Notes on Set-Theory

David Pierce

2004.2.20

0 Introduction

0.1

The book of Landau [11] that influences these notes begins with two prefaces, one for the student and one for the teacher. The first asks the student not to read the second. Perhaps Landau hoped to *induce* the student to read the Preface for the Teacher, but not to worry about digesting its contents. I have such a hope concerning § 0.2 below.

An earlier version of these notes[1] began immediately with a study of the natural numbers. The set-theory in those notes was somewhat *naïve*, that is, non-axiomatic. Of the usual so-called Zermelo–Fraenkel Axioms with Choice, the notes *did* mention the Axioms of Foundation, Infinity and Choice, but not (explicitly) the others. The present notes *do* give all of the axioms[2] of **ZFC**.

What is a set? First of all, a set is many things that can be considered as one; it is a multitude that is also a unity; it is something like a *number*[3]. Therefore, set-theory might be an appropriate part of the education of the guardians of an ideal city—namely, the city that Plato's Socrates describes in the *Republic*. The following translation from Book VII (524d–525b) is mine, but depends on the translations of Shorey [14] and Waterfield [15]. I have inserted some of the original Greek words, especially[4] those that are origins of English words. (See Table I below for transliterations.)

> 'So this is what I [Socrates] was just trying to explain: Some things are thought-provoking, and some are not. Those things are

[1] I prepared the earlier version for the first-year course at METU called 'Fundamentals of Mathematics' (Math 111); but those notes contained much more than that course had time for.

[2] In their order of appearance here, they are: Extensionality (p. 7), Pairing (p. 11), Comprehension (p. 11), Power-set (p. 12), Union (p. 12), Replacement (p. 15), Infinity (p. 42) and Foundation (p. 43).

[3] I may set technical terms in a slanted font thus, by way of acknowledging that they *are* technical terms.

[4] I have also included certain derivatives of the present participle ὄντ- corresponding to the English *being*. Addition of the abstract-noun suffix -ία yields οὐσία; the corresponding Turkish might be `olurluk`. The Greek οὐσία is sometimes translated as *substance*, and indeed both words can connote wealth. Putting the definite article in front of the nominative neuter form of ὄντ- creates τὸ ὄν.

called **thought-provoking** that strike our sense together with their opposites. Those that do not, do not tend to awaken reflection.'

'Ah, now I understand' he [Glaucon] said. 'It seems that way to me, too.'

'Okay then. Which of these do *multiplicity* (ἀριθμός) and *unity* (τὸ ἕν) seem to be?'

'I can't imagine' he said.

'Well,' I said 'reason it out from what we said. If unity is fully grasped alone, in itself, by sight or some other sense, then it must be [an object] like a finger, as we were explaining: it does not draw us towards *being-ness* (οὐσία). But if some discrepancy is always seen with it, so as to appear not rather *one* (ἕν) than its opposite, then a decision is needed—indeed, the *soul* (ψυχή) in itself is compelled to be puzzled, and to cast about, arousing thought within itself, and to ask: What then is unity as such? And so the *study* (μάθησις) of unity must be among those that lead and guide [the soul] to the sight of *that which is* (τὸ ὄν).'

'But certainly' he said 'vision is especially like that. For, the same thing is seen as one and as *indefinite multitude* (ἄπειρα τὸ πλῆθος).'

'If it is so with unity,' I said 'is it not so with every *number* (ἀριθμός)?'

'How could it not be?'

'But *calculation* (λογιστική) and *number-theory* (ἀριθμητική) are entirely about number.'

'Absolutely.'

'And these things appear to lead to truth.'

'Yes, and extremely well.'

'So it seems that these must be some of the *studies* (μαθημάτα) that we are looking for. Indeed, the *military* (πολεμικόν) needs to learn them for deployment [of troops],—and the philosopher, because he has to rise out of [the world of] *becoming* (γένεσις) in order to take hold of being-ness, or else he will never *become a calculator* (λογιστικῷ γενέσθαι).'

'Just so' he said.

Table I: The Greek alphabet

Α α	**a**lpha	Η η	**ē**ta	Ν ν	**nu**	Τ τ	**t**au
Β β	**b**eta	Θ ϑ	**th**eta	Ξ ξ	**xi**	Υ υ	**u**psilon
Γ γ	**g**amma	Ι ι	**i**ota	Ο ο	**o**micron	Φ φ	**ph**i
Δ δ	**d**elta	Κ κ	**k**appa	Π π	**p**i	Χ χ	**ch**i
Ε ε	**e**psilon	Λ λ	**l**ambda	Ρ ρ	**r**ho	Ψ ψ	**ps**i
Ζ ζ	**z**eta	Μ μ	**mu**	Σ σ/ς	**s**igma	Ω ω	**ō**mega

The first letter or two of the (Latin) name provides a transliteration for the Greek letter. In texts, the rough-breathing mark ῾ over an initial vowel (or ρ) corresponds to a preceeding h; the smooth-breathing mark ᾿ and the three tonal accents can be ignored.

> 'And our guardian happens to be both military man and philosopher.'
>
> 'Of course.'
>
> 'So, Glaucon, it is appropriate to require this study by law and to persuade those who intend to take part in the greatest affairs of the city to go into calculation and to engage in it not *as a pastime* (ἰδιωτικῶς), but until they have attained, by thought itself, the vision of the nature of numbers, not [for the sake of] buying and selling, as if they were preparing to be merchants or shopkeepers, but for the sake of war and an easy turning of the soul itself from becoming towards truth and being-ness.'
>
> 'You speak superbly' he said.

(In reading this passage from Plato, and in particular the comments on war, one can hardly be sure that Socrates is not pulling Glaucon's leg. Socrates previously (369b–372c) described a primitive, peaceful, vegetarian city, which Glaucon rejected (372c–d) as being fit only for pigs.)

The reader of the present notes is not assumed to have much knowledge 'officially'. But the reader should have some awareness of the Boolean connectives of propositional logic and their connexion with the Boolean operations on sets. (A dictionary of the connectives is in Table II below.)

One theme of these notes is the relation between *definition by recursion* and *proof by induction*. The development of propositional logic already requires recursion and induction.[5] For example, **propositional formulas**[6] are defined recursively:

($*$) *Propositional variables* and 0 and 1 are propositional formulas.

(†) If F is a propositional formula, then so is $\neg F$.

(‡) If F and G are propositional formulas, then so is $(F \mathbin{\square} G)$, where $\square$ is $\wedge$, $\vee$, $\rightarrow$ or $\leftrightarrow$.

The **sub-formulas** of a formula are also defined recursively:

($*$) Every formula is a sub-formula of itself.

[5]Here I use the words 'recursion' and 'induction' in a more general sense than in the definitions on pp. 21 and 23.

[6]Words in bold-face in these notes are being defined.

Table II: Boolean connectives

$\wedge$	*and*	conjunction
$\vee$	*or*	disjunction
$\neg$	*not*	negation
$\rightarrow$	*implies*	implication
$\leftrightarrow$	*if and only if*	biconditional

(†) Any sub-formula of a formula F is a sub-formula of $\neg F$.

(‡) Any sub-formula of F or G is a sub-formula of $(F \mathbin{\square} G)$.

Now, two formulas are **equivalent** if they have the same *truth-table*. For example, $(P \to Q)$ and $(\neg P \lor Q)$ are equivalent, because their truth-tables are, respectively:

P	$\to$	Q
0	1	0
1	0	0
0	1	1
1	1	1

$\neg$	P	$\lor$	Q
1	0	1	0
0	1	0	0
1	0	1	1
0	1	1	1

Suppose F and G are equivalent; this is denoted

$$F \sim G.$$

Suppose also F is a sub-formula of H, and H' is the result of replacing F in H with G. The **Substitution Theorem** is that

$$H \sim H'.$$

Because of the recursive definition of propositional formulas, we can prove the Substitution Theorem by induction as follows:

($*$) The claim is trivially true when H is a propositional variable or 0 or 1, since then F *is* H, so H' is G.

(†) Suppose, as an inductive hypothesis, that the claim is true when H is H_0. Then we can show that the claim is true when H is $\neg H_0$.

(‡) Suppose, as an inductive hypothesis, that the claim is true when H is H_0 and when H is H_1. Then we can show that the claim is true when H is $(H_0 \mathbin{\square} H_1)$, where $\square$ is as above.

Such a proof is sometimes said to be a 'proof by induction on the complexity of propositional formulas'.

A conjunction corresponds to an *intersection* of sets, and so forth, but this is spelled out in § 1 below. I shall also use formulas of *first-order* logic, and in particular the *quantifiers* (given in Table III below). For emphasis, instead of $\to$ and $\leftrightarrow$, I may use the arrows $\Longrightarrow$ and $\Longleftrightarrow$ between formulas.

My own research-interests lie more in model-theory than in set-theory. I aim here just to set down some established mathematics as precisely as possible,

Table III: Quantifiers

$\forall$	*for all*	universal
$\exists$	*there exists... such that*	existential

without much discussion. (There is a textbook that has been in use for over two thousand years, but that contains no discussion at all, only axioms, definitions, theorems and proofs. This is Euclid's *Elements* [8].) I do think that explicit reference to models can elucidate some points. The reader should also consult texts by people who *are* set-theorists, for other points of view, for historical references, and to see how the field has developed beyond what is given in these notes. Also, the reader should remember that these notes are still a rough draft. There are not yet many exercises, and some of them are difficult or lacking in clear answers.

0.2

Any text on axiomatic set-theory will introduce the set ω, which is the smallest set that contains $\varnothing$ and that contains $x \cup \{x\}$ whenever it contains x. The text *may* (but need not) mention that ω is a model of the Peano axioms for the natural numbers. The present notes differ from some published texts in two ways:

- I prove facts about the natural numbers *from the Peano axioms*, not just *in* ω.
- I mention structures that are models of some, but not all, of the Peano axioms.

I set out a minimum of set-theory in § 1, enough so that the properties of natural numbers can be derived from the Peano axioms, starting in § 3. Some set-theory books, such as Ciesielski [2, § 3.1], will immediately give ω as a model of these axioms. Certain properties of natural numbers are easier to prove *in this model* than *by the Peano axioms.* I prefer to follow the axiomatic approach for several reasons:

One reason is practice. It is worthwhile to have experience with the Peano axioms as well as **ZFC**, especially since, unlike **ZFC**, the Peano axioms include a second-order statement. (It may be that some writers assume that the reader has already had sufficient practice with the Peano axioms; I do not make such an assumption.)

The Peano axioms are more natural than their specific model ω. The elements of ω (as well as ω itself) are so-called *von Neumann ordinals*, that is, *transitive* sets that are *well-ordered* by *containment.* In a slightly different context, the model-theorist Poizat [16, § 8.1] observes:

> We meet some students who are allergic to ordinals as 'well-ordering types' and who find the notion of von Neumann ordinals easier to digest; that is a singular consequence of dogmatic teaching, which confuses formalism with rigor, and which favors technical craft to the detriment of the fundamental idea: It takes a strangely warped mind to find the notion of a transitive set natural!

A third reason for taking the axiomatic approach to the natural numbers is that it can bring out a distinction that is often ignored. The structure of the natural numbers admits *proof by induction* and *definition by recursion*. Vaught [18, ch. 2, § 4], for example, says that recursion *is* 'the same thing as definition by induction'. Since it is just about terminology, the statement is not wrong. But definition by 'induction' or recursion[7] works *only* in models of the Peano axioms, while there are other structures in which *proof* by induction[8] works.

There are 'strong' versions of induction and recursion. There is proof by strong induction, and definition by strong recursion. Admission of either of *these* is equivalent to admission of the other; the structures that admit them are precisely the well-ordered sets. Some basic undergraduate texts suggest confusion on this point. For example, in talking about the integers, one book[9] says:

> It is apparent that if the principle of strong mathematical induction is true, then so is the principle of ordinary mathematical induction... It can also be shown that if the principle of ordinary mathematical induction is true, then so is the principle of strong mathematical induction. A proof of this fact is sketched in the exercises...

Both statements about induction here are literally false. The second statement is correct if it is understood to mean simply that the natural numbers satisfy the principle of strong induction. The 'proof' that is offered for the first statement uses implicitly that every integer is a *successor*, something that does not follow from strong induction.

Finally, by emphasizing the axiomatic development of the natural numbers, I hope to encourage the reader to watch out for unexamined assumptions, in these notes and elsewhere. The Hajnal text [9] defines ω on the first page of § 1 as 'the set of nonnegative integers'. Then come a hundred pages of the set-theory covered in the present notes, and more. The Preface says that this work 'is carried out on a quite precise, but intuitive level'; only after *this* does the reader get, in an appendix, on p. 127, a rigorous definition of ω. To my mind, the precise but intuitive way to treat the natural numbers is by means of the Peano axioms. Perhaps the reader of Hajnal is supposed to have seen such a treatment before, since, according to the index, the term 'Peano' appears only once, on p. 133, and there is no definition.

Devlin [6] seems never to mention the natural numbers as such at all, though early on (p. 6), he asserts the existence of sets $\{a_1, \ldots, a_n\}$. (Later he defines the symbol ω, naïvely on p. 24, rigorously on p. 66.) Like Hajnal, Moschovakis [13] *names* the set of natural numbers on the first page of text; but then he discusses set-theory for only fifty pages before devoting a chapter to a rigorous treatment of the natural numbers.

1 Sets and classes

A set has **members**, or **elements**. A set **contains** its elements, and the elements **compose** the set. To say that a set A has an element b, we may write

$$b \in A,$$

[7] In the sense defined on p. 23 below.

[8] In the sense defined on p. 21.

[9] Namely, Epp [7, § 4.4, p. 213], used sometimes in Math 111 and 112.

using between b and A a symbol derived from the Greek minuscule epsilon, which can be understood as standing for the Latin word ELEMENTVM. A set is not *distinct* from its elements in the way that a box is distinct from its contents. A set may be distinct from any *particular* element. But I propose to say that a set *is* its elements, and the elements *are* the set.

This is a paradoxical statement. How can one thing be many, and many, one? The difficulty of answering this is perhaps reflected in the difficulties of set-theory. In any case, if a set is its elements, then the elements *uniquely determine* the set. This is something whose meaning we can express mathematically; it is perhaps the most fundamental axiom of set-theory:

1.1 Axiom (Extensionality). *If two sets A and B have the same members, then $A = B$.*

The converse of this axiom is trivially true: If two sets have different members, then of course the sets themselves are different.

A set is also the sort of thing that can *be* an element: If A and B are sets, then the statement $A \in B$ is meaningful, and the statement

$$A \in B \vee A \notin B$$

is true.

Are all elements sets themselves? We do not answer this question; we avoid it:

1.2 Definition. A property P of sets is **hereditary**, provided that, if A is a set with property P, then all elements of A are *sets* with property P. A *set* is **hereditary** if it has a hereditary property.

We shall ultimately restrict our attention to hereditary sets.[10] Now, we shall not assert, as an axiom, that all sets are hereditary. We cannot now formulate such an axiom precisely, since we do not yet have a definition of a 'property' of sets. The *language* with which we talk about sets will end up ensuring that our sets are hereditary:

Everything that we shall say about sets can be said with the symbol $\in$, along with $=$ and the logical symbols given in Tables II and III of § 0, and with variables and names *for individual sets.*

1.3 Definition. The **$\in$-formulas** are recursively defined[11] as follows:

(∗) If x and y are variables, and A and B are names, then $x \bigcirc y$, $x \bigcirc A$, $A \bigcirc x$ and $A \bigcirc B$ are **atomic** $\in$-formulas, where $\bigcirc$ is $\in$ or $=$.

(†) If ϕ and ψ are $\in$-formulas, then so are $\neg\phi$ and $(\phi \square \psi)$, where $\square$ is one of $\wedge$, $\vee$, $\to$ and $\leftrightarrow$.

[10] See also Kunen [10, ch. 1, § 4] for discussion of this point.

[11] This definition uses also brackets (parentheses) in the formulas, but the brackets do not carry meaning in the way that the other symbols do. The brackets are meaningful in the way that the *order* of the symbols in a formula is meaningful. Indeed, we could dispense with the brackets by using the so-called Polish or Łukasiewicz notation, writing, say, $\wedge\phi\psi$ instead of $\phi \wedge \psi$. I shall use the infix notation instead, but shall omit brackets where they are not needed.

(‡) If ϕ is an $\in$-formula and x is a variable, then $(\mathsf{Q}x\ \phi)$ is an $\in$-formula, where Q is $\forall$ or $\exists$.

