the structure of the set-theoretic hierarchy, called the *cumulative hierarchy*, or simply V , more precise. Each stage of the hierarchy is a set of sets. The first stage, V_0 , is the set of individuals. In the pure hierarchy, this is the empty set:

$$V_0 = \emptyset.$$

From any stage V_k , we recursively define the next stage V_{k+1} as the set of all sets whose elements are in V_k . This is just the power set of V_k :

$$V_{k+1} = \mathcal{P}(V_k).$$

In the pure hierarchy, $V_1 = \mathcal{P}(\emptyset) = \{\emptyset\}$, $V_2 = \mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$, and so on. This yields an infinite sequence $V_0, V_1, V_2, \dots$ of ever-larger sets, all ultimately built from the empty set.

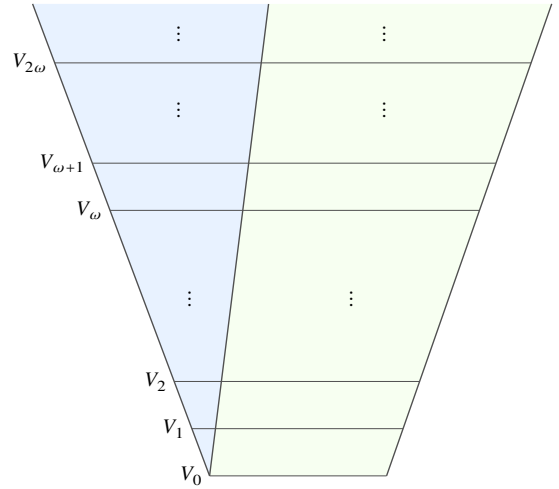
But we don't stop there. After all the stages $V_0, V_1, V_2, \dots$, there is another stage V_ω ("V omega"). V_ω contains all sets that have appeared at any earlier stage. That is, V_ω is the union of all earlier stages:

$$V_\omega = \bigcup_{k < \omega} V_k.$$

While all sets in the sequence $V_0, V_1, V_2, \dots$ are finite, V_ω has infinitely many elements.

From V_ω , we can form yet further sets by repeating the previous recipes. At stage $V_{\omega+1}$, we collect all the subsets of V_ω . (Many of these are infinite and thus didn't appear at any earlier stage.) That is, $V_{\omega+1} = \mathcal{P}(V_\omega)$. We then form $V_{\omega+2} = \mathcal{P}(V_{\omega+1})$, and so on. After all the stages $V_\omega, V_{\omega+1}, V_{\omega+2}, \dots$, there is another stage $V_{\omega+\omega}$, or $V_{\omega \cdot 2}$. It contains all sets that have appeared at any earlier stage: $V_{\omega \cdot 2} = \bigcup_{k < \omega \cdot 2} V_k$. From $V_{\omega \cdot 2}$, we construct $V_{\omega \cdot 2+1}$, $V_{\omega \cdot 2+2}$, etc. by taking power sets. Then we construct $V_{\omega \cdot 3}$ by taking the union of all earlier stages. And so on and on.

And we don't stop there. After all the stages $V_\omega, \dots, V_{\omega \cdot 2}, \dots, V_{\omega \cdot 3}, \dots$, there is another stage $V_{\omega \cdot \omega}$, or V_{ω^2} , where we take the union of all previous stages. From this, we construct further stages by taking power sets and unions. Eventually, we reach V_{ω^3} , then V_{ω^4} , etc. Then we take the union of all these stages to get V_{ω^ω} , and so on and on,



through $V_{\omega^{\omega^{\omega}}}$, through stages with infinitely high towers of ω , and much, much further. The cumulative hierarchy is *vast*.

Exercise 4.9 Consider the pure hierarchy. How many sets are in V_3 ? How many are in V_4 ?

Exercise 4.10 Is the cardinality of $V_{\omega+1}$ greater than the cardinality of V_ω ?

Let's now try to axiomatize this conception of sets. That is, we'll try to find a set of sentences in a suitable first-order language that describes the structure of the cumulative hierarchy. The description I just gave, with its 'and so on's and 'after all these stages' can't be directly translated into first-order logic. We have to take a more indirect approach.