The $\in$-formulas are the *first-order* formulas in the *signature* consisting of $\in$ alone. Other signatures are discussed later (see Definition 2.3). In any signature, the first-order formulas are defined recursively as $\in$-formulas are, but the atomic formulas will be different. In a first-order formula, only variables can follow quantifiers; otherwise, the distinction between a variable and a name is not always clear (see Exercise 2.6). Also, in a first-order formula, variables and names refer only to individual objects, rather than, say, sets of objects. In set-theory, our objects *are* sets, so it would not make much sense to have more than one kind of variable.[12]

Variables and names in $\in$-formulas are also called **terms**. (In other signatures, there will be a more general definition of *term*.) Names used in formulas may be called **parameters**.

In an $\in$-formula, instead of a sub-formula $\neg\, x \in y$, we can write

$$x \notin y;$$

and instead of $\neg\, x = y$, we can write

$$x \neq y;$$

here, x and y are terms.

If a first-order formula contains no quantifiers, then its variables are **free** variables. The free variables of $\exists x\ \phi$ and $\forall x\ \phi$ are those of ϕ, *except* x. The free variables of $\neg\phi$ are just those of ϕ. Finally, the free variables of $\phi \mathrel{\square} \psi$ (where $\square$ is one of $\wedge$, $\vee$, $\rightarrow$ and $\leftrightarrow$) are those of ϕ or ψ. A **sentence** is a formula with no free variables.

We can now attempt to write the Extensionality Axiom as the sentence

$$\forall x\ \forall y\ (\forall z\ (z \in x \leftrightarrow z \in y) \rightarrow x = y). \qquad (*)$$

Now, if the variables x, y and z can refer to any sets at all, and if some sets contain objects that are not sets, then $(*)$ is actually stronger than Axiom 1.1. Indeed, if A is a set, and b is an object that is not a set, then there might be a set $\{A, b\}$ containing A and b and nothing else, and a set $\{A\}$ containing A and nothing else. Then for all *sets* z, we have

$$z \in \{A, b\} \iff z \in \{A\}.$$

From this, $(*)$ seems to imply $\{A, b\} = \{A\}$, which is evidently false. Our solution to this problem will be to restrict the variables in formulas like $(*)$ to *hereditary* sets. In this way, $(*)$ becomes merely a special case of the Extensionality Axiom. In particular, since the set $\{A, b\}$ is not hereditary, $(*)$ says nothing about it.

1.4 Exercise. Alternatively, we might let our variables range over all (mathematical) objects, even if some of these might not be sets. If a is not a set, then we should require $\forall x\ x \notin a$. In this case, if $(*)$ is still true, prove that there is at most one object that is not a set.

[12] See also the comments of Levy [12, ch. 1, § 1, p. 4].

In the 'Platonic' view of set-theory, when the logical symbols in an $\in$-sentence are interpreted as in Tables II and III of § 0, and when terms are understood to refer to hereditary sets, then the sentence is either true or false. (A 'relative' notion of truth is given in Definition 2.5.) Then we are looking for the true $\in$-sentences; in particular, we are looking for some 'obviously' true sentences—*axioms*—from which all other true sentences about hereditary sets follow logically.[13]

A first-order formula ϕ with at most one free variable is called **unary**; if that free variable is x, then the formula might be written

$$\phi(x).$$

If this is an $\in$-formula, it expresses a **property** that sets might have. If A has that property, then we can assert

$$\phi(A).$$

Formally, we obtain the sentence $\phi(A)$ from ϕ by replacing each *free occurrence* of x with A. A precise recursive definition of $\phi(A)$ is possible. Here it is, for thoroughness, although we shall not spend time with it:

1.5 Definition. For any first-order formula ϕ, variable x and term t, the formula

$$\phi^x_t$$

is the result of *freely* replacing each free occurrence of x in ϕ with t; it is determined recursively as follows:

- ($*$) If ϕ is atomic, then ϕ^x_t is the result of replacing *each* instance of x in ϕ with t.
- (†) $(\neg\phi)^x_t$ is $\neg(\phi^x_t)$, and $(\phi \mathbin{\square} \psi)^x_t$ is $\phi^x_t \mathbin{\square} \psi^x_t$.
- (‡) $(\mathsf{Q}x\ \phi)^x_t$ is $\mathsf{Q}x\ \phi$.
- (§) If y is not x and does not appear in t, then $(\mathsf{Q}y\ \phi)^x_t$ is $\mathsf{Q}y\ \phi^x_t$.
- (¶) If y is not x, but y does appear in t, then $(\mathsf{Q}y\ \phi)^x_t$ is $\mathsf{Q}z\ (\phi^y_z)^x_t$, where z is a variable that does not appear in t or ϕ.

If ϕ is $\phi(x)$, then ϕ^x_t can be denoted

$$\phi(t).$$

The point is that if, for example, ϕ is $\psi(x) \wedge \exists x\ \chi(x)$, then $\phi(A)$ is $\psi(A) \wedge \exists x\ \chi(x)$. Alternatively, ϕ might be $\exists y\ (\psi(x) \wedge \chi(y))$, in which case $\phi(y)$ is $\exists z\ (\psi(y) \wedge \chi(z))$, not $\exists y\ (\phi(y) \wedge \chi(y))$.

The sets with the property given by $\phi(x)$ compose a **class**, denoted

$$\{x : \phi(x)\}. \tag{†}$$

[13] This project must fail. By Gödel's Incompleteness Theorem, we cannot define a list of axioms from which all truths of set-theory follow. We can still hope to identify axioms from which *some* interesting truths follow. One purpose of these notes is to develop some of these interesting truths.

This is the class of sets that **satisfy** ϕ, the class of x such that $\phi(x)$. But not every class is a set; not every class is a unity to the extent that it can be considered as a member of sets:

1.6 Theorem (Russell Paradox). *The class $\{x : x \notin x\}$ is not a set.*

Proof. Suppose A is a set such that

$$x \in A \implies x \notin x \tag{‡}$$

for all sets x. Either $A \notin A$ or $A \in A$, but in the latter case, by (‡), we still have $A \notin A$. Therefore A is a member of $\{x : x \notin x\}$, but not of A itself; so

$$A \neq \{x : x \notin x\}.$$

Hence $\{x : x \notin x\}$ cannot be a set. □

1.7 Remark. The Russell Paradox is often established by contradiction: If the class $\{x : x \notin x\}$ is a set A, then both $A \in A$ and $A \notin A$, which is absurd. However, the proof given above shows that a false assumption is not needed.

1.8 Exercise. The sets that we are considering compose the class $\{x : x = x\}$. It is logically true that

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

Explain how this is a proof that all of our sets are hereditary.

Not every class is a set; but every set A is the class $\{x : x \in A\}$. A *disjunction* of formulas gives us the **union** of corresponding classes:

$$\{x : \phi(x) \lor \psi(x)\} = \{x : \phi(x)\} \cup \{x : \psi(x)\}.$$

Likewise, a *conjunction* gives an **intersection**:

$$\{x : \phi(x) \land \psi(x)\} = \{x : \phi(x)\} \cap \{x : \psi(x)\};$$

and a *negation* gives a **complement**:

$$\{x : \neg\phi(x)\} = \{x : \phi(x)\}^{\mathrm{c}}.$$

Finally, we can form a **difference** of classes, not corresponding to a single Boolean connective from our list:

$$\{x : \phi(x) \land \neg\psi(x)\} = \{x : \phi(x)\} \smallsetminus \{x : \psi(x)\}.$$

If $\forall x\, (\phi(x) \to \psi(x))$, then $\{x : \phi(x)\}$ is a **sub-class** of $\{x : \psi(x)\}$. We can write

$$\forall x\, (\phi(x) \to \psi(x)) \iff \{x : \phi(x)\} \subseteq \{x : \psi(x)\}.$$

We now have several abbreviations to use in writing $\in$-formulas:

$$\begin{aligned} x \in A \cup B &\iff x \in A \lor x \in B; \\ x \in A \cap B &\iff x \in A \land x \in B; \\ x \in A \smallsetminus B &\iff x \in A \land x \notin B; \\ A \subseteq B &\iff \forall x\, (x \in A \to x \in B). \end{aligned}$$

If sets exist at all, then any two sets ought to be members of some set:

1.9 Axiom (Pairing). *For any two sets, there is a set that contains them:*

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

The set given by the axiom might have elements other than those two sets; we can cast them out by means of:

1.10 Axiom (Comprehension). *A sub-class of a set is a set: For any* $\in$-*formula* $\phi(x)$,

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

Note that this axiom is not a single $\in$-sentence, but a *scheme* of $\in$-sentences.

A sub-class of a set A can now be called a **subset** of A. A set **includes** its subsets. A subset B of A that is distinct from A is a **proper** subset of A, and we may then write

$$B \subset A.$$

We now have that, for any x and y, there is a set

$$\{x, y\}$$

whose members are *just* x and y; if $x = y$, then this set is

$$\{x\},$$

which is sometimes called a **singleton**.

1.11 Exercise. Prove that the class of all sets is not a set.

If A is a set and ϕ is a (unary) formula, then the set $A \cap \{x : \phi(x)\}$ can be written

$$\{x \in A : \phi(x)\}.$$

In particular, if B is also a set, then

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

As long as **some** set A exists, we have the **empty set**,

$$\varnothing,$$

which can be defined as $\{x \in A : x \neq x\}$. Does some set exist? I take this as a logical axiom:

$$\exists x\ x = x. \tag{§}$$

Indeed, *something* exists, as we might argue along with Descartes [4, II, ¶ 3]:

> Therefore I will suppose that all I see is false... But certainly I should exist, if I were to persuade myself of something... Thus it must be granted that, after weighing everything carefully and sufficiently, one must come to the considered judgment that the statement '*I am, I exist* (EGO SVM, EGO EXISTO)' is necessarily true every time it is uttered by me or conceived in my mind [5, p. 17].

Of course, we are claiming that *hereditary sets* exist. But I take (§) to be implicit in the assertion of any sentence, such as (∗).

For any x and y, the **ordered pair** (x, y) is the set

$$\{\{x\}, \{x, y\}\}.$$

All that we require of this definition is that it allow us to prove the following:

1.12 Theorem. $(x, y) = (u, v) \iff x = u \wedge y = v.$

1.13 Exercise. Prove the theorem.

Given two classes $\mathbf{C}$ and $\mathbf{D}$, we can now form their **cartesian product**:

$$\mathbf{C} \times \mathbf{D} = \{(x, y) : x \in \mathbf{C} \wedge y \in \mathbf{D}\}.$$

1.14 Lemma. *A cartesian product of classes is a well-defined class, that is, can be written as $\{x : \phi(x)\}$ for some $\in$-formula ϕ.*

1.15 Exercise. Prove the lemma.

To prove that the cartesian product of *sets* is a set, we can use:

1.16 Axiom (Power-set). *If A is a set, then there is a set B such that*

$$x \subseteq A \implies x \in B$$

for all sets x. That is,

$$\forall x\ \exists y\ \forall z\ (\forall w\ (w \in z \to w \in x) \to z \in y).$$

Hence, for any set A, its **power-set** $\{x : x \subseteq A\}$ is a set; this is denoted

$$\mathcal{P}(A).$$

In particular, $(x, y) \in \mathcal{P}(\mathcal{P}(\{x, y\}))$, so $A \times B \subseteq \mathcal{P}(\mathcal{P}(A \cup B))$.

If A is a set, its **union** is $\{x : \exists y\ (y \in A \wedge x \in y)\}$, denoted

$$\bigcup A.$$

In particular, for any sets A and B,

$$A \cup B = \bigcup \{A, B\}.$$

1.17 Exercise. What are $\bigcup \varnothing$ and $\bigcup \{\varnothing\}$?

1.18 Axiom (Union). *The union of a set is a set:*

$$\forall x\ \exists y\ \forall z\ (\exists w\ (z \in w \wedge w \in x) \to z \in y).$$

The union of a set A might be denoted also

$$\bigcup_{x\in A} x.$$

Suppose that, for each x in A, there is a set B_x. We shall soon be able to define a union

$$\bigcup_{x\in A} B_x. \tag{¶}$$

This will be the union of $\{B_x : x \in A\}$. But for now, we don't even know that this thing is a well-defined *class*, much less a set.

1.19 Theorem. *The cartesian product of sets is a set.*

1.20 Exercise. Prove the theorem.

If A is a set, its **intersection** is $\{x : \forall y\ (y \in A \to x \in y)\}$, denoted

$$\bigcap A.$$

If A contains a set B (that is, if $B \in A$), then $\bigcap A \subseteq B$, so $\bigcap A$ is a set. Also, for any sets A and B,

$$A \cap B = \bigcap\{A, B\}.$$

1.21 Exercise. What are $\bigcap\varnothing$ and $\bigcap\{\varnothing\}$?

A **relation** between A and B is a subset of $A \times B$. If $R \subseteq A \times B$, then

$$R^{-1} = \{(y, x) : (x, y) \in R\},$$

a relation between B and A. If also $S \subseteq B \times C$, then

$$S \circ R = \{(x, z) : \exists y\ ((x, y) \in R \land (y, z) \in S)\},$$

a relation between A and C.

A relation between A and itself is a **binary** relation on A. The set

$$\{(x, x) : x \in A\}$$

is the **diagonal** Δ_A on A. A binary relation R on A is:

- **reflexive**, if $\Delta_A \subseteq R$;
- **irreflexive**, if $\Delta_A \cap R = \varnothing$;
- **symmetric**, if $R^{-1} = R$;
- **anti-symmetric**, if $R \cap R^{-1} \subseteq \Delta_A$;
- **transitive**, if $R \circ R \subseteq R$.

Then R is:

- an **equivalence-relation**, if it is reflexive, symmetric and transitive;

- a **partial ordering**, if it is anti-symmetric and transitive and either reflexive or irreflexive;
- a **strict** partial ordering, if it is an irreflexive partial ordering;
- a **total ordering**, if it is a partial ordering and $R \cup \Delta_A \cup R^{-1} = A \times A$.

A relation f between A and B is a **function**, or **map**, from A to B if

$$f \circ f^{-1} \subseteq \Delta_B \wedge \Delta_A \subseteq f^{-1} \circ f.$$

Suppose f is thus. We may refer to the function $f : A \to B$. For each x in A, there is a unique element $f(x)$ of B such that $(x, f(x)) \in f$. Here $f(x)$ is the **value** of f at x. We may refer to f as

$$x \longmapsto f(x) : A \longrightarrow B.$$

The **domain** of f is A, and f is a function **on** A. The **range** of f is the set $\{y \in B : \exists x\ (x, y) \in f\}$, that is, $\{f(x) : x \in A\}$, which is denoted

$$f''A.$$

If $C \subseteq A$, then $f \cap (C \times B)$ is a function on C, denoted $f|_C$ and having range $f''C$. This set is also the **image** of C under f.

The function $f : A \to B$ is:

- **surjective** or **onto**, if $\Delta_B \subseteq f \circ f^{-1}$;
- **injective** or **one-to-one**, if $f^{-1} \circ f \subseteq \Delta_A$;
- **bijective**, if surjective and injective (that is, one-to-one and onto).

All of the foregoing definitions involving relations make sense even if A and B are merely classes.

To discuss functions in the most general sense, it is convenient to introduce a new quantifier,

$$\exists!,$$

read 'there exists a unique...such that'; this quantifier is defined by

$$\exists!\, x\ \phi(x) \iff \exists x\ \phi(x) \wedge \forall y\ (\phi(y) \to y = x).$$

A formula ψ with free variables x and y at most can be written

$$\psi(x, y);$$

it is a **binary** formula. Then a function is a class $\{(x, y) : \psi(x, y)\}$ such that

$$\forall x\ (\exists y\ \psi(x, y) \to \exists!\, y\ \psi(x, y)).$$

The domain of this function is $\{x : \exists y\ \psi(x, y)\}$. If the function itself is called f, and if its domain includes a set A, then the image $f''A$ or $\{f(x) : x \in A\}$ is the class

$$\{y : \exists x\ (x \in A \wedge \psi(x, y)\}.$$

That this class is a *set* is the following:

1.22 Axiom (Replacement). *The image of a set under a function is a set: For all classes* $\{(x,y) : \psi(x,y)\}$ *that are* functions,

$$\forall x \; \exists y \; \forall z \; \forall w \; (z \in x \wedge \psi(z,w) \rightarrow w \in y).$$

Like Comprehension, the Replacement Axiom is a scheme of $\in$-sentences. Indeed, for each binary formula $\psi(x,y)$, we have

$$\forall x \; (\exists y \; \psi(x,y) \rightarrow \exists! \, y \; \psi(x,y)) \rightarrow \forall x \; \exists y \; \forall z \; \forall w \; (z \in x \wedge \psi(z,w) \rightarrow w \in y).$$

If we have a function $x \mapsto B_x$ on a set A, then the union (¶) above is now well-defined.

Other set-theoretic axioms will arise in the course of the ensuing discussion.

2 Model-theory

A **unary** relation on a set is just a subset. A unary **operation** on a set is a function from the set to itself. A **binary** operation on a set A is a function from $A \times A$ to A. We can continue. A *ternary* relation on A is a subset of

$$A \times A \times A,$$

that is, $(A \times A) \times A$, also denoted A^3. A ternary operation on A is a function from A^3 to A. More generally:

2.1 Definition. The **cartesian powers** of a set A are defined recursively:

(∗) A is a cartesian power of A.

(†) If B is a cartesian power of A, then so is $B \times A$.

A **relation** on A is a subset of a cartesian power of A. An **operation** on A is a function from a cartesian power of A into A.

Note that we do not (yet) assert the existence of a *set* containing the cartesian powers of A.

2.2 Exercise. Is there a *class* containing the cartesian powers of a given set and nothing else?

2.3 Definition. A **structure** is an ordered pair

$$(A, T),$$

where A is a non-empty set, and T is a set (possibly empty) whose elements are operations and relations on A and elements of A. The set A is the **universe** of the structure. The structure itself can then be denoted

$$\mathfrak{A}$$

(or just A again). A **signature** of $\mathfrak{A}$ is a set S of symbols for the elements of T: This means:

(∗) There is a bijection $s \mapsto s^{\mathfrak{A}} : S \to T$.

(†) Different structures can have the same signature.

The element $s^{\mathfrak{A}}$ of T is the **interpretation** in $\mathfrak{A}$ of the symbol s. Usually one doesn't bother to write the superscript for an interpretation, so s might really mean $s^{\mathfrak{A}}$.

In the next section, we shall assert as an axiom—the Peano Axiom—the existence of a structure

$$(\mathbb{N}, \{^+, 0\})$$

having certain properties. The universe $\mathbb{N}$ will be the set of *natural numbers*, and $^+$ will be the unary operation $x \mapsto x+1$. The structure is more conveniently written as

$$(\mathbb{N}, ^+, 0);$$

we shall also look at structures $(\mathbb{N}, ^+, 0, P)$, where P is a unary relation on $\mathbb{N}$.

An $\in$-sentence is supposed to be a statement about the world of (hereditary) sets. Structures live in this world. The signature of a structure allows us to write sentences that are true or false *in the structure*. The Peano Axiom will be that certain sentences of the signatures $\{^+, 0\}$ and $\{^+, 0, P\}$ are *true in* $(\mathbb{N}, ^+, 0)$ and $(\mathbb{N}, ^+, 0, P)$.

2.4 Definition. The **terms** of $\{^+, 0\}$ and $\{^+, 0, P\}$ are defined recursively:

(∗) Variables and names and 0 are terms.

(†) If t is a term, then so is t^+.

The **atomic** formulas of $\{^+, 0\}$ are equations $t = u$ of terms; the signature $\{^+, 0, P\}$ also has atomic formulas

$$P(t),$$

where t is a term. From the atomic formulas, formulas are built up as in Definition 1.3.

The definition can be generalized to other signatures. If for example the signature has a binary operation-symbol $+$, and t and u are terms of the signature, then so is $(t + u)$.

2.5 Definition. An atomic *sentence* σ becomes **true** or **false in** a structure $\mathfrak{A}$, once interpretations $c^{\mathfrak{A}}$ are chosen for any names c appearing in σ; if σ is true in $\mathfrak{A}$, then we write

$$\mathfrak{A} \models \sigma, \tag{$\|$}$$

and we say that $\mathfrak{A}$ is a **model** of σ. Note that ($\|$) could be written out as an $\in$-sentence. For arbitrary sentences, we define:

$$\mathfrak{A} \models \neg\sigma \iff \neg(\mathfrak{A} \models \sigma),$$
$$\mathfrak{A} \models \sigma \mathbin{\square} \tau \iff \mathfrak{A} \models \sigma \mathbin{\square} \mathfrak{A} \models \tau,$$

where $\square$ is $\wedge$, $\vee$, $\to$ or $\leftrightarrow$. Finally,

$$\mathfrak{A} \models \forall x\ \phi(x) \tag{$**$}$$

if and only if $\mathfrak{A} \models \phi(a)$ for all a in A; and

$$\mathfrak{A} \models \forall x\ \phi(x) \iff \mathfrak{A} \models \neg\exists x\ \neg\phi(x).$$

2.6 Exercise. In the definition of $(**)$, is a a variable or a name?

3 The Peano axioms

The five so-called Peano axioms amount to the following five-part assertion:

3.1 Axiom (Peano). *There is a set* $\mathbb{N}$*,*

- $(*)$ *containing a distinguished element* 0 *(called* ***zero****), and*
- $(\dagger)$ *equipped with a unary operation* $x \mapsto x^+$ *(the* ***successor-operation****), such that*
- $(\ddagger)$ $(\mathbb{N}, {}^+, 0) \models \forall x\ x^+ \neq 0$*;*
- $(\S)$ $(\mathbb{N}, {}^+) \models \forall x\ \forall y\ (x^+ = y^+ \to x = y)$*;*
- $(\P)$ $(\mathbb{N}, {}^+, 0, P) \models P(0) \wedge \forall x\ (P(x) \to P(x^+)) \to \forall x\ P(x)$*, for every unary relation* P *of* $\mathbb{N}$*.*

Thus, in one sense, there is a single 'Peano axiom', asserting that a structure $(\mathbb{N}, {}^+, 0)$ exists with certain properties. Its properties are that it satisfies the following three *axioms*—where now 'axiom' is used in a slightly different sense:

Axiom Z $\forall x\ x^+ \neq 0$;

Axiom U $\forall x\ \forall y\ (x^+ = y^+ \to x = y)$;

Axiom I $0 \in X \wedge \forall x\ (x \in X \to x^+ \in X) \to \forall x\ x \in X$, for every subset X.

The set-theoretic axioms given in § 1 are supposed to be true in the mathematical world. The three axioms just above are supposed to be true *in a particular structure* in the mathematical world. Note that Axiom I, considered as a single sentence, is not a first-order sentence, but is *second-order*, since the variable X refers to sub*sets* of a model, and not to elements. (Axiom Z and Axiom U are first-order.)

3.2 Remark. In first-order logic, Axiom I is replaced by a *scheme* of axioms, consisting of one sentence

$$\phi(0) \wedge \forall x\ (\phi(x) \to \phi(x^+)) \to \forall x\ \phi(x) \tag{$*$}$$

for each unary first-order formula ϕ in the signature $\{{}^+, 0\}$ with parameters. This scheme of axioms is *weaker* than Axiom I, because not every subset of $\mathbb{N}$ is defined by a first-order formula. (Later we shall be able to prove this:

There are *countably* many formulas $\phi(x)$, but $\mathbb{N}$ has *uncountably* many subsets.) This scheme of axioms ($*$), together with Axiom Z and Axiom U, might be denoted **PA**. It is a consequence of Gödel's Incompleteness Theorem that **PA** is an *incomplete* theory. This means that some first-order sentences are true in $(\mathbb{N}, {}^+, 0)$, but are not logical consequences of **PA**. In fact, there are models of **PA** that are not models of Axiom I.

To talk more about the Peano Axioms, we make the following:

3.3 Definition. A natural number is called a **successor** if it is x^+ for some x in $\mathbb{N}$. We have special names for certain successors:

x	0	1	2	3	4	5	6	7	8
x^+	1	2	3	4	5	6	7	8	9

A natural number x is an **immediate predecessor of** y if $x^+ = y$.

Later we shall define the binary operation $(x, y) \mapsto x + y$ so that $x^+ = x + 1$.

Our names for the Peano Axioms are tied to their meanings (although these names are not in general use):

- Axiom Z is that *Z*ero is not a successor.
- Axiom U is that immediate predecessors are *U*nique when they exist.
- Axiom I is the Axiom of **Induction**: a set contains all natural numbers, provided that it contains 0 and contains the successor of each natural number that it contains.

Also, Axiom Z is that the immediate predecessor of 0 does *not* exist.[14] Axiom U is that the successor-operation is injective.

We may henceforth write $\mathbb{N}$ instead of $(\mathbb{N}, {}^+, 0)$. As first examples of the Induction Axiom in action, we have:

3.4 Lemma. *Every non-zero natural number is a successor. Symbolically,*

$$\mathbb{N} \models \forall x \ (x = 0 \lor \exists y \ y^+ = x).$$

Proof. Let A be the set of natural numbers comprising 0 and the successors. That is, $A = \{0\} \cup \{x \in \mathbb{N} : \exists y \ y^+ = x\}$. Then $0 \in A$ by definition. Also, if $x \in A$, then x^+ is a successor, so $x^+ \in A$. By induction, $A = \mathbb{N}$. $\square$

3.5 Theorem. *The successor-operation is a bijection between* $\mathbb{N}$ *and* $\mathbb{N} \smallsetminus \{0\}$.

3.6 Exercise. Prove the theorem.

3.7 Lemma. *Every natural number is distinct from its successor:*

$$\mathbb{N} \models \forall x \ x^+ \neq x.$$

[14] Peano did not count 0 as a natural number, so his original axioms included the assertion that 1 had no immediate predecessor.

Proof. Let $A = \{x \in \mathbb{N} : x^+ \neq x\}$. Now, 0^+ is a successor and is therefore distinct from 0 by Axiom Z. Hence $0 \in A$. Suppose $x \in A$. Then $x^+ \neq x$. Therefore $(x^+)^+ \neq x^+$ by the contrapositive of Axiom U; so $x^+ \in A$. By induction, $A = \mathbb{N}$. □

We can spell out Axiom I more elaborately thus: *For every unary relation P on $\mathbb{N}$, in order to prove $\mathbb{N} \models \forall x\ P(x)$, it is enough to prove two things:*

(∗) $\mathbb{N} \models P(0)$ *(the* ***base step****);*

(†) $\mathbb{N} \models \forall x\ (P(x) \to P(x^+))$ *(the* ***inductive step****), that is, $P(x^+)$ is true under the assumption that x is a natural number and $P(x)$ is true.*

In the inductive step of a proof, the assumption that $x \in \mathbb{N}$ and $\mathbb{N} \models P(x)$ is called the **inductive hypothesis**. In the proof of Lemma 3.4, the full inductive hypothesis was not needed; only $x \in \mathbb{N}$ was needed.

4 Binary operations on natural numbers

To able to say much more about the natural numbers, we should introduce the usual arithmetic operations. But how? We do not need new axioms; the axioms that we already have are enough to enable us to *define* the arithmetic operations.

Let's start with **addition**. This is a binary operation $+$ on $\mathbb{N}$ whose values can be arranged in an (infinite) matrix as follows, in which $m+n$ is the entry (m,n), that is, the entry in row m and column n, the counts starting at 0:

$$\begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \\ 1 & 2 & 3 & 4 & \\ 2 & 3 & 4 & 5 & \\ 3 & 4 & 5 & 6 & \\ \vdots & & & & \ddots \end{array}$$

Then row m of this matrix is the sequence of values of a unary operation f_m on $\mathbb{N}$ such that $f_m(0) = m$ and $f_m(n^+) = f_m(n)^+$ for all n in $\mathbb{N}$. So we can *define* $m + n$ as $f_m(n)$. To do this rigorously, we need to know two facts:

(∗) that the functions f_m exist (so that an addition can be defined); and

(†) that the f_m are unique (so that there is only one addition).

Each of these facts is established by induction, as follows:

4.1 Theorem. *There is a unique binary operation $+$ on $\mathbb{N}$ such that $x + 0 = x$ and*

$$x + y^+ = (x + y)^+$$

for all x and y in $\mathbb{N}$.

Proof. Let A be the set of natural numbers x for which there is a unary operation f_x on $\mathbb{N}$ such that $f_x(0) = x$ and

$$f_x(y^+) = f_x(y)^+$$

for all y in $\mathbb{N}$. We can define f_0 by

$$f_0(y) = y.$$

So $0 \in A$. Suppose $x \in A$. Define f_{x^+} by

$$f_{x^+}(y) = f_x(y)^+.$$

Then $f_{x^+}(0) = f_x(0)^+ = x^+$, and

$$f_{x^+}(y^+) = f_x(y^+)^+ = (f_x(y)^+)^+ = f_{x^+}(y)^+;$$

so $x^+ \in A$. By induction, $A = \mathbb{N}$. This establishes the *existence* of the desired operation $+$, since we can define $x + y = f_x(y)$.

For the uniqueness of $+$, it is enough to note the uniqueness of the functions f_x. If f'_x has the properties of f_x, then $f'_x(0) = x = f_x(0)$, and if $f'_x(y) = f_x(y)$, then $f'_x(y^+) = f'_x(y)^+ = f_x(y)^+ = f_x(y^+)$. By induction, $f'_x = f_x$. □

4.2 Lemma. $\mathbb{N}$ *satisfies*

(∗) $\forall x\ 0 + x = x$,

(†) $\forall x\ \forall y\ y^+ + x = (y + x)^+$.

4.3 Exercise. Prove the lemma. (For part (†), this can be done by showing $\mathbb{N} = \{x : \forall y\ y^+ + x = (y + x)^+\}$.)

4.4 Theorem. $\mathbb{N}$ *satisfies*

(‡) $\forall x\ x^+ = x + 1$,

(§) $\forall x\ \forall y\ x + y = y + x$ *[that is, + is* ***commutative****]*,

(¶) $\forall x\ \forall y\ \forall z\ (x + y) + z = x + (y + z)$ *[that is, + is* ***associative****]*.

4.5 Exercise. Prove the theorem.

We can uniquely define **multiplication** on $\mathbb{N}$ just as we did addition: We can show that the multiplication-table

$$\begin{array}{ccccc} 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 2 & 3 & \\ 0 & 2 & 4 & 6 & \\ 0 & 3 & 6 & 9 & \\ \vdots & & & & \ddots \end{array}$$

can be written in exactly one way:

4.6 Theorem. *There is a unique binary operation $\cdot$ on $\mathbb{N}$ such that $x \cdot 0 = 0$ and*

$$x \cdot y^+ = x \cdot y + x$$

for all x and y in $\mathbb{N}$.

4.7 Exercise. Prove the theorem.

Multiplication is also indicated by juxtaposition, so that $x \cdot y$ is xy.

4.8 Lemma. $\mathbb{N}$ *satisfies*

($*$) $\forall x\ 0x = 0$,

($\dagger$) $\forall x\ \forall y\ y^+x = yx + x$.

4.9 Exercise. Prove the lemma.

4.10 Theorem. $\mathbb{N}$ *satisfies*

($\ddagger$) $\forall x\ 1x = x$,

($\S$) $\forall x\ \forall y\ xy = yx$ *[that is, $\cdot$ is commutative],*

($\P$) $\forall x\ \forall y\ \forall z\ (x + y)z = xz + yz$ *[that is, $\cdot$ **distributes** over $+$],*

($\|$) $\forall x\ \forall y\ \forall z\ (xy)z = x(yz)$ *[that is, $\cdot$ is associative].*

4.11 Exercise. Prove the theorem.

In establishing addition and multiplication as operations with the familiar properties, we used only that $\mathbb{N}$ satisfies the Induction Axiom. Other structures satisfy this axiom as well, so they too have addition and multiplication:

4.12 Example. Let $A = \{0, 1, 2\}$, and define $s : A \to A$ by

x	0	1	2
$s(x)$	1	2	0

Then $(A, s, 0)$ satisfies Axiom I, so it must have addition and multiplication—which in fact are given by the matrices

$$\begin{matrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{matrix} \quad \text{and} \quad \begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{matrix}$$

But $(A, s, 0)$ does not satisfy Axiom Z.

If a structure satisfies Axiom I, we may say that the structure **admits (proof by) induction**. So all structures that admit induction have unique operations of addition and multiplication with the properties given above.