The most popular axiomatization of set theory is *ZFC*, for 'Zermelo-Fraenkel set theory with the Axiom of Choice'. Its only primitive concept is the membership relation. So we have a single non-logical symbol: the binary predicate symbol ' $\in$ '. From this, other concepts are defined. For example, we can define the subset relation $\subseteq$ as follows:

$$t_1 \subseteq t_2 \text{ abbreviates } \forall x(x \in t_1 \rightarrow x \in t_2).$$

Let's go through the axioms of ZFC. The quantifiers are assumed to range over the pure sets. Our first axiom is known as the axiom of *extensionality*.

$$\text{Z1 } \forall x \forall y ((\forall z(z \in x \leftrightarrow z \in y)) \rightarrow x = y).$$

This says that a set is determined by its elements: no two sets have the same elements. Unlike Frege's Axiom V, Z1 doesn't imply that for any formula $A(x)$ there is a corresponding set $\{x : A(x)\}$. Instead of this unrestricted comprehension principle, we have a more restricted principle, called the *separation axiom*. It's actually a schema:

$$\text{Z2 } \forall y \exists z \forall x (x \in z \leftrightarrow (x \in y \wedge A(x)))$$

This says that for any set y and any formula $A(x)$, there is a set z that contains just those elements of y of which $A(x)$ is true. That is, provided that we already have a set y , we can use any formula to carve out a subset of y containing those elements of y of which the formula is true.

The next axiom postulates the existence of the empty set, the base level of the hierarchy.

$$Z3 \quad \exists x \forall y (y \notin x).$$

This says that there is something (a set) that has no elements. By the extensionality axiom, there is only one such thing. It's convenient to have a name for it: ' $\emptyset$ '. But ' $\emptyset$ ' isn't officially part of the language. The only singular terms in the language of set theory are variables. So we can't say that ' $\emptyset$ ' is shorthand for some more complex term in the language, in the way we could treat ' 3 ' as shorthand for ' $s(s(s(0)))$ '. What we can do instead is give a *contextual* or *syncategorematic* definition of ' $\emptyset$ ', as follows:

$$A(\emptyset) \text{ abbreviates } \exists x (\forall y (y \notin x \wedge A(x))).$$

Here, $A(x)$ is an expression with one free variable x , and $A(\emptyset)$ is that expression with ' $\emptyset$ ' in place of x . For example, consider the expression

$$\forall x (\emptyset \subseteq x).$$

By the convention for $\emptyset$, it is shorthand for

$$\exists z (\forall y (y \notin z) \wedge \forall x (z \subseteq x)).$$

By the convention for $\subseteq$, this is in turn shorthand for

$$\exists z (\forall y (y \notin z) \wedge \forall x \forall v (v \in z \rightarrow v \in x)).$$

The same trick is needed to talk about operations on sets. To define the union operation $\cup$, for example, we need to find a formula that is true of sets x, y , and z iff z is the union of sets x and y . Such a formula is not hard to find:

$$\forall v (v \in x \vee v \in y \leftrightarrow v \in z).$$

With this, we can give a contextual definition of ' $\cup$ ':

$$A(t_1 \cup t_2) \text{ abbreviates } \exists x (\forall y (y \in t_1 \vee y \in t_2 \leftrightarrow y \in x) \wedge A(x)),$$

where x and y do not occur in A .

We can similarly define the intersection operation $\cap$:

$$A(t_1 \cap t_2) \text{ abbreviates } \exists x (\forall y (y \in t_1 \wedge y \in t_2 \leftrightarrow y \in x) \wedge A(x)).$$

Exercise 4.11 Give contextual definitions of $\bigcup t$ and $\mathcal{P}(t)$. $\bigcup t$ is the union of all sets in t ; $\mathcal{P}(t)$ is the set of all subsets of t .

The next two axioms guarantee that for every set x , there is a set $\bigcup x$ comprising all elements of elements of x , and a set $\mathcal{P}(x)$ comprising all subsets of x . Z4 is the *union axiom*, Z5 the *powerset axiom*.

$$\text{Z4} \quad \forall x \exists u \forall y (y \in u \leftrightarrow \exists z (z \in x \wedge y \in z)).$$

$$\text{Z5} \quad \forall x \exists p \forall y (y \in p \leftrightarrow y \subseteq x).$$

Next, we have the *pairing axiom*:

$$\text{Z6} \quad \forall x \forall y \exists z \forall v (v \in z \leftrightarrow (v = x \vee v = y))$$

This says that for any sets x, y there is a set $\{x, y\}$ that contains exactly x and y . This is needed, for example, to ensure that $x \cup y$ exists whenever x and y exist: the pairing axiom gives us $\{x, y\}$, from which we get $x \cup y = \bigcup \{x, y\}$ by Z4.