Exponentiation on $\mathbb{N}$ is a binary operation $(x, y) \mapsto x^y$ whose values compose the matrix

$$\begin{array}{ccccc} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 1 & \\ 1 & 2 & 4 & 8 & \\ 1 & 3 & 9 & 27 & \\ \vdots & & & & \ddots \end{array}$$

The formal properties are that $x^0 = 1$ and

$$x^{y^+} = x^y \cdot x$$

for all x and y in $\mathbb{N}$. By induction, there can be no more than one such operation:

4.13 Exercise. Prove this.

Nonetheless, we shall need more than induction to prove that such an operation exists at all:

4.14 Example. In the induction-admitting structure $(A, s, 0)$ of Example 4.12, if we try to define exponentiation, we get $2^0 = 1$, $2^1 = 2$, $2^2 = 1$, $2^{s(2)} = 2^2 \cdot 2 = 2$; but $s(2) = 0$, so $2^{s(2)} = 2^0 = 1$. Since $1 \neq 2$, our attempt fails.

For any x in $\mathbb{N}$, we want to define $y \mapsto x^y$ as an operation g such that $g(0) = 1$, and $g(n^+) = g(n) \cdot x$. We have just seen that induction is *not* enough to allow us to do this. In the next section, we shall see that *recursion* is enough, and that this is equivalent to Axiom Z, Axiom U and Axiom I together.

5 Recursion

We want to be able to define functions g on $\mathbb{N}$ by specifying $g(0)$ and by specifying how to obtain $g(n^+)$ from $g(n)$. The next theorem is that we can do this. The proof is difficult, but the result is powerful:

5.1 Theorem (Recursion). *Suppose B is a set with an element c. Suppose f is a unary operation on B. Then there is a* unique *function $g : \mathbb{N} \to B$ such that $g(0) = c$ and*

$$g(x^+) = f(g(x)) \tag{$*$}$$

for all x in $\mathbb{N}$.

Proof. Let $\mathcal{S}$ be the set whose members are the subsets R of $\mathbb{N} \times B$ that have the following two properties:

(†) $(0, c) \in R$;

(‡) $(x, t) \in R \implies (x^+, f(t)) \in R$, for all (x, t) in $\mathbb{N} \times B$.

So the members of $\mathcal{S}$ have the properties required of g, except perhaps the property of being a function on $\mathbb{N}$.

The set $\mathcal{S}$ is non-empty, since $\mathbb{N} \times B$ itself is in $\mathcal{S}$. Let g be the intersection $\bigcap \mathcal{S}$. Then $g \in \mathcal{S}$ (why?).

We shall show that g is a function with domain $\mathbb{N}$. To do this, we shall show by induction that, for all x in $\mathbb{N}$, there is a unique t in B such that $(x, t) \in g$.

For the base step of our induction, we note first that $(0, c) \in g$. To finish the base step, we shall show that, for every t in $\mathbb{N}$, if $(0, t) \in g$, then $t = c$. Suppose $t \neq c$. Then neither property (†) nor property (‡) requires $(0, t)$ to be in a given member of $\mathcal{S}$. That is, if $R \in \mathcal{S}$, then $R \smallsetminus \{(0, t)\}$ still has these two properties; so, this set is in $\mathcal{S}$. In particular, $g \smallsetminus \{(0, t)\} \in \mathcal{S}$. But g is the smallest member of $\mathcal{S}$, so

$$g \subseteq g \smallsetminus \{(0, t)\},$$

which means $(0, t) \notin g$. By contraposition, the base step is complete.

As an inductive hypothesis, let us suppose that $x \in \mathbb{N}$ and that there is a unique t in B such that $(x, t) \in g$. Then $(x^+, f(t)) \in g$. To complete our inductive step, we shall show that, for every u in B, if $(x^+, u) \in g$, then $u = f(t)$. There are two possibilities for u:

(§) If $(x^+, u) = (y^+, f(v))$ for some (y, v) in g, then $x^+ = y^+$, so $x = y$ by Axiom U; this means $(x, v) \in g$, so $v = t$ by inductive hypothesis, and therefore $u = f(v) = f(t)$.

(¶) If $(x^+, u) \neq (y^+, f(v))$ for any (y, v) in g, then (as in the base step) $g \smallsetminus \{(x^+, u)\} \in \mathcal{S}$, so $g \subseteq g \smallsetminus \{(x^+, u)\}$, which means $(x^+, u) \notin g$.

Therefore, if $(x^+, u) \in g$, then $(x^+, u) = (y^+, f(v))$ for some (y, v) in g, in which case $u = f(t)$. Therefore $f(t)$ is unique such that $(x^+, f(t)) \in g$.

Our induction is now complete; by Axiom I, we may conclude that g is a function on $\mathbb{N}$ with the required properties (†) and (‡). If h is also such a function, then $h \in \mathcal{S}$, so $g \subseteq h$, which means $g = h$ since both are functions on $\mathbb{N}$. So g is unique. □

5.2 Exercise. If g and $\mathcal{S}$ are as in the proof of the Recursion Theorem, prove that $g \in \mathcal{S}$.

Equation $(*)$ in the statement of Theorem 5.1 is depicted in the following diagram:

$$\begin{tikzcd} \mathbb{N} \arrow[r, "+"] \arrow[d, "g"'] & \mathbb{N} \arrow[d, "g"] \\ B \arrow[r, "f"'] & B \end{tikzcd}$$

From the $\mathbb{N}$ on the left to the B on the right, there are two different routes, but each one yields the same result.

A **definition by recursion** is a definition of a function on $\mathbb{N}$ that is justified by Theorem 5.1. Informally, we can define such a function g by specifying $g(0)$ and by specifying how $g(x^+)$ is obtained from $g(x)$.

5.3 Remark. Sections 6 and 8 will provide several important examples of recursive definitions.

5.4 Theorem. *The Induction Axiom is a logical consequence of the Recursion Theorem.*

Proof. Suppose $A \subseteq \mathbb{N}$, and $0 \in A$, and $x^+ \in A$ whenever $x \in A$. Using the Recursion Theorem alone, we shall show $A = \mathbb{N}$.

Let $\mathbb{B} = \{0, 1\}$, and define a function $g_0 : \mathbb{N} \to \mathbb{B}$ by the rule

$$g_0(x) = \begin{cases} 0, & \text{if } x \in A; \\ 1, & \text{if } x \in \mathbb{N} \smallsetminus A. \end{cases}$$

Then g_0 is a function $g : \mathbb{N} \to \mathbb{B}$ such that $g(0) = 0$ and $g(x^+) = g(x)$ for all x in $\mathbb{N}$ (why?). But the function g_1 such that $g_1(x) = 0$ for all x in $\mathbb{N}$ is also such a function g. By the Recursion Theorem, there is only one such function g. Therefore $g_0 = g_1$, so $g_0(x)$ is never 1, which means $A = \mathbb{N}$. □

5.5 Exercise. Supply the missing detail in the proof.

However, there are models of the Induction Axiom which do not satisfy the Recursion Theorem:

5.6 Example. Again let $\mathbb{B} = \{0, 1\}$, and let $\neg$ be the unary operation on $\mathbb{B}$ such that $\neg 0 = 1$ and $\neg 1 = 0$. Then $(\mathbb{B}, \neg, 0)$ admits induction, but there is *no* function $g : \mathbb{B} \to \mathbb{N}$ such that $g(0) = 0$ and $g(\neg x) = (g(x)) + 1$ for all x in $\mathbb{B}$.

5.7 Remark. Apparently Peano himself did not recognize the distinction between proof by induction and definition by recursion; see the discussion in Landau [11, p. x]. Burris [1, p. 391] does not acknowledge the distinction. Stoll [17, p. 72] uses the term 'definition by weak recursion', although he observes that the validity of such a definition does *not obviously* follow from the Induction Axiom. However, Stoll does not *prove* (as we have done in Example 5.6) that the Induction Axiom is consistent with the negation of the Recursion Theorem.

5.8 Remark. The structure $(\mathbb{B}, s, 0)$ in Example 5.6 also satisfies Axiom U, but not Axiom I. If we define $t : \mathbb{B} \to \mathbb{B}$ so that $t(x) = 1$ for each x in $\mathbb{B}$, then $(\mathbb{B}, t, 0)$ satisfies the Induction Axiom and Axiom Z, but not Axiom U. Later (see Remark 20.5) we shall have natural examples of structures satisfying Axiom Z and Axiom U, but not Induction. We shall also observe (in Remark 8.3) that Axiom U is a consequence of the Recursion Theorem.

5.9 Exercise. Prove that Axiom Z is a consequence of the Recursion Theorem.

6 Binary operations by recursion

The Recursion Theorem guarantees the existence of certain *unary* functions on $\mathbb{N}$. As in Theorem 4.1, we can get the binary operation of addition by obtaining the unary operations $y \mapsto x + y$. By recursion, we can define addition as the unique operation such that

$$x + 0 = x \wedge x + y^+ = (x + y)^+$$

for all x and y in $\mathbb{N}$. In the same way, we can define multiplication by

$$x \cdot 0 = 0 \land x \cdot y^+ = x \cdot y + x.$$

The definition of exponentiation can follow this pattern:

6.1 Definition. The binary operation $(x, y) \mapsto x^y$ on $\mathbb{N}$ is given by:

$$x^0 = 1 \land x^{y^+} = x^y \cdot x. \tag{$*$}$$

In fact, we have something a bit more general. A **monoid** is a structure $(A, \cdot, 1)$ in which $\cdot$ is associative, and $a \cdot 1 = a = 1 \cdot a$ for all a in A. The monoid is **commutative** if $\cdot$ is commutative.

6.2 Theorem. *Suppose $\mathfrak{A}$ is a monoid. For every y in $\mathbb{N}$, there is a unique operation $x \mapsto x^y$ on A such that $(*)$ holds for all x in A and all y in $\mathbb{N}$.*

Proof. Let c be the operation $x \mapsto 1$ on A, let B be the set of unary operations on A, and let f be the operation

$$h \longmapsto (x \mapsto h(x) \cdot x)$$

on B. By recursion, there is a function $g : \mathbb{N} \to B$ such that $g(0) = c$ and $g(y^+) = f(g(y))$ for all y in $\mathbb{N}$. Now define $x^y = g(y)(x)$. $\square$

6.3 Theorem. *For all x and w in a commutative monoid, and for all y and z in $\mathbb{N}$, the following hold:*

- $(*)$ $x^{y+z} = x^y x^z$;
- $(\dagger)$ $(x^y)^z = x^{yz}$;
- $(\ddagger)$ $(xw)^z = x^z w^z$.

6.4 Exercise. Prove the theorem.

The **binomial coefficient** $\binom{m}{n}$ is entry (m, n) in the following matrix:

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & 0 & \\ 1 & 2 & 1 & 0 & 0 & \\ 1 & 3 & 3 & 1 & 0 & \\ 1 & 4 & 6 & 4 & 1 & \\ \vdots & & & & & \ddots \end{matrix}$$

We can give a formal definition by recursion:

6.5 Definition. The binary operation $(x, y) \mapsto \binom{x}{y}$ on $\mathbb{N}$ is given by:

$$\binom{x}{0} = 1 \land \binom{0}{y^+} = 0 \land \binom{x^+}{y^+} = \binom{x}{y} + \binom{x}{y^+} \tag{$\dagger$}$$

6.6 Exercise. Show precisely that this is a valid definition by recursion.

As with exponentiation, we can define the binomial coefficients in a more general setting. The proof uses the same technique as the proof of the Recursion Theorem:

6.7 Theorem. *For any structure $(A, {}^{+}, 0)$ that satisfies Axiom U and admits induction, for every y in $\mathbb{N}$, there is a unique operation $x \mapsto \binom{x}{y}$ on A such that*(†) *holds for all x in A and all y in $\mathbb{N}$.*

Proof. Let c be the operation $x \mapsto 1$ on A, and let B be the set of unary operations on A. We first prove that, for every h in B, there is a unique operation $f(h)$ in B given by

$$f(h)(0) = 0 \wedge f(h)(x^{+}) = h(x) + f(h)(x).$$

Say $h \in B$, and let $\mathcal{S}$ be the set whose members are the subsets R of $A \times A$ such that:

(∗) $(0, 0) \in R$;

(†) $(x, t) \in R \implies (x^{+}, h(x) + t) \in R$, for all (x, t) in $A \times A$.

Then $\bigcap \mathcal{S}$ is the desired operation $f(h)$. (Why?) By recursion, there is a function $g : \mathbb{N} \to B$ such that $g(0) = c$ and $g(y^{+}) = f(g)(y)$ for all y in $\mathbb{N}$. Now define $\binom{x}{y} = g(y)(x)$. □

6.8 Exercise. Supply the missing detail in the proof.

6.9 Exercise. Prove that $\binom{x}{1} = x$ for all x in $\mathbb{N}$.

See also Exercises 9.14 and 9.15.

In the proof of the last theorem, it was essential that the successor-operation on A be injective:

6.10 Example. Let $A = \{0, 1, 2\}$, and define s on A by

x	0	1	2
$s(x)$	1	2	1

If we attempt to define $(x, y) \to \binom{x}{y}$ on $A \times \mathbb{N}$, we get a matrix

$$\begin{matrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{matrix}$$

That is, $\binom{1}{2}$ should be both 0 and 1. So our attempt fails.

7 The integers and the rational numbers

Arithmetic on the integers is determined by arithmetic on the natural numbers. Given $\mathbb{N}$, we could just *define* the negative integers by somehow attaching minus-signs. A neater approach is motivated as follows.

For each natural number a, we want there to be an integer x such that

$$0 = a + x.$$

Then for each b in $\mathbb{N}$, we should have

$$b = a + b + x.$$

By these equations, the pairs $(0, a)$ and $(b, a + b)$ determine the same integer; so we can define integers to be equivalence-classes of such pairs.

7.1 Lemma. *On $\mathbb{N} \times \mathbb{N}$, let $\sim$ be the relation given by*

$$(a, b) \sim (c, d) \iff a + d = b + c.$$

Then $\sim$ is an equivalence-relation. If $(a_0, b_0) \sim (a_1, b_1)$ and $(c_0, d_0) \sim (c_1, d_1)$, then

(∗) $(a_0 + c_0, b_0 + d_0) \sim (a_1 + c_1, b_1 + d_1)$;

(†) $(b_0, a_0) \sim (b_1, a_1)$;

(‡) $(a_0c_0 + b_0d_0, b_0c_0 + a_0d_0) \sim (a_1c_1 + b_1d_1, b_1c_1 + a_1d_1)$.

7.2 Exercise. Prove the lemma. (For part (‡), show that each member is equivalent to $(a_1c_0 + b_1d_0, b_1c_0 + a_1d_0)$.)

7.3 Definition. Let $\sim$ be as in Lemma 7.1. We define $\mathbb{Z}$ to be $\mathbb{N} \times \mathbb{N}/\sim$. Let the $\sim$-class of (a, b) be denoted

$$a - b.$$

By Lemma 7.1, we can define the operations $+$, $-$ and $\cdot$ on $\mathbb{Z}$ by the following rules, where $a, b, c, d \in \mathbb{N}$:

(∗) $(a - b) +^{\mathbb{Z}} (c - d) = (a +^{\mathbb{N}} c) - (b +^{\mathbb{N}} d)$;

(†) $-^{\mathbb{Z}}(a - b) = b - a$;

(‡) $(a - b) \cdot^{\mathbb{Z}} (c - d) = (a \cdot^{\mathbb{N}} c +^{\mathbb{N}} b \cdot^{\mathbb{N}} d) - (b \cdot^{\mathbb{N}} c +^{\mathbb{N}} a \cdot^{\mathbb{N}} d)$.

Note that, by the definition, an integer like $5 - 3$ is *not* the natural number 2; it is not a natural number at all; it is the equivalence-class

$$\{(2, 0), (3, 1), (4, 2), (5, 3), \dots\},$$

which is $\{(x, y) \in \mathbb{N}^2 : x = y + 2\}$.

7.4 Theorem. *The function $x \mapsto x - 0 : \mathbb{N} \to \mathbb{Z}$ is injective and preserves $+$ and $\cdot$, that is,*

(∗) $(x +^{\mathbb{N}} y) - 0 = (x - 0) +^{\mathbb{Z}} (y - 0)$;

(†) $(x \cdot^{\mathbb{N}} y) - 0 = (x - 0) \cdot^{\mathbb{Z}} (y - 0)$

for all x and y in $\mathbb{N}$. On $\mathbb{Z}$, addition and multiplication are commutative and associative, and multiplication distributes over addition. Finally,

$$x +^{\mathbb{Z}} (-^{\mathbb{Z}} x) = 0 - 0$$

for all x in $\mathbb{Z}$.

7.5 Exercise. Prove the theorem.

7.6 Definition. On $\mathbb{Z}$, define the binary operation $-$ by

$$x -^{\mathbb{Z}} y = x +^{\mathbb{Z}} (-^{\mathbb{Z}} y).$$

7.7 Lemma. *If $x, y \in \mathbb{N}$, then the integer $x - y$ is $(x - 0) -^{\mathbb{Z}} (y - 0)$.*

Now we can identify the natural numbers with their images in $\mathbb{Z}$, considering the natural number x to be equal to the integer $x - 0$.

We can define the **rational numbers** similarly:

7.8 Lemma. *On $\mathbb{Z} \times (\mathbb{Z} \smallsetminus \{0\})$, let $\sim$ be the relation given by*

$$(a, b) \sim (c, d) \iff ad = bc.$$

Then $\sim$ is an equivalence-relation. If $(a_0, b_0) \sim (a_1, b_1)$ and $(c_0, d_0) \sim (c_1, d_1)$, then

(∗) $(a_0 d_0 \pm b_0 c_0, b_0 d_0) \sim (a_1 d_1 \pm b_1 c_1, b_1 d_1)$;

(†) $(a_0 c_0, b_0 d_0) \sim (a_1 c_1, b_1 d_1)$;

(‡) $(b_0, a_0) \sim (b_1, a_1)$ *and* $(0, a_0) \sim (0, 1)$ *if* $a_0 \neq 0$.

7.9 Exercise. Prove the lemma.

7.10 Definition. Let $\sim$ be as in Lemma 7.8. We define $\mathbb{Q}$ to be $\mathbb{Z} \times (\mathbb{Z} \smallsetminus \{0\})/{\sim}$. Let the $\sim$-class of (a, b) be denoted

$$\frac{a}{b}$$

or a/b. By Lemma 7.8, we can define the operations $+$, $-$ and $\cdot$ on $\mathbb{Q}$, and $x \mapsto x^{-1}$ on $\mathbb{Q} \smallsetminus \{0/1\}$, by the following rules, where $a, b, c, d \in \mathbb{Z}$:

(∗) $a/b \pm c/d = (ad \pm bc)/bd$;

(†) $(a/b)(c/d) = ac/bd$;

(‡) $(a/b)^{-1} = b/a$ if $a \neq 0$.

7.11 Theorem. *The function* $x \mapsto x/1 : \mathbb{Z} \to \mathbb{Q}$ *is injective and preserves* $+$, $-$ *and* $\cdot$. *On* $\mathbb{Q}$, *addition and multiplication are commutative and associative, and multiplication distributes over addition. Finally,*

$$x \cdot x^{-1} = \frac{1}{1}$$

for all x *in* $\mathbb{Q}$.

7.12 Exercise. Prove the theorem.

Now we can identify the integers with their images in $\mathbb{Q}$, considering the integer x to be equal to the rational number $x/1$.

8 Recursion generalized

How can we define n-**factorial**, $(n!)$? Informally, we write

$$n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n.$$

For a formal recursive definition, we can try

$$0! = 1 \wedge (x^+)! = x^+ \cdot x! \tag{$*$}$$

—but for this to be valid by the Recursion Theorem, we need an operation f on $\mathbb{N}$ so that $f(x!) = (x^+ \cdot x!)$. Such an operation exists, but it is not clear how we can define it before we have defined $(x!)$.

The definition $(*)$ is valid by the following:

8.1 Theorem (Recursion with parameter). *Suppose* B *is a set with an element* c. *Suppose* F *is a function from* $\mathbb{N} \times B$ *to* B. *Then there is a* unique *function* $G : \mathbb{N} \to B$ *such that* $G(0) = c$ *and*

$$G(x^+) = F(x, G(x)) \tag{$\dagger$}$$

for all x *in* $\mathbb{N}$.

Proof. Let f be the function

$$(x, b) \longmapsto (x^+, F(x, b)) : \mathbb{N} \times B \longrightarrow \mathbb{N} \times B.$$

By recursion, there is a unique function g from $\mathbb{N}$ to $\mathbb{N} \times B$ such that $g(0) = (0, c)$ and

$$g(x^+) = f(g(x))$$

for all x in $\mathbb{N}$. Now let G be $\pi \circ g$, where π is the function

$$(x, b) \longmapsto b : \mathbb{N} \times B \longrightarrow B.$$

Then for each x in $\mathbb{N}$ we have $g(x) = (y, G(x))$ for some y in $\mathbb{N}$. We can prove by induction that $y = x$. Indeed, this is the case when $x = 0$, since $g(0) = (0, c)$. Suppose $g(x) = (x, G(x))$ for some x in $\mathbb{N}$. Then

$$g(x^+) = f(x, G(x)) = (x^+, F(x, G(x))). \tag{$\ddagger$}$$

In particular, the first entry in the value of $g(x^+)$ is x^+. This completes our induction.

We now know that $g(x) = (x, G(x))$ for all x in $\mathbb{N}$. Hence in particular $g(x^+) = (x^+, G(x^+))$. But we also have (‡). Therefore we have (†), as desired. Finally, each of g and G determines the other. Since g is unique, so is G. □

8.2 Example. We can define a function f on $\mathbb{N}$ by requiring $f(0) = 0$ and $f(x^+) = x$. This is a valid recursive definition, by Theorem 8.1. Note that f picks out the immediate predecessor of a natural number, when this exists.

8.3 Remark. In the example, since f is unique, we see that Axiom U follows from the Recursion Theorem.

8.4 Definition. For any function $f : \mathbb{N} \to M$, where M is a set equipped with addition and multiplication, we define the sum $\sum_{k=0}^{n} f(k)$ and the product $\prod_{k=0}^{n} f(k)$ recursively as follows:

- $\displaystyle\sum_{k=0}^{0} f(k) = f(0)$ and $\displaystyle\sum_{k=0}^{n^+} f(k) = \sum_{k=0}^{n} f(k) + f(n^+)$;
- $\displaystyle\prod_{k=0}^{0} f(k) = f(0)$ and $\displaystyle\prod_{k=0}^{n^+} f(k) = \left(\prod_{k=0}^{n} f(k)\right) f(n^+)$.

8.5 Exercise. Prove the following.

(∗) $\sum_{k=0}^{n}(k+1) = (n^2 + 3n + 2)/2$

(†) $\sum_{k=0}^{n}(k+1)^2 = (2n^3 + 9n^2 + 13n + 6)/6$

(‡) $\sum_{k=0}^{n} b^k = (b^{n+1} - 1)/(b-1)$

(§) $\sum_{k=0}^{n}(2k+1) = (n+1)^2$

(¶) $\prod_{k=0}^{n}((k+1)/(k+2)) = 1/(n+2)$

9 The ordering of the natural numbers

In $\mathbb{N}$, if $x^+ = y$, then x is an immediate predecessor of y, and we know that x is unique. More generally, we should like to say that z is a *predecessor* of y if y is z^+, or $(z^+)^+$, or $((z^+)^+)^+$, or $(((z^+)^+)^+)^+$, or .... We can take care of the dots using recursion.

9.1 Definition. Let the function $x \mapsto \overline{x} : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ be given by the rule:

$$\bar{0} = \varnothing \land \forall x\ \overline{x^+} = \overline{x} \cup \{x\}.$$

The elements of $\overline{x}$ are the **predecessors of** x.

We shall prove in this section that the binary relation

$$\{(x, y) : x \in \overline{y}\}$$

on $\mathbb{N}$ is a strict total ordering. It will be important in § 12 that everything proved in this section is a consequence of just two facts:

- $\mathbb{N}$ admits induction.
- A function $x \mapsto \overline{x} : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ does exist as given by Definition 9.1.

We shall have to be precise with the relations $\in$ and $\subseteq$, which are *containment* and *inclusion* respectively. The relation $\subset$ is *proper* inclusion (the intersection of $\subseteq$ and $\neq$). We have:

$x \in A \iff$	x is an element of A	$\iff A$ contains x
$x \subseteq A \iff$	x is a subset of A	$\iff A$ includes x

We shall show first that $y \in \overline{x} \iff \overline{y} \subset \overline{x}$ for all x and y in $\mathbb{N}$.

9.2 Lemma. $\mathbb{N}$ *satisfies*

$$\forall y\ (y \in \overline{x} \to \overline{y} \subset \overline{x} \wedge \overline{y^+} \subseteq \overline{x}) \tag{$*$}$$

whenever $x \in \mathbb{N}$. Hence $\forall x\ x \notin \overline{x}$; also, the map $x \mapsto \overline{x}$ is injective.

Proof. The formula $(*)$ is satisfied when $x = 0$. Suppose it is satisfied when $x = z$. By contraposition, this means that, since $\overline{z} \not\subset \overline{z}$, we have $z \notin \overline{z}$. Therefore $\overline{z} \subset \overline{z^+}$. Say $y \in \overline{z^+}$. Then either $y \in \overline{z}$ or $y = z$. In the former case, $\overline{y} \subset \overline{z}$ by inductive hypothesis. Hence in either case, $\overline{y} \subseteq \overline{z}$. Therefore $\overline{y} \subset \overline{z^+}$; also, $\{y\} \subseteq \overline{z^+}$, so $\overline{y^+} \subseteq \overline{z}$. So $(*)$ holds when $x = z^+$. By induction, $(*)$ holds for all x in $\mathbb{N}$.

Since $\overline{x} \not\subset \overline{x}$, we have $x \notin \overline{x}$, again by the contrapositive of $(*)$. For the injectivity of $x \mapsto \overline{x}$, note first that $\overline{x} = \overline{0} \iff x = 0$. Suppose $\overline{x^+} = \overline{y^+}$, that is, $\overline{x} \cup \{x\} = \overline{y} \cup \{y\}$. Then either $y \in \overline{x}$ or $y = x$. In the first case, $\overline{y^+} \subseteq \overline{x} \subset \overline{x^+}$ (since $x \notin \overline{x}$), contradicting $\overline{x^+} = \overline{y^+}$; therefore $y = x$.

By Lemma 3.4 (whose proof uses only induction), we are done. □

9.3 Lemma. $\mathbb{N}$ *satisfies*

$$\forall y\ (\overline{y} \subset \overline{x} \to y \in \overline{x}) \tag{$\dagger$}$$

for each natural number x.

Proof. The formula $(\dagger)$ holds when $x = 0$. Suppose $(\dagger)$ is true when $x = z$. Say $y \in \mathbb{N}$ and $\overline{y} \subset \overline{z^+}$. Then $\overline{z^+} \not\subseteq \overline{y}$, so $z \notin \overline{y}$ by Lemma 9.2. Hence $\overline{y} \subseteq \overline{z}$. If $\overline{y} \subset \overline{z}$, then $y \in \overline{z}$ by inductive hypothesis. If $\overline{y} = \overline{z}$, then $y = z$, so $y \in \{z\}$. In either case, $y \in \overline{z^+}$. Thus $(\dagger)$ holds when $x = z^+$. □

9.4 Definition. If $x, y \in \mathbb{N}$, we write

$$x < y$$

instead of $x \in \bar{y}$; so $<$ is a binary relation on $\mathbb{N}$. We write

$$x \leqslant y$$

just in case $x < y \vee x = y$, equivalently, $\overline{x} \subseteq \overline{y}$.

To prove that $\leqslant$ has the properties we expect, we need a new proof-technique:

9.5 Theorem (Strong Induction). *If $A \subseteq \mathbb{N}$, then $\mathbb{N}$ satisfies*

$$\forall x\ (\overline{x} \subseteq A \to x \in A) \to \forall x\ x \in A. \tag{‡}$$

Proof. Suppose $A \subseteq \mathbb{N}$, and $\overline{x} \subseteq A \implies x \in A$ for all x in $\mathbb{N}$. We shall show that $\overline{x} \subseteq A$ for all x in $\mathbb{N}$. This is trivially true when $x = 0$, since $\overline{0} = \varnothing$. Suppose $\overline{z} \subseteq A$. Then $z \in A$ by assumption, so

$$\overline{z^+} = \overline{z} \cup \{z\} \subseteq A.$$

Hence, by induction, $\overline{x} \subseteq A$ for all x. In particular, $x \in \overline{x^+}$, but $\overline{x^+} \subseteq A$, so $x \in A$. □

As a consequence of the Strong Induction Theorem, we have the following method of proof. *For any unary relation P on $\mathbb{N}$, in order to prove $\mathbb{N} \models \forall x\ P(x)$, it is enough to prove one thing:*

(∗) $\mathbb{N} \models \forall x\ (\forall y\ (y < x \to P(y)) \to P(x))$, *that is, $x \in P$ under the assumption that x is a natural number and every predecessor of x is in P.*

The assumption that $x \in \mathbb{N}$ and $\forall y\ (y < x \to P(y))$ is the **strong inductive hypothesis**.

9.6 Theorem. $(\mathbb{N}, \leqslant)$ *is a total order.*

Proof. $(\mathbb{N}, \leqslant)$ is a partial order since

$$x \leqslant y \iff \overline{x} \subseteq \overline{y}$$

for all x and y in $\mathbb{N}$, and $x \mapsto \overline{x}$ is injective, by the preceding lemmas. It remains to show that $(\mathbb{N}, \leqslant)$ is a total order, equivalently, $\mathbb{N}$ satisfies

$$\forall y\ (\overline{y} \not\subseteq \overline{x} \to \overline{x} \subseteq \overline{y}). \tag{§}$$

We shall prove this by strong induction, that is, Theorem 9.5. Let A be the set of x in $\mathbb{N}$ such that (§) holds.

Suppose $\overline{z} \subseteq A$, that is, (§) holds whenever $x \in \overline{z}$. We shall show that (§) holds when $x = z$.

Suppose $\overline{y} \not\subseteq \overline{z}$; we shall show $\overline{z} \subseteq \overline{y}$. Say $x \in \overline{z}$. By strong inductive hypothesis, (§) holds. But $\overline{x} \subset \overline{z}$, by Lemma 9.2, so $\overline{y} \not\subseteq \overline{x}$, hence $\overline{x} \subseteq \overline{y}$ by (§). But $\overline{x} \neq \overline{y}$, so $\overline{x} \subset \overline{y}$, whence $x \in \overline{y}$. Thus $\overline{z} \subseteq \overline{y}$. Therefore (§) holds when $x = z$. By strong induction, the proof is complete. □

9.7 Lemma. $\mathbb{N} \models \forall x\ \forall y\ (x < y \to x^+ < y^+)$.

9.8 Exercise. Prove the lemma directly (without induction) using the previous lemmas.

9.9 Theorem. $\mathbb{N}$ *satisfies:*

(∗) $\forall x\ 0 \leqslant x$;

(†) $\forall x\,\forall y\,\forall z\,(x < y \leftrightarrow x + z < y + z)$;

(‡) $\forall x\,\forall y\,\forall z\,(x < y \to x \cdot z^+ < y \cdot z^+)$;

(§) $\forall x\,\forall y\,\exists z\,(x \leqslant y \leftrightarrow x + z = y)$.

9.10 Exercise. Prove the theorem.

9.11 Definition. If $x, y \in \mathbb{N}$, and $x \leqslant y$, then

$$y - x$$

is the natural number z (which exists and is unique by Theorem 9.9, parts (†) and (§)) such that $x + z = y$.

9.12 Exercise. Prove the following in $\mathbb{N}$.

(∗) $\forall x\,\forall y\;1 + xy \leqslant (1 + x)^y$

(†) $\forall x\,(3 < x \to x^2 < 2^x)$

9.13 Exercise. Find the flaw in the following argument, where max is the function from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$ such that $\max(x, y) = y$ if $x \leqslant y$, and otherwise $\max(x, y) = x$.

> If $\max(x, y) = 0$, then $x = y$. Suppose that $x = y$ whenever $\max(x, y) = n$. Suppose $\max(z, w) = n+1$. Then $\max(z-1, w-1) = n$, so $z - 1 = w - 1$ by inductive hypothesis; therefore $z = w$. Therefore all natural numbers are equal.