The sets x and y in Z6 needn't be different. For the case where $x = y$, the axiom says that for every set x there is a set $\{x, x\} = \{x\}$ that contains exactly x . This is called the *singleton* set of x . We'll help ourselves to $\{t\}$ as a contextually defined term:

$$A(\{t\}) \text{ abbreviates } \exists x (\forall y (y \in x \leftrightarrow y = t) \wedge A(x)).$$

We make use of this abbreviation in our next axiom, the *axiom of infinity*:

$$\text{Z7} \quad \exists x (\emptyset \in x \wedge \forall y (y \in x \rightarrow y \cup \{y\} \in x)).$$

Without the axiom of infinity, we couldn't guarantee the existence of any infinite set. In the next section, we'll see that the set x whose existence is guaranteed by Z7 can be understood as the set $\mathbb{N}$ of natural numbers.

Exercise 4.12 List three members of the set whose existence is guaranteed by Z7.

Next, we have the axiom of *foundation* (or *regularity*). It ensures that every set (every object in the domain) is part of the cumulative hierarchy. Consider any nonempty set x . The elements of x are other sets. If x is in the cumulative hierarchy, then its elements must

have appeared at earlier stages in the construction, and there must be some stage at which the first of them appeared. Let y be one of these earliest elements. Since all elements of y appear strictly before y , it follows that none of the elements of y are elements of x . That is, every nonempty set x of sets must have an element y that is disjoint from x :

$$\text{Z8} \quad \forall x(x \neq \emptyset \rightarrow \exists y(y \in x \wedge x \cap y = \emptyset))$$

There are two more axioms. Next is the *axiom of replacement*, due to Abraham Fraenkel. It is motivated by the observation that the naive comprehension principle only seems to go wrong in cases where the formula $A(x)$ from which it allows defining a set $\{x : A(x)\}$ is true of every set, or of things of which there are as many as there are sets. For example, Russell's $x \notin x$ is true of all the sets. Objects of which there are as many as there are sets are sometimes said to form a *proper class*. (This concept is formalized in some extensions of ZFC, such as the von Neumann-Bernays-Gödel set theory NBG.) In essence, the axiom of replacement says that if there are no more objects of a certain kind than there are members of some set x , then these objects also form a set. After all, they can't form a proper class if there are no more of them than there are members of x , which is known to be a set.

To see how this may be used, suppose that we have a construction that defines a set s_n for each natural number. So there are as many sets $s_0, s_1, s_2, \dots$ as there are natural numbers. If we identify the natural numbers with the elements of the set whose existence is guaranteed by Z7, we know that the natural numbers form a set. The replacement axiom allows us to conclude that the $s_0, s_1, s_2, \dots$ also form a set. It is called 'replacement' because it allows replacing all members i of a known set by other things $f(i)$.

Let's review how we compare the sizes of infinite collections. By the standards of section 3.1, a set x is no larger than a set y iff there is an injective function from x to y : a function that maps each element of x to an element of y , without mapping different elements of x to the same element of y . In set theory, we don't have functions as separate objects. But we can simulate them by sets. Since functions are fully determined by which outputs they return for which inputs, we can identify them with sets of input-output pairs. For example, the square function on the natural numbers would be identified with the set of ordered pairs $\langle 0, 0 \rangle, \langle 1, 1 \rangle, \langle 2, 4 \rangle, \langle 3, 9 \rangle$, etc.

To complete this definition, we need to provide a set-theoretic surrogate for the concept of an ordered pair. An ordered pair $\langle x, y \rangle$ isn't simply the set $\{x, y\}$: we want to distinguish $\langle 2, 4 \rangle$ from $\langle 4, 2 \rangle$. In general, ordered pairs $\langle x_1, y_1 \rangle$ and $\langle x_2, y_2 \rangle$ are identical iff $x_1 = x_2$ and $y_1 = y_2$. Can we find set-theoretic constructs that satisfy this condition? Easy. The standard construction, due to Kazimierz Kuratowski, identifies $\langle x, y \rangle$ with the set

$\{\{x\}, \{x, y\}\}$. You can easily show that, on this definition, $\langle x_1, y_1 \rangle = \langle x_2, y_2 \rangle$ iff $x_1 = x_2$ and $y_1 = y_2$.

Exercise 4.13 Find another set-theoretic construction of $\langle x, y \rangle$ that satisfies the identity condition for ordered pairs.

Return to the axiom of replacement. We want to use the axiom show that certain objects form a set, given that there are no more of them than there are members of some other set. Unfortunately, we can't assume in this context we have already established the existence of any functions from the objects to the other set. (If we knew that there is a set of ordered pairs $\langle x, y \rangle$ in which each of our objects figures as a first member, we could infer that the objects form a set by the axioms of union and separation: we wouldn't need replacement).

Instead of invoking functions, the replacement axiom therefore uses *formulas* to express that there is a functional relationship. First, a convenient abbreviation:

$\exists! x A(x)$ abbreviates $\exists x(A(x) \wedge \forall y(A(y) \rightarrow y=x))$,

$\exists! x A(x)$ says that there is exactly one x such that $A(x)$. So $\forall x \exists! y A(x, y)$ says that $A(x, y)$ expresses a functional relationship: it relates each x to exactly one y . If there is such a functional relationship between the members of some set v and some y s, there can be no more y s than there are members of v . The axiom of replacement, which is really an axiom schema, allows us to conclude that there is a set w that contains all these y s:

$$Z9 \quad \forall v((\forall x(x \in v \rightarrow \exists! y A(x, y)) \rightarrow \exists w \forall x(x \in v \leftrightarrow \exists y(y \in w \wedge A(x, y))))))$$

Replacement is needed to ensure that the cumulative hierarchy extends beyond the finite stages $V_0, V_1, V_2, \dots$. From Z1–Z8 (sometimes called *Zermelo set theory* or Z), we get the finite stages, but we can't prove the existence of V_ω . With Replacement, we can show that there is a set $\{V_n : n \in \mathbb{N}\}$ of all finite stages: the formula $A(x, y)$ in Z9 says that y is obtained from $\emptyset$ by x applications of the power set operation. V_ω is the union of $\{V_n : n \in \mathbb{N}\}$.

Finally, we have Zermelo's Axiom of Choice. This says that if we have a set x of non-empty sets, then there is a set y that contains exactly one element from each set in x .

$$Z10 \quad \forall x[\forall z(z \in x \rightarrow z \neq \emptyset) \rightarrow \exists y \forall z(z \in x \rightarrow \exists! v(v \in z \wedge v \in y))]$$

Unlike the other axioms, the Axiom of Choice states that a certain set exists without describing how it can be constructed: we are not told *which* element of each set in x is in y . For this reason (as well as certain strange consequences in the theory of measures), the axiom has long been controversial. Nowadays, it is generally accepted, as many important mathematical results depend on it.

Exercise 4.14 The axioms of ZFC guarantee that for any three things a, b, c , there is a set $\{a, b, c\}$. Explain how.

Exercise 4.15 Explain why the separation axiom implies that there is no set of all sets.

4.3 Sets and numbers

The Axiom of Infinity draws attention to an infinite sequence of sets, called the *finite von Neumann ordinals*, or simply the *finite ordinals*:

- $\emptyset$
- $\{\emptyset\}$
- $\{\emptyset, \{\emptyset\}\}$
- $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$
- ...

This sequence has the structure of the natural numbers. We can think of $\emptyset$ as 0, $\{\emptyset\}$ as 1, $\{\emptyset, \{\emptyset\}\}$ as 2, and so on. The successor of any number n is $n \cup \{n\}$. (Note that, conveniently, each number n in this construction has exactly n elements.)

More formally, we can use the finite ordinals to define a model of arithmetical theories like $\text{Th}(\mathfrak{N})$ and PA. Recall that a model of a theory is a structure consisting of a domain and an interpretation of the non-logical symbols in which all sentences in the theory are true. For a model of $\text{Th}(\mathfrak{N})$, we can choose as the domain the set ω of finite ordinals. The interpretation function maps the ‘0’ symbol to $\emptyset$ and the successor symbol ‘s’ to the function that maps each set $x \in \omega$ to $x \cup \{x\}$. The standard recursive definitions of addition and multiplication then determine the interpretation of ‘+’ and ‘ $\times$ ’. (If n and m are in ω , $n + m$ will be the unique set in ω that has exactly $n + m$ elements, and $n \times m$ the unique set with $n \times m$ elements.) This shows that the natural number structure can

be embedded in the structure of sets. The same is true for almost every mathematical structure.