9.14 Exercise. Prove that, if $y \leqslant x$, then $\dbinom{x}{y} = \dfrac{x!}{y!\,(x-y)!}$.

9.15 Exercise. Prove the **Binomial Theorem**:

$$(x + y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i.$$

9.16 Exercise. If $x, y \in \mathbb{N}$, we write $x \mid y$ if $\exists z\; xz = y$; in this case we say that x is a **divisor** of y. A natural number is **prime** if its only divisors are 1 and itself, and these are distinct. Show that every natural number different from 1 has a prime divisor.

9.17 Definition. If $x \in \mathbb{N}$, then for the set $\overline{x}$, we may write

$$\{0, \dots, x - 1\}.$$

Here, the notation $x - 1$ has no independent meaning if $x = 0$; in this case, $\{0, \dots, x - 1\} = \varnothing$. For $\overline{x^+}$, we may write

$$\{0, \dots, x\}.$$

Similarly, if G is a function on $\mathbb{N}$, then, recalling the definition on p. 14, we may write

$$G''\overline{x} = \{G(0), \dots, G(x - 1)\},$$
$$G''\overline{x^+} = \{G(0), \dots, G(x)\}.$$

In this notation, by strong induction, a subset A of $\mathbb{N}$ is equal to $\mathbb{N}$, provided

$$\{0, \dots, x-1\} \subseteq A \implies x \in A$$

for all x in A. This condition is logically *stronger*—harder to satisfy—than the condition

$$x \in A \implies x^+ \in A$$

in the Induction Axiom. To make this precise, first note that the following agrees with Definition 9.4 in case $(X, \leqslant)$ is $(\mathbb{N}, \leqslant)$:

9.18 Definition. If $(X, \leqslant)$ is a total order, and $x \in X$, let

$$\overline{x} = \{y \in X : y < x\}.$$

In this section, we have *used* the Strong-Induction Theorem to prove that $(\mathbb{N}, \leqslant)$ is a total order. But if we already have a total order, then we can say that it **admits (proof by) strong induction** if it satisfies (‡) of Theorem 9.5.

9.19 Theorem. *A structure $\mathfrak{A}$ that admits induction and has a total ordering admits strong induction, provided also that*

$$x < x^+$$

for all x in A.

9.20 Exercise. Prove the theorem.

Section 11 will introduce totally ordered sets in which strong induction works, but ordinary induction may not.

10 The real numbers

Recall from § 7 that every integer is a difference $x - y$ of two natural numbers, and every rational number is a quotient u/v of two integers.

10.1 Lemma. *There is a well-defined subset P of $\mathbb{Z}$ consisting of those differences $a - b$ of natural numbers a and b such that $b < a$. There is a unique strict linear ordering $<$ of $\mathbb{Z}$ such that*

$$x < y \iff y - x \in P$$

for all x and y in $\mathbb{Z}$. The embedding $x \mapsto x - 0 : \mathbb{N} \to \mathbb{Z}$ preserves $<$.

10.2 Lemma. *There is a well-defined subset P of $\mathbb{Q}$ consisting of those quotients a/b of integers a and b such that $0 < ab$. There is a unique strict linear ordering $<$ of $\mathbb{Q}$ such that*

$$x < y \iff y - x \in P$$

for all x and y in $\mathbb{Q}$. The embedding $x \mapsto x/1 : \mathbb{Z} \to \mathbb{Q}$ preserves $<$.

10.3 Definition. A **cut** of a linear order $(X, \leqslant)$ is a subset $\mathfrak{a}$ such that:

(∗) $\varnothing \subset \mathfrak{a} \subset X$;

(†) $x < y \wedge y \in \mathfrak{a} \implies x \in \mathfrak{a}$;

(‡) $\forall y\ (y < x \to y \in \mathfrak{a}) \implies x \in \mathfrak{a}$.

The set $\mathbb{R}$ of **real numbers** is the set of cuts of $\mathbb{Q}$.

10.4 Exercise. Define $+$, $\cdot$ and $<$ on $\mathbb{R}$. Show that the function $x \mapsto \{y \in \mathbb{Q} : y \leqslant x\} : \mathbb{Q} \to \mathbb{R}$ is an injection that preserves $+$, $\cdot$ and $<$.

If $a, b \in \mathbb{R}$, then $[a, b)$ is the set $\{x \in \mathbb{R} : a \leqslant x < b\}$.

10.5 Theorem. *Suppose* $\mathfrak{a}$ *in* $[0, 1)$*, there is a unique function* $n \mapsto a_n : \mathbb{N} \to \{0, 1\}$ *such that*

$$\sum_{k=0}^{n} \frac{a_k}{2^k} \leqslant a < \sum_{k=0}^{n} \frac{a_k}{2^k} + \frac{1}{n}.$$

10.6 Exercise. Prove the theorem.

11 Well-ordered sets

11.1 Definition. A total order is called **well-ordered** if every non-empty subset has a least element. The least element of a subset A can be denoted

$$\min A.$$

11.2 Definition. A total order $(X, \leqslant)$ **admits (definition by) strong recursion** if, for every set B and function $h : \mathcal{P}(B) \to B$, there is a unique function $G : X \to B$ such that

$$G(x) = h(G''\overline{x})$$

for all x in X.

11.3 Theorem. *The following are equivalent conditions on a total order:*

(∗) *It is well-ordered.*

(†) *It admits strong induction.*

(‡) *It admits strong recursion.*

Proof. Let $(X, \leqslant)$ be a total order.

Suppose $(X, \leqslant)$ is well-ordered. If $A \subset X$ and $x = \min(X \smallsetminus A)$, then $\overline{x} \subseteq A$, but $x \notin A$. By contraposition, if $\overline{y} \subseteq A \implies y \in A$ for all y in X, then $A = X$; that is, $(X, \leqslant)$ admits strong induction.

Suppose $(X, \leqslant)$ admits strong induction. Say h is a function from $\mathcal{P}(B)$ to B. Let A be the subset of X consisting of those x for which there is a unique function $G_x : \overline{x} \cup \{x\} \to B$ such that

$$G_x(y) = h(G_x''\overline{y})$$

for all y in $\overline{x} \cup \{x\}$. Say $\overline{x} \subseteq A$. To show that $x \in A$, we can define $G_x : \overline{x} \cup \{x\} \to B$ by

$$G_x(y) = \begin{cases} G_y(y), & \text{if } y < x; \\ h(\{G_z(z) : z < x\}), & \text{if } y = x. \end{cases}$$

Then G_x is a function witnessing that $x \in A$ (why?). By strong induction, $A = X$. Now we can let G be $x \mapsto G_x(x) : X \to B$. This shows that $(X, \leqslant)$ admits strong recursion. (Why is G unique?)

Finally, suppose $(X, \leqslant)$ is *not* well-ordered. In particular, suppose $\varnothing \subset A \subseteq X$, but A has no least element. Let $\mathbb{B} = \{0, 1\}$, and let $h : \mathcal{P}(\mathbb{B}) \to \mathbb{B}$ be given by

$$h(x) = 1 \iff 1 \in x.$$

For each i in $\mathbb{B}$, let $G_i : X \to \mathbb{B}$ be given by

$$G_i(x) = \begin{cases} 0, & \text{if } x \notin A; \\ i, & \text{if } x \in A. \end{cases}$$

Then $G_i(x) = h(G_i''\overline{x})$ for all x in X. Since there are two functions G_i, the order $(X, \leqslant)$ does not admit strong recursion. □

11.4 Exercise. Supply the missing details in the last proof.

11.5 Remark. That X is a *set* is not used in the proof of the theorem. We shall later (in § 18) consider well-ordered *classes*.

11.6 Corollary. $(\mathbb{N}, \leqslant)$ *is well-ordered. In particular, suppose B is a set, and h is a function from $\mathcal{P}(B)$ to B. Then there is a unique function $G : \mathbb{N} \to B$ such that*

$$G(x) = h(\{G(0), \dots, G(x-1)\})$$

for all x in $\mathbb{N}$.

12 A model of the Peano axioms

We have assumed the existence of natural numbers that satisfy the Peano axioms. We have made no assumptions about each natural number in itself. Now we shall construct a *model* of the Peano axioms. We shall be able to describe each element of this model.

Definition 9.1 suggests a way to proceed. Why not define the natural numbers so that each one is *identical* to the set of its predecessors? We can do this recursively, provided that we have *some* model $\mathbb{N}$ of the Peano axioms. Indeed, we can define a function f on $\mathbb{N}$ by the rule

$$f(0) = \varnothing \wedge \forall x\, f(x^+) = f(x) \cup \{f(x)\}.$$

Then the range of f determines a model of the Peano axioms in which 0 is the empty set, the successor-operation is the map $x \mapsto x \cup \{x\}$, and each number *is* the set of its predecessors. The first five elements—namely 0, 1, 2, 3 and 4—of this model are:

$$\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}\}.$$

An alternative way to construct this model is to forget the Peano axioms and proceed as follows.

12.1 Definition. The **successor** of any set is the smallest set that contains and includes it. So, the successor of a set A is the set

$$A \cup \{A\}.$$

We shall denote this set by A'.

Thus, for the moment, the successor of a natural number and the successor of a set are called by the same word, but have different symbols.

We propose to assume:

12.2 Axiom (Infinity). *There is a set Ω ('Omega') of sets such that $\varnothing \in \Omega$, and $A' \in \Omega$ whenever $A \in \Omega$.*

It seems reasonable to assume that, given a set, we can always form its successor. After all, we can do this symbolically, as in Definition 12.1. The Axiom of Infinity is that we—or some being—can have started with the empty set, and can have repeatedly taken successors *until no more successors can be taken.* Expressed in these terms, the Axiom is a philosophically problematic assumption. Nonetheless, like most (though not all) mathematicians, we shall make this assumption.

In a more benign formulation, the Axiom of Infinity is just that some set Ω contains $\varnothing$ and includes the image of itself under the successor-operation $A \mapsto A'$. Then we can form the structure $(\Omega, ', \varnothing)$. There may be more than one such set Ω, but the intersection of such sets is still such a set. Hence there is a *smallest* such set. We give it a name:

12.3 Definition. We denote by

$$\omega$$

('omega') the smallest set of sets that contains $\varnothing$ and includes its own image under the successor-operation.

12.4 Exercise. Verify that there is exactly one set ω as given by the definition.

12.5 Lemma. *$(\omega, ', \varnothing)$ satisfies the Induction Axiom.*

12.6 Exercise. Prove the lemma.

12.7 Lemma. *Every element of ω is a subset of ω.*

12.8 Exercise. Prove the lemma.

12.9 Theorem. *$(\omega, ', \varnothing)$ is a model of the Peano axioms.*

Proof. It is obvious that $(\omega, ', \varnothing)$ satisfies Axiom Z. By the last lemma, we have a map, namely
$$x \longmapsto x : \omega \longrightarrow \mathcal{P}(\omega),$$
that takes x' to $x \cup \{x\}$. Therefore, by § 9, the structure $(\omega, \subseteq)$ is a total order, and (because of Lemma 9.7), the map $x \mapsto x'$ is injective, that is $(\omega, ')$ satisfies Axiom U. □

12.10 Exercise. Write a more detailed proof.

If there is one model of the Peano axioms, then there are others. We now have a notational distinction: $(\mathbb{N}, ^+, 0)$ is an arbitrary model of the axioms, but $(\omega, ', \varnothing)$ is the specific model that we have defined. If we want to be precise, we may refer to the elements of ω as the **von Neumann** natural numbers. In the rest of these notes though, natural numbers will always be von Neumann natural numbers. so we shall have $0 = \varnothing$, and $1 = \{0\}$, and so forth. (Also, in Definition 22.1, we shall give a new meaning to the symbol $^+$.)

In one sense, it doesn't matter *which* model of the Peano axioms we use:

12.11 Theorem. *Every model* $(\mathbb{N}, ^+, 0)$ *of the Peano axioms is uniquely isomorphic to* $(\omega, ', \varnothing)$*, that is, there is a unique bijection* $f : \mathbb{N} \to \omega$ *such that* $f(0) = \varnothing$*, and* $f(x^+) = f(x)'$ *for all* x *in* $\mathbb{N}$*.*

Proof. By recursion, there is a unique function f on $\mathbb{N}$ such that $f(0) = \varnothing$, and $\forall x\ f(x^+) = f(x)'$. then $f(0) \in \omega$, and if $f(x) \in \omega$, then $f(x^+) \in \omega$; so ω includes the range of f. For the same reason, there is a function $g : \omega \to \mathbb{N}$ such that $g(\varnothing) = 0$ and $g(x') = g(x)^+$. By induction, $g \circ f$ is the identity on $\mathbb{N}$, and $f \circ g$ is the identity on ω. So f is a bijection from $\mathbb{N}$ to ω. □

Nonetheless, as we have seen, the set of von Neumann natural numbers has the peculiar property that proper inclusion and containment are the same relation on it, and this relation is the relation $<$ induced by the Peano axioms.

Since a natural number is now a set of natural numbers, we must be careful with functional notation. Suppose for example that $f : \omega \to \omega$ is the doubling function, $x \mapsto 2 \cdot x$. Then $f(4) = 8$, but $f''4 = f''\{0, 1, 2, 3\} = \{0, 2, 4, 6\}$.

13 Numbers in ordinary language

We shall come to understand the von Neumann natural numbers both as cardinal and as ordinal numbers.

In ordinary languages like Turkish and English (or Latin and Greek), there is a one-to-one correspondence between the cardinal and the ordinal numbers.

Turkish constructs the ordinal numbers from the cardinals by adding the suffix `-(#)nc#`, where `#` is chosen from the set {`ı`, `i`, `u`, `ü`} according to the rules of vowel harmony. In English, the regular way to get the ordinals from the cardinals is to add -*(e)th*, but there are irregularities. We have:

English cardinal:	one	two	three	four	five
Türkçesi:	bir	iki	üç	dört	beş
its numeral:	1	2	3	4	5
related ordinal:	first	second	third	fourth	fifth
its abbreviation:	1st	2nd	3rd	4th	5th
Türkçesi:	birinci	ikinci	üçüncü	dördüncü	beşinci
kısaltması:	1.	2.	3.	4.	5.

Also, for example, from *twenty-one* (21) we get *twenty-first* (21st), although, historically, the cardinal has been written *one and twenty*, with corresponding ordinal *one-and-twentieth*.

There is evidently no formal connection between *one* and *first*, or between *two* and *second*. At its roots, *first* means *foremost*, that is, 'coming before everything else.' Indeed, the *fir-* in *first* is related to *fore* (as in *before*), and the *-st* of *first* and *most* is related to the suffix *-est* used to form regular superlatives like *biggest* and *soonest*. Also, *second* comes from the Latin SECVNDVS, meaning 'following'.

Thus, it would not do violence to English if we treated zero as the *first* natural number, and one as the second. But two as the *third* number might be strange. In any case, the word *zeroth* (or sıfırıncı) has been coined as a label for the position of zero on the list of numbers.

14 Natural numbers as cardinals

Cardinal numbers name the sizes of sets. Each natural number (that is, von Neumann natural number) is a set of a certain size. So we can use a natural number as a cardinal number for itself and for other sets of the same size. For such a convention to be most useful, we should make sure that different natural numbers have different sizes. To do this, we must be precise about what we mean by having the same or different sizes.

14.1 Definition. If A and B are sets, then we write:

($*$) $A \preccurlyeq B$, if there is an injection from A into B;

(†) $A \approx B$, if there is a bijection from A onto B;

(‡) $A \not\approx B$, if there is no bijection between A and B;

(§) $A \prec B$, if $A \preccurlyeq B$ and $A \not\approx B$.

We say that A and B have the **same size**, or are **equipollent**[15], if $A \approx B$; otherwise, A and B have different sizes. If $A \prec B$, then B is **strictly larger than** A.

14.2 Lemma. *On any set of sets, the relation $\approx$ is an equivalence-relation, and is a refinement of $\preccurlyeq$ (that is, $A \approx B \implies A \preccurlyeq B$). Also, $\preccurlyeq$ is reflexive and transitive.*

[15] That is, have 'equal power'.

14.3 Exercise. Prove the lemma.

Certainly $\preccurlyeq$ is not anti-symmetric, since $\{1\} \preccurlyeq \{0\}$, and $\{0\} \preccurlyeq \{1\}$, but $\{0\} \neq \{1\}$. We *shall* show later (in Theorem 16.1) that $\preccurlyeq$ is anti-symmetric on $\approx$-classes, that is,

$$A \preccurlyeq B \land B \preccurlyeq A \implies A \approx B.$$

However, this implication is not obvious. It *is* obvious that $A \subseteq B \implies A \preccurlyeq B$.