There is more. Suppose we read

- ‘0’ as an abbreviation of ‘ $\emptyset$ ’,
- ‘ $s(t)$ ’ as an abbreviation of ‘ $t \cup \{t\}$ ’,
- ‘ $t_1 + t_2$ ’ and ‘ $t_1 \times t_2$ ’ as abbreviations of the corresponding operations on sets,

and we restrict all quantifiers in PA to range over ω , so that ‘ $\forall xA$ ’ becomes ‘ $\forall x(x \in \omega \rightarrow A)$ ’. All axioms of PA are then provable in ZFC. We say that PA is *interpretable* in ZFC. In general, a theory T is interpretable in ZFC if there is a translation scheme of the kind I’ve sketched under which all sentences in T are provable in ZFC.

A wide range of mathematical theories are interpretable in ZFC. In that sense, ZFC is *at least as strong* as these other theories: whatever they can prove, ZFC can prove as well (if only under the appropriate translation scheme).

I’m not going to prove that PA is interpretable in ZFC. The proof isn’t hard, but a little fiddly. To get a sense of what needs to be shown, consider the second axiom of PA:

$$Q2 \quad \forall x 0 \neq s(x)$$

Under the above translation scheme, this turns into $\forall x(x \in \omega \rightarrow (\emptyset \neq x \cup \{x\}))$. And that’s easily provable in ZFC.

Exercise 4.16 Sketch a proof of the translated Q2 axiom (from the axioms of ZFC).

Let’s now have a closer look at the finite ordinals. They have some interesting properties.

For one, every member of a finite ordinal is also a subset of it. Sets of this kind are called *transitive*. That’s because a transitive set z is a set such that whenever $x \in y$ and $y \in z$ then $x \in z$.

Another special property of the finite ordinals is that they are *linearly ordered by $\in$* : any two members of x are related one way or the other by $\in$. I’ll say, for short, that the finite ordinals are *$\in$ -ordered*.

In ZFC, the finite ordinals can be defined as the transitive and $\in$ -ordered sets with finitely many elements. Now suppose we drop the finiteness condition. Let’s define an *ordinal* as a transitive and $\in$ -ordered set. The finite ordinals are ordinals, but they are

not the only ones. For example, ω , the set of finite ordinals, is itself an ordinal. (As you can confirm, it is transitive and $\in$ -ordered.) ω is an *infinite ordinal*. So is $\omega \cup \{\omega\}$: the set we get from ω by adding ω itself as an element. Following our earlier definition of the successor relation, we can see $\omega \cup \{\omega\}$ as the “successor” of ω . The successor of $\omega \cup \{\omega\}$ is $\omega \cup \{\omega\} \cup \{\omega \cup \{\omega\}\}$, and so on.

The ordinals form a *transfinite* sequence. If we identify the finite ordinals with the natural numbers, the transfinite sequence of ordinals look like this:

$$0, 1, 2, \dots, \omega, \omega + 1, \omega + 2, \dots, \omega + \omega, \dots$$

Like 0, ω is not the successor of any ordinal. Infinite ordinals that are not successors are called *limit ordinals*. The next limit ordinal after ω is $\omega + \omega$, or $\omega \cdot 2$. It is the union of all ordinals $\omega + n$, where n is a finite ordinal. The next limit ordinal after $\omega \cdot 2$ is $\omega \cdot 3$. After all the limit ordinals $\omega \cdot n$ and all their successors comes their union $\omega \cdot \omega$, or ω^2 – another limit ordinal. Much later we reach ω^ω , ω^{ω^ω} , and so on.

The ordinals extend the idea of “counting” beyond the finite. This has many mathematical applications. Above, I’ve used the ordinals to label stages in the cumulative hierarchy. I used limit ordinals to label stages at which we take the union of the earlier stages, and successor ordinals to label stages at which we take power sets.

Exercise 4.17 Show that ω is transitive and $\in$ -ordered.

Exercise 4.18 Show from the axioms of ZFC that every ordinal has a successor.

Exercise 4.19 Is the set of all ordinals an ordinal?

The ordinals can also be used to interpret the theory of cardinals that I outlined in the previous chapter. Remember that two sets have the same cardinality iff there is a bijection between them. For finite sets, cardinalities are naturally identified with natural numbers: {Athens, Berlin, Cairo} has cardinality 3. But what kind of thing is the cardinality of an infinite set? In section 3.1, we gave them names: we called them $\aleph_0$, $\aleph_1$, etc. But I didn’t say more about what these things might be.

The standard answer in contemporary set theory identifies the cardinals with certain ordinals: the cardinality of any set x is defined as *the least ordinal that is equinumerous with x* .

For finite sets, this yields the expected results. $\{\text{Athens, Berlin, Cairo}\}$ is equinumerous with $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$, which is the ordinal 3. So the cardinality of $\{\text{Athens, Berlin, Cairo}\}$ is 3. The cardinality of ω is ω . That's because ω is the least ordinal that is equinumerous with ω . Since ω is countably infinite, and we defined $\aleph_0$ as the cardinality of any countably infinite set, this means that $\omega = \aleph_0$.

Beyond $\aleph_0$, things get interesting. The cardinality of $\omega + 1$ is still $\aleph_0$. Remember that $\omega + 1$ is $\omega \cup \{\omega\}$: it is ω with one extra element. If you add a single element to a countably infinite set, you always get another countably infinite set. So $\omega + 1$ is equinumerous with ω . Since the cardinality of a set is the *least* ordinal equinumerous with it, the cardinality of $\omega + 1$ is ω (a.k.a. $\aleph_0$).

After ω , the *ordinal numbers* and the *cardinal numbers* diverge. ω is both an ordinal and a cardinal. But $\omega + 1$ is only an ordinal. We've introduced ' $\aleph_1$ ' to name the next cardinal after $\aleph_0$. By our current definition, $\aleph_1$ is first ordinal in the transfinite sequence of ordinals that is not equinumerous with ω . It comes surprisingly late. It's not $\omega + 1$, or $\omega \cdot 2$, or ω^2 , or ω^ω , or ω^{ω^ω} . All of these are equinumerous with ω . $\aleph_1$ comes much later. And yet we know, from Cantor's theorem, that there are infinitely many different cardinalities. Indeed, for every ordinal κ , there is a distinct cardinal $\aleph_\kappa$, which is itself an ordinal!

By Cantor's theorem, the cardinality of $\mathcal{P}(\omega)$ is greater than $\aleph_0$. How much greater? Cantor conjectured, but was unable to prove, that the cardinality of $\mathcal{P}(\omega)$ is $\aleph_1$. Since $\mathcal{P}(\omega)$ is equinumerous with the set of real numbers $\mathbb{R}$, which is also known as the *continuum*, Cantor's conjecture was that there is no set whose cardinality is strictly between that of the natural numbers and that of the real numbers. This became known as the *continuum hypothesis*.

In 1938, Gödel proved that the continuum hypothesis is consistent with ZFC (assuming ZFC itself is consistent): it can't be disproved from the axioms of ZFC. In 1963, Paul Cohen showed that the negation of the continuum hypothesis is also consistent with ZFC (assuming ZFC is consistent). So the continuum hypothesis can be neither proved nor disproved in ZFC.

I find this odd. Take the set of real numbers $\mathbb{R}$. We know that this set is uncountable. We can get a countable set by removing sufficiently many elements from $\mathbb{R}$. Can we also remove elements from $\mathbb{R}$ so that we get a set that's still uncountable, but smaller than $\mathbb{R}$? I would expect this simple question to have a definite answer. But it can't be answered from the standard axioms of set theory. We could, of course, add the continuum hypothesis as a further axiom. But we could equally add its negation. Neither leads to a contradiction.

Early set theorists assumed that all questions about pure sets have definite answers that

can be established by an extended kind of logic. The status of the continuum hypothesis casts doubt on this picture. By now, hundreds of other statements are known that can neither be proved nor disproved in ZFC. We can investigate structures in which they hold and structures in which they fail. Perhaps there is no “true” structure of sets after all. When we describe the cumulative hierarchy, we seem to describe a unique structure. We say that $V_{\omega+2}$ contains *all subsets* of $V_{\omega+1}$. But we can’t tell whether these subsets include sets with a cardinality between $\aleph_0$ and the continuum. If the concept of ‘all subsets’ has a definite meaning, this meaning seems impossible to pin down.