14.4 Lemma. *Distinct natural numbers have different sizes.*

Proof. By Theorem 9.9, it is enough to show that

$$\forall y\ (x \approx x + y \to y = 0)$$

for all x in ω. The claim is true if $x = 0$, since the only function on $\varnothing$ is the empty function. Suppose the claim is true when $x = z$. Say $f : z' \to z' + y$ is a bijection. Then so is $g : z \to z + y$, where

$$g(x) = \begin{cases} f(x), & \text{if } f(x) \neq z + y; \\ f(z), & \text{if } f(x) = z + y. \end{cases}$$

Hence $y = 0$ by inductive hypothesis. So the claim is true when $x = z'$. □

14.5 Theorem. *On ω, the relation $\preccurlyeq$ is the total ordering $\leqslant$.*

14.6 Exercise. Prove the theorem.

The following is now justified:

14.7 Definition. If $A \approx n$ for some n in ω, then we call n the **cardinality** of A, and we write $|A| = n$; we also call A a **finite** set. The natural numbers are the **finite cardinal numbers**.

Note that $|n| = n$ for all n in ω. We shall ultimately come up with a definition of $|A|$ for all sets A. This is not a trivial matter, since some sets are not finite:

14.8 Theorem. *Suppose $A \preccurlyeq B$. If B is finite, then A is finite.*

Proof. It is enough to show that if $n \in \omega$, and $A \subseteq n$, then A is finite.

The claim is trivially true if $n = 0$. Suppose it is true when $n = k$. Say $A \subseteq k'$. If $A = k'$, then A is finite by definition. If $A \subseteq k$, then A is finite by inductive hypothesis. In the remaining case, $k \in A$, but there is m in $k \smallsetminus A$. Then $A \cup \{m\} \smallsetminus \{k\} \subseteq k$, so the set is finite by inductive hypothesis. But A and $A \cup \{m\} \smallsetminus \{k\}$ are equipollent. □

14.9 Theorem. *A set A is finite if and only if $A \prec \omega$.*

Proof. Suppose A is finite; this means $A \approx n$ for some n in ω. Then $A \prec n+1 \preccurlyeq \omega$ by Theorem 14.5.

Now suppose A is not finite, but $A \preccurlyeq \omega$. We may assume $A \subseteq \omega$. If $x \in \omega$, then $A \not\subseteq x$, by the last theorem. Hence we can define $g : \omega \to A$ by

$$g(0) = \min A \land g(x') = \min(A \smallsetminus g(x)').$$

Then $g(x) < g(x')$ for all x in ω, so g is injective (why?). Therefore $\omega \preccurlyeq A$, so $A \approx \omega$. □

14.10 Exercise. Supply the missing detail in the proof.

The following is immediate:

14.11 Corollary. *ω is not finite.*

14.12 Theorem. *The union of two finite sets is finite. In fact, if A and B are finite, then*

$$|A \cup B| + |A \cap B| = |A| + |B|.$$

14.13 Exercise. Prove the theorem.

15 Infinite sets

Commonly, an infinite set is simply a non-finite set—a set that is not finite. However, another definition is preferable, for reasons to be mentioned presently.

15.1 Definition. A set is **infinite** if it is equipollent with a proper subset of itself.

15.2 Theorem. *Suppose $A \preccurlyeq B$. If A is infinite, then B is infinite.*

Proof. If $f : A \to B$ and $g : A \to A$ are injections, then so is $h : B \to B$, where

$$h(x) = f(g(y)),$$

if $x = f(y)$ for some y in A, and otherwise $h(x) = x$. If g is not surjective, then neither is h. □

Is Theorem 15.2 the contrapositive of Theorem 14.8? Or is there a set that is neither finite nor infinite, or that is both finite and infinite?

15.3 Lemma. *A set A is infinite if and only if it can be equipped with a unary operation s such that, for some a in A, the structure (A, s, a) is a model of Axiom Z and Axiom U.*

15.4 Exercise. Prove the lemma.

We immediately have:

15.5 Theorem. *ω is infinite.*

15.6 Lemma. *Any model (A, s, a) of Axiom Z and Axiom U has a substructure that is a model of all of the Peano axioms.*

15.7 Exercise. Prove the lemma.

By the last lemma and Theorem 12.11, we have:

15.8 Theorem. *A set A is infinite if and only if $\omega \preccurlyeq A$.*

15.9 Corollary. *No set is both finite and infinite.*

15.10 Exercise. Prove the theorem and corollary.

15.11 Remark. Lemmas 15.3 and 15.6 justify the name of the Axiom of Infinity. By this axiom, the infinite set ω exists. But if *any* infinite set exists, then a model of the Peano axioms exists; hence the specific model $(\omega, ', \varnothing)$ exists, as shown at the beginning of § 12. So, as we stated it, the Axiom of Infinity is equivalent to the assumption that some infinite set exists.

We can show that a set A is infinite if we can find an injective function $G : \omega \to A$. That G is injective means precisely that

$$G(x) \in A \smallsetminus \{G(0), \dots, G(x-1)\}$$

for all x in ω. Now, if A is *not* finite, then in each case the set

$$A \smallsetminus \{G(0), \dots, G(x-1)\}$$

is not empty, so there is some hope that the function G exists.

15.12 Definition. A **choice-function** for a set A is a function $f : \mathcal{P}(A) \to A$ such that

$$f(X) \in X$$

for all X in $\mathcal{P}(A) \smallsetminus \{\varnothing\}$.

15.13 Theorem. *If A has a choice-function, then $A \prec \omega \vee \omega \preccurlyeq A$.*

Proof. Suppose A has a choice-function f, but A is not finite. Let h be $X \mapsto f(A \smallsetminus X) : \mathcal{P}(A) \to A$. By strong recursion, there is $G : \omega \to A$ given by

$$G(x) = h(G''\overline{x}).$$

Then $G(x) \notin G''\overline{x}$, so G is injective. □

By the theorem, all sets are finite or infinite, provided we assume:

15.14 Axiom (Choice). *Every set has a choice-function.*

The Axiom of Choice is **AC** for short. There are a number of equivalent formulations, such as *Zorn's Lemma*. It is a remarkable result of twentieth-century mathematics that neither **AC** nor its negation **AC** is a consequence of our earlier axioms.

15.15 Exercise. If all sets are finite or infinite, do you think the Axiom of Choice follows?

I shall try to be explicit about when I use **AC**. For example:

15.16 Theorem. *Every set is either finite or infinite (assuming* **AC***).*

Using Theorem 15.8, we can prove:

15.17 Theorem. *If A is infinite, then $A' \approx A$.*

Proof. Let $f : \omega \to A$ be an injection. Define $g : A' \to A$ given by:

$$g(x) = \begin{cases} f(0), & \text{if } x = A; \\ x, & \text{if } x \in A \smallsetminus f''\omega; \\ f(f^{-1}(x)+1), & \text{if } x \in f''\omega. \end{cases}$$

Then g is a bijection. □

Is the converse of this theorem true? It is, *if* a set is always a proper subset of its successor. Suppose if possible that $A = \{A\}$. Then $|A| = 1$, but $A = A'$. By the following, such sets do not exist:

15.18 Axiom (Foundation). *Every non-empty set A contains a set X such that $A \cap X = \varnothing$.*

15.19 Theorem. *If $A' \approx A$, then A is infinite.*

Proof. By the Foundation Axiom, $\{A\} \cap A = \varnothing$, so $A \notin A$, which means $A \subset A'$. □

16 The ordering of cardinalities

Before defining $|A|$ for sets A in general, we can still write

$$|A| = |B| \tag{$*$}$$

instead of $A \approx B$. We can think of $|A|$ as *something*, if only an $\approx$-class of sets; so in $(*)$ we can call $|A|$ the *cardinality* of A. By Lemma 14.2, the relation $\preccurlyeq$ induces a relation on cardinalities, so we can write

$$|A| \leqslant |B|$$

instead of $A \preccurlyeq B$.

We haven't yet proved that $\preccurlyeq$ is a partial ordering of the cardinalities. This we now do.

The appropriate name of the following is uncertain:

16.1 Theorem (Schröder–Bernstein). $A \preccurlyeq B \land B \preccurlyeq A \implies A \approx B$ *for all sets A and B.*

Proof. Suppose $f : A \to B$ and $g : B \to A$ are injections. We recursively define a function

$$n \mapsto (A_n, B_n) : \omega \to \mathcal{P}(A) \times \mathcal{P}(B)$$

by requiring $(A_0, B_0) = (A, B)$, and $(A_{n+1}, B_{n+1}) = (g''B_n, f''A_n)$. Since f and g are injective, we have

$$f''(A_n \smallsetminus A_{n+1}) = f''A_n \smallsetminus f''A_{n+1} = B_{n+1} \smallsetminus B_{n+2},$$

and likewise $g''(B_n \smallsetminus B_{n+1}) = A_{n+1} \smallsetminus A_{n+2}$. Also

$$f''\bigcap\{A_n : n \in \omega\} = \bigcap\{B_{n+1} : n \in \omega\}.$$

Now define $h : A \to B$ by

$$h(x) = \begin{cases} f(x), & \text{if } x \in A_{2n} \smallsetminus A_{2n+1}; \\ g^{-1}(x), & \text{if } x \in A_{2n+1} \smallsetminus A_{2n+2}; \\ f(x), & \text{if } x \in \bigcap\{A_n : n \in \omega\}. \end{cases}$$

Then h is a bijection. □

So $\preccurlyeq$ is anti-symmetric on cardinalities, that is,

$$|A| \leqslant |B| \wedge |B| \leqslant |A| \implies |A| = |B|.$$

By Lemma 14.2 then, $\preccurlyeq$ induces a partial ordering of cardinalities.

From the preceding sections, we know that if A is finite, and B is infinite, then A' is finite, and

$$A \prec A' \prec \omega \preccurlyeq B.$$

So $|\omega|$ is the least infinite cardinality, and is a least upper bound for the finite cardinalities.

Does $\preccurlyeq$ induce a total ordering of cardinalities? We shall ultimately (with Theorems 18.13 and 21.1) show that it does, by **AC**. First we shall show how to produce, from given sets, strictly larger sets. Taking Cartesian products does not generally accomplish this:

16.2 Exercise. If $A \preccurlyeq \omega$ and $B \approx \omega$, show that $A \times B \approx \omega$.

17 Uncountable sets

17.1 Definition. If A and B are sets, then

$${}^{B}A$$

is the set of functions from B to A.

17.2 Lemma. $\mathcal{P}(A) \approx {}^{A}2$ *for all sets* A.

Proof. The function $f \mapsto \{x \in A : f(x) = 1\} : {}^{A}2 \to \mathcal{P}(A)$ is a bijection. □

We now have several unary operations on sets:

(∗) the constant-map $A \mapsto \varnothing$,

(†) the successor-map $A \to A'$,

(‡) the power-set operation $A \mapsto \mathcal{P}(A)$,

(§) the map $A \mapsto {}^{A}A$, and

(¶) the maps $A \mapsto {}^{B}A$ and $A \mapsto {}^{A}B$, where B is a set fixed in advance.

If the sets are hereditary (as ours are), then we also have $A \mapsto \bigcup A$, and $A \mapsto \bigcup\bigcup A$, and so forth, and likewise with $\bigcap$. By Theorem 15.17, operation (†) does not produce bigger sets than it starts with, if it starts with infinite sets. Operation (‡) does produce bigger sets:

17.3 Theorem. *If A is any set, then $A \prec \mathcal{P}(A)$.*

Proof. We have an injection $x \mapsto \{x\} : A \to \mathcal{P}(A)$, so $A \preccurlyeq \mathcal{P}(A)$. Suppose f is an arbitrary injection from A into $\mathcal{P}(A)$. Let B be the subset $\{x \in A : x \notin f(x)\}$ of A. Then B is not in the range of f. For, suppose $x \in A$. If $x \in B$, then $x \notin f(x)$, so $B \neq f(x)$. If $x \notin B$, then $x \in f(x)$, so again $B \neq f(x)$. So there is no bijection between A and $\mathcal{P}(A)$. □

If the natural numbers are precisely the counting-numbers, and if a set should be called 'countable' if its elements can be labelled with the counting-numbers, then the following definition makes sense.

17.4 Definition. If $A \preccurlyeq \omega$, then A is called **countable**. If $A \approx \omega$, then ω is the **cardinality** of A, that is, $|A| = \omega$. If $\omega \prec A$, then A is called **uncountable**.

By Theorem 17.3, we know that uncountable sets exist, at least in principle.

17.5 Definition. The set $\mathbb{R}$ of real numbers is called the **continuum**, and its cardinality is denoted by $\mathfrak{c}$; that is, $|\mathbb{R}| = \mathfrak{c}$.

17.6 Theorem. $\mathfrak{c} = |\mathcal{P}(\omega)|$*; in particular,* $\mathbb{R}$ *is uncountable.*

Proof. There is an injection $f : \mathbb{R} \to [0, 1)$ given by

$$f(x) = \begin{cases} (x-1)/(x-2), & \text{if } x \leqslant 0; \\ 1/(x+2), & \text{if } 0 \leqslant x. \end{cases}$$

So it is enough to show $[0, 1) \approx {}^{\omega}2$.

Theorem 10.5 gives a map $\mathfrak{a} \mapsto (n \mapsto a_n) : [0, 1) \to {}^{\omega}2$. In fact, this map is a bijection between $[0, 1)$ and ${}^{\omega}2 \smallsetminus A$, where A is the set of functions $\sigma : \omega \to 2$ such that, for some m in ω,

$$m \leqslant n \implies \sigma(n) = 1$$

for all n in ω (why?). Hence $[0, 1) \preccurlyeq {}^{\omega}2$. But the set A is countable (why?). Therefore we can define an injection from ${}^{\omega}2$ into $[0, 1)$ (how?). By the Schröder–Bernstein Theorem, we are done. □

17.7 Exercise. Supply the missing details in the proof.

17.8 Exercise. Show that $|{}^{\omega}\omega| = \mathfrak{c}$.

17.9 Exercise. Show that $\mathbb{R} \times \mathbb{R} \approx \mathbb{R}$.

18 Ordinal numbers

According to the ordinary use of the term, the *ordinal numbers* should serve as labels for the items on a list in such a way that the label determines the position of the item. In these notes, list-items are labelled with symbols like $(*)$ and $(\dagger)$ and $(\ddagger)$; these *distinguish* list-items, but do not indicate position. We shall not define the word *list*; but we propose:

- $(*)$ that every list be well-ordered;
- $(\dagger)$ that the assignment of ordinals to the items of a list be uniquely determined by the ordering of the items.

We shall define ordinals so that they satisfy these requirements, and so that the natural numbers are ordinals.

18.1 Definition. A class is called **transitive** if it properly includes each of its elements.

18.2 Examples. Each natural number is a transitive set. The set of natural numbers is a transitive set. The set

$$\{0, 1, \{1\}\},$$

that is, $\{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}\}$, is a transitive set, but the relation of containment $(\in)$ on this set is not a transitive relation, since $0 \in 1$ and $1 \in \{1\}$, but $0 \notin \{1\}$. Containment is a transitive relation on the set

$$\{1, \{1\}, \{1, \{1\}\}\},$$

but this set is not a transitive set, since the element 1 is not a subset.

18.3 Lemma. *No transitive set contains itself. Every transitive set includes the successor of each of its elements. The successor of every transitive set is transitive.*

Proof. Suppose A is transitive. If $B \not\subset A$, then $B \notin A$, by the definition of transitivity; therefore $A \notin A$. If $x \in A$, then $\{x\} \subseteq A$, but also $x \subset A$ by transitivity of A, so that $x' \subseteq A$. If $y \in A'$, then either $y = A$ or $y \in A$; in either case, $y \subset A'$. Thus A' is transitive. □

18.4 Definition. An **initial segment** of a well-ordered set $(A, \leqslant)$ is a subset B of A such that

$$x \leqslant y \wedge y \in B \implies x \in B$$

for all x and y in A. An initial segment is **proper** if it is not the whole set.