4.4 Unintended models, again

In section 4.1, we looked at non-standard models of Q: models in which all axioms of Q are true but whose structure is clearly not that of the natural numbers. I didn’t emphasize it at the time, but Peano Arithmetic also has non-standard models. These are harder to construct directly. But we know that they exist, from the compactness theorem.

Theorem 4.1

There are non-standard models of Peano Arithmetic.

Proof. Let c be an individual constant other than 0. Let Γ be the set of sentences consisting of the axioms of PA together with all the sentences

$$c \neq 0, c \neq s(0), c \neq s(s(0)), \dots$$

Every finite subset of Γ is true in the standard model of arithmetic: just interpret c as a sufficiently large natural number. By the compactness theorem, Γ has a model. All axioms of PA are true in this model. But the object denoted by c (in this model) can’t be a natural number: it lies outside the number sequence 0,1,2,3, etc. $\square$

Intuitively, Peano Arithmetic doesn’t “know” that there are no numbers besides 0,1,2,3, etc.: its axioms are compatible with the existence of further numbers. We know from theorem 3.5 that there’s no way to add the missing information to PA, in the form of further axioms: even the set of all truths in the language of arithmetic, $\text{Th}(\mathcal{N})$, has non-standard models.

Exercise 4.20 PA rules out structures in which the “non-standard numbers” form either a loop or a second chain $a, s(a), s(s(a)), \dots$. What else could a non-standard model look like?

Are there also non-standard models of ZFC? Let’s first clarify the standard model. A model of ZFC consists of a set D of objects and an interpretation function I that assigns some relation on D to the symbol ‘ $\in$ ’. In the *intended* model, D is the set of all sets, and I maps ‘ $\in$ ’ to ... Wait. There is no set of all sets!

In a sense, every model of ZFC is a non-standard model. For every model has a set as its domain, but there is no set of all sets. The real sets form a proper class.

The problem is that we’ve formalized our semantic concepts in set-theoretic terms. We’ve define models as set-theoretic structures. The intended interpretation of ZFC can’t be formalized in this way.

You may wonder how ZFC can have models in our set-theoretic sense at all. In any set-theoretic model of ZFC, the domain is a set, but ZFC entails that there is no set of all sets. We can strengthen this puzzle. Let’s take for granted that ZFC is consistent. By the completeness theorem, it follows that ZFC has a model. By the (downward) Löwenheim-Skolem theorem, it follows from this that ZFC has a *countable* model. Call that model $\mathfrak{M}$. The domain of $\mathfrak{M}$ contains only countably many objects. Yet all sentences in ZFC are true in $\mathfrak{M}$, including sentences saying that there are uncountably many things in $\mathcal{P}(\omega)$, even more in $\mathcal{P}(\mathcal{P}(\omega))$, and so on. This is known as “Skolem’s Paradox”.

It’s not a real paradox. A set x is countable if there is an injective function from x to ω ; x is uncountable if there is no such function. ZFC proves that $\mathcal{P}(\omega)$ is uncountable by proving that there is no injective function from $\mathcal{P}(\omega)$ to ω . Remember that functions are represented as sets of ordered pairs. To say that there is an injective function from x to ω is to say that there is a set f of pairs $\langle y, n \rangle$ such that for each $y \in x$ there is exactly one $n \in \omega$ for which $\langle y, n \rangle \in f$, and for each $n \in \omega$ there is at most one $y \in x$ for which $\langle y, n \rangle \in f$. ZFC proves that there is no such set f for $x = \mathcal{P}(\omega)$. In the countable model $\mathfrak{M}$, there may, in fact, be a bijection between the objects denoted by ω and $\mathcal{P}(\omega)$. That is, we may be able to construct such a bijection. But it need not be an object in the domain of $\mathfrak{M}$. If it is not, the statement that there is no bijection of the given type is *true* in $\mathfrak{M}$.

There is still a puzzle, however. It is related to the puzzle from the end of the previous section. How do we manage to latch onto the set-theoretic universe? We could program an AI to interpret the language of ZFC in a countable model. The AI would *say* that there are uncountably many sets. It would say all the right things. But its conception of

sets would seem to be radically different from ours. For we can see that there are, in fact, only countably many of the things it calls ‘sets’. Given that this is possible, how can we be sure that we are not equally mistaken about the true sets? How do we know that there isn’t an outside perspective from which one can see that there only countably many of the things we call ‘sets’?