18.5 Example. Every natural number is a proper initial segment of $(\omega, \subseteq)$.

18.6 Lemma. *For every proper initial segment B of a well-ordered set $(A, \leqslant)$, there is an element x of A such that $B = \{y \in A : y < x\} = \overline{x}$.*

Proof. Let x be the least element of $A \smallsetminus B$. □

18.7 Definition. A transitive set is an **ordinal (number)** if it is strictly well-ordered by containment.

18.8 Exercise. Show that the ordinals compose a class.

18.9 Definition. The class of ordinals is $\mathbf{ON}$.

18.10 Example. $\omega \subseteq \mathbf{ON}$ and $\omega \in \mathbf{ON}$.

We shall let lower-case letters from the beginning of the Greek alphabet—such as α, β, γ, δ and ζ—refer to ordinals.

18.11 Lemma. *Suppose α is an ordinal, and x is a set. The following are equivalent:*

($*$) $x \in \alpha$

(†) *x is a proper initial segment of α*

(‡) *x is an ordinal, and $x \subset \alpha$*

Proof. ($*$) $\Longrightarrow$ (‡): Suppose $x \in \alpha$. Then $x \subset \alpha$, by transitivity of α. Hence x is strictly well-ordered by containment, since α is. Say $y \in x$. Then $y \neq x$, since $\in$ is irreflexive on α. Also, if $z \in y$, then $z \in \alpha$, by transitivity of α, so $z \in x$ by transitivity of $\in$ on α. Thus $y \subset x$. Therefore x is transitive.

(‡) $\Longrightarrow$ (†): Suppose x is an ordinal, and $x \subset \alpha$. Say $y \in x$ and $z \in y$. Then $z \in x$ by transitivity of x. Hence x is a proper initial segment of α.

(†) $\Longrightarrow$ ($*$): Suppose x is a proper initial segment of α. Then

$$x = \{z \in \alpha : z \in y\}$$

for some y in α, by Lemma 18.6. But if $z \in y$, then $z \in \alpha$, by transitivity of α. Hence $x = y$. □

18.12 Lemma. *Suppose α and β are distinct ordinals such that $\alpha \notin \beta$. Then $\beta \in \alpha$.*

Proof. Since $\alpha \notin \beta$, we have $\alpha \not\subset \beta$ by the previous Lemma. Since also $\alpha \neq \beta$, there is an element of $\alpha \smallsetminus \beta$. Let γ be the least element of $\alpha \smallsetminus \beta$. Then $\gamma \subseteq \beta$, but $\gamma \notin \beta$, so $\gamma \not\subset \beta$, and therefore $\gamma = \beta$. □

18.13 Theorem. *Every class of ordinals is strictly well-ordered by containment.*

Proof. Let $\mathbf{C}$ be a class of ordinals. Then containment is transitive on $\mathbf{C}$, since every element of $\mathbf{C}$ is transitive. So containment is a strict total ordering of $\mathbf{C}$, by the last two lemmas. If $\alpha \in \mathbf{C}$, then either α is the least element of $\mathbf{C}$, or $\mathbf{C} \cap \alpha$ has a least element, which is the least element of $\mathbf{C}$. □

18.14 Corollary (Burali-Forti Paradox). $\mathbf{ON}$ *is not a set.*

Proof. The class $\mathbf{ON}$ is transitive by Lemma 18.11. Suppose A is a transitive set of ordinals that is strictly well-ordered by containment. Then A is an ordinal, and $A \in \mathbf{ON} \smallsetminus A$. In particular, $A \neq \mathbf{ON}$. □

As noted in Remark 11.5, being well-ordered, the class **ON** admits strong induction and recursion. It is also said that **ON** admits **transfinite** induction and recursion.

19 Order-types

We are ready to show that the items in every list can be uniquely labelled by ordinals.

19.1 Definition. Two totally ordered sets **have the same order-type** if they are isomorphic, that is, there is an order-preserving bijection between them. An **order-type** for a *well-ordered* set is an ordinal that is isomorphic to it. That is, the ordinal α is an order-type for a well-ordered set $(A, \leqslant)$, provided there is a bijection $f : A \to \alpha$ such that

$$x < y \to f(x) \in f(y)$$

for all x and y in A.

19.2 Lemma. *No well-ordered set has more than one order-type, or has more than one isomorphism onto its order-type.*

Proof. It is enough to show that every ordinal has exactly one order-type, namely itself, and that the only isomorphism from an ordinal onto itself is the identity. Suppose $f : \alpha \to \beta$ is a surjective map of ordinals which is *not* the identity. Let γ be the least element of α such that $f(\gamma) \neq \gamma$. If $f(\gamma) \in \gamma$, then $f(f(\gamma)) = f(\gamma)$, by minimality of γ. If $\gamma \in f(\gamma)$, then (by surjectivity of f) there is ζ in α such that $\gamma \in \zeta$ and $f(\zeta) = \gamma$. In either case, f is not an isomorphism. □

19.3 Theorem. *Every well-ordered set has exactly one order-type.*

Proof. Suppose $(A, \leqslant)$ is a well-ordered set. Suppose $x \in A$. If $\overline{x}$ has an order-type α, let f be the isomorphism from $\overline{x}$ onto α; if $y \in \overline{x}$, then $f(y)$ is an order-type for $\overline{y}$.

By uniqueness of order-types, if, for every y in $\overline{x}$, there is an order-type $f(y)$ for $\overline{y}$, then the set $\{f(y) : y \in \overline{x}\}$ is transitive, so it is an ordinal, by Theorem 18.13; hence it is the order-type of $\overline{x}$. By strong induction, every proper initial segment of A has an order-type; the set of these order-types is the order-type of A. □

20 Kinds of ordinals

Let us feel free to write

$$\alpha < \beta,$$

if $\alpha \in \beta$; and $\alpha \leqslant \beta$, if $\alpha \subseteq \beta$.

20.1 Theorem. *α' is the least ordinal greater than α.*

Proof. Note first that α' *is* an ordinal. Also, $\alpha < \alpha'$, and if $\alpha < \beta$, then $\alpha' \leqslant \beta$, by Lemma 18.3. $\square$

20.2 Corollary. *Every successor-ordinal has a unique predecessor.*

Proof. If $\alpha < \beta$, then $\alpha' \leqslant \beta < \beta'$. Hence, by Lemma 18.12, if $\alpha' = \beta'$, then $\alpha = \beta$. $\square$

20.3 Definition. An ordinal is **positive** if it is not 0. A positive ordinal which is not a successor is called a **limit** ordinal.

20.4 Theorem. *ω is the least limit ordinal.*

Proof. ω *is* a limit, since $\omega \neq 0$, and if $n < \omega$, then $n' < \omega$ by definition of ω. So ω is the least limit ordinal by Lemma 18.12. $\square$

We can write

$$\omega' = \{0, 1, 2, \dots; \omega\},$$

where the semicolon (;) indicates that ω is a limit.

20.5 Remark. If α is a limit ordinal, then $(\alpha,', 0)$ is a model of Axiom Z and Axiom U. The next section will show that there are limit ordinals strictly larger than ω.

21 Cardinality

Every *well-ordered* set is equipollent with some ordinal, by § 19. *Every* set has a choice-function, by **AC**.

21.1 Theorem. *Do not assume* **AC**. *A non-empty set has a choice-function if and only if the set is equipollent with some ordinal.*

Proof. Suppose $f : \mathcal{P}(A) \to A$ is a choice-function for A. By strong recursion, for every ordinal α, there is a unique function $g_\alpha : \alpha' \to A$ such that

$$g_\alpha(\beta) = f(A \smallsetminus g_\alpha''\beta)$$

for all β in α'. By definition of a choice-function, each g_α is either injective or surjective. If g_α is always injective, then the function $\alpha \mapsto g_\alpha(\alpha)$ orders a subset of A with the order-type of **ON**, which is absurd. So let α be least such that g_α is surjective. Then g_α is a bijection between α and A.

Conversely, if $A \approx \alpha \in \mathbf{ON}$, then $X \mapsto \min X$ on $\mathcal{P}(A) \smallsetminus \{\varnothing\}$ extends to a choice-function of A. $\square$

21.2 Exercise. Look up and prove other equivalent forms of **AC**.

The following is consistent with Definition 17.4:

21.3 Definition. The **cardinality** of a set is the least ordinal that is equipollent with it. The cardinality of A is denoted $|A|$. An ordinal is a **cardinal** if it is the cardinality of some set.

21.4 Theorem. *Infinite cardinals are limit ordinals.*

21.5 Corollary. *There are limit ordinals strictly larger than ω (assuming* **AC***.)*

21.6 Exercise. Prove the theorem and its corollary.

21.7 Exercise. Find an example of a limit ordinal that is not a cardinal.

22 The list of cardinals

Every finite ordinal is a cardinal, but some infinite ordinals, such as ω', are not cardinals. However, for every cardinal there is a larger cardinal; so—since cardinals are ordinals—there is a *least* larger.

22.1 Definition. If κ is a cardinal, then κ^+ is the least element of $\{\alpha \in |\mathcal{P}(\kappa)| : \kappa < |\alpha|\}$.

The following is an instance of *trans-finite recursion*:

22.2 Definition. The ordinals $\aleph_\alpha$ as follows:

- ($*$) $\aleph_0 = \omega$,
- (†) $\aleph_{\beta'} = \aleph_\beta^+$,
- (‡) $\aleph_\delta = \bigcup_{\gamma<\delta} \aleph_\gamma$, if δ is a limit-ordinal.

($\aleph$ is the Hebrew letter *aleph*.)

22.3 Lemma. *The infinite cardinals are precisely the ordinals $\aleph_\alpha$, and*

$$\alpha < \beta \leftrightarrow \aleph_\alpha < \aleph_\beta.$$

In particular, the assignment $\alpha \mapsto \aleph_\alpha$ is an order-preserving bijection between **ON** *and the class of infinite cardinals.*

Proof. We first prove that each ordinal $\aleph_\alpha$ is a cardinal, and

$$\forall\beta\ (\beta < \alpha \rightarrow \aleph_\beta < \aleph_\alpha). \qquad (*)$$

This claim is true when $\alpha = 0$. Suppose it is true when $\alpha = \gamma$. Then by definition, $\aleph_{\gamma'}$ is the least ordinal whose cardinality is greater than $\aleph_\gamma$. In particular, $\aleph_{\gamma'}$ is a cardinal, and

$$\aleph_\gamma < \aleph_{\gamma'}.$$

Hence ($*$) holds when $\alpha = \gamma'$.

Now suppose that δ is a limit ordinal. Say $\aleph_\alpha$ is a cardinal, and ($*$) holds, whenever $\alpha < \delta$. By definition, $\aleph_\delta$ is the least ordinal that includes each cardinal $\aleph_\alpha$ such that $\alpha < \delta$. If an ordinal includes a cardinal, then its cardinality includes that cardinal; therefore $\aleph_\delta$ is a cardinal. Also, if $\beta < \delta$, then

$$\aleph_\beta < \aleph_{\beta'} \leqslant \aleph_\delta$$

by inductive hypothesis, since $\beta < \beta' < \delta$; so (∗) holds when $\alpha = \delta$.

Finally, suppose κ is an infinite cardinal. Let A be the class

$$\{\alpha \in \mathbf{ON} : \aleph_\alpha < \kappa\}.$$

By the Replacement Axiom, A is a *set*, since it is in one-to-one correspondence with the subset

$$\{\beta < \kappa : \exists\alpha\ \aleph_\alpha = \beta\}$$

of κ. Hence A is not **ON**. In particular, there is a least ordinal β such that $\kappa \leqslant \aleph_\beta$. Hence $\aleph_\alpha < \kappa$ when $\alpha < \beta$. But this property of κ is shared by $\aleph_\beta$, which is the least cardinal with this property. Therefore $\kappa = \aleph_\beta$. □

23 The Continuum Hypothesis

We can now say that $\aleph_1$ is the least or first uncountable cardinal. What else can we say about $\aleph_1$? The **Continuum Hypothesis** (or **CH**) is that

$$\aleph_1 = \mathfrak{c}.$$

It turns out that, just as **AC** is independent of **ZF**, so **CH** is independent of **ZFC**.

Now, calculus can be developed using only **ZFC**. Therefore calculus will never answer the question of whether

$$\aleph_0 < |A| < \mathfrak{c}$$

for some subset A of $\mathbb{R}$. In a sense, this question has no answer.

In another sense, this question can have whatever answer we like. We assume **AC** because we can, and because it seems to yield good mathematics (such as our theorem that all sets are finite or infinite). In the same way, we could assume **CH**, or ¬**CH**. Some logicians are recommending the latter.

Whether **CH** is assumed or not, we can make the following.

23.1 Definition. The assignment $\alpha \mapsto \beth_\alpha$ of ordinals to infinite cardinals is made as follows.

(∗) $\beth_0 = \omega$;

(†) $\beth_{\beta'} = |\mathcal{P}(\beth_\beta)|$;

(‡) $\beth_\delta = \bigcup_{\gamma<\delta} \beth_\gamma$, if δ is a limit ordinal.

($\beth$ is the Hebrew letter *beth.*)

The Continuum Hypothesis is that

$$\aleph_1 = \beth_1;$$

the **Generalized Continuum Hypothesis**, or **GCH**, is that $\aleph_\alpha = \beth_\alpha$ for all α.

24 Ordinal arithmetic

If we have two lists, we can put one after the other to get a new list. If we have a list of lists, then we can make one big list by first listing all items of the first list, then listing all items of the second list, and so forth. These ideas make sense for well-ordered sets in general.

24.1 Definition. Let $(A, <)$ and $(B, <)$ be totally ordered sets. The **lexicographic** (or **dictionary-**) **order** on $A \times B$ is given by

$$(a, b) < (c, d) \leftrightarrow b < d \vee (b = d \wedge a < c).$$

24.2 Example. Say A is the Arabic alphabet, equipped with its alphabetical order. The lexicographic order on $A \times A$ gives the order in which all two-letter words would appear in a dictionary. (The point of using Arabic is that it is read from right to left.)

24.3 Lemma. *If $(A, <)$ and $(B, <)$ are well-ordered sets, then $A \times B$ is well-ordered by the lexicographic order.*

Proof. It is clear that the lexicographic order on $A \times B$ is total. Say C is a non-empty subset of $A \times B$. Let b be the least element of

$$\{y \in B : \exists x\ (x \in A \wedge (x, y) \in C)\},$$

and let a be the least element of $\{x \in A : (x, b) \in C\}$. Then (a, b) is the least element of C. □

24.4 Definition. The **ordinal sum** $\alpha + \beta$ (the result of **adding** β to α) is the order-type of

$$(\alpha \times \{0\}) \cup (\beta \times \{1\}),$$

considered as a subset of $(\alpha \cup \beta) \times 2$ with the lexicographic order. The **ordinal product** $\alpha\beta$ (the result of **multiplying** α by β) is the order-type of $\alpha \times \beta$.

We shall see presently that ordinal addition and multiplication, applied to natural numbers, agree with the operations defined earlier. In general though, the ordinal operations are not commutative.

24.5 Examples. $\omega + 1 = \omega'$, and $\omega + \omega = \omega 2$, but $1 + \omega = \omega$ and $2\omega = \omega$.

24.6 Theorem.

(∗) $\alpha' = \alpha + 1$

(†) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$

(‡) $0 + \alpha = \alpha + 0 = \alpha$

(§) $0\alpha = \alpha 0 = 0$

(¶) $(\alpha\beta)\gamma = \alpha(\beta\gamma)$

(‖) $1\alpha = \alpha 1 = \alpha$

(**) $\alpha(\beta+\gamma) = \alpha\beta + \alpha\gamma$

24.7 Exercise. Prove the theorem.

24.8 Corollary. *Applied to natural numbers, the ordinal operations agree with the earlier definitions.*

Proof. It is enough to note:

$$\begin{aligned} \alpha + 0 &= \alpha, \\ \alpha + \beta' &= \alpha + (\beta + 1) = (\alpha + \beta) + 1 = (\alpha+\beta)', \\ \alpha 0 &= 0, \\ \alpha\beta' &= \alpha(\beta+1) = \alpha\beta + \alpha 1 = \alpha\beta + \alpha, \end{aligned}$$

where the operations are the ordinal ones; so these agree with the operations defined on natural numbers. □

We can now write the following initial segment of **ON**:

$$\{0, 1, 2, \dots; \omega, \omega+1, \omega+2, \dots; \omega 2, \omega 2 + 1, \dots; \omega 3, \dots; \omega\omega\}.$$

Here the ordinals following the semicolons (;) are limits. Note also that there are limits between $\omega 3$ and $\omega\omega$. We can continue the initial segment of **ON** by writing $\omega\omega$ as ω^2, and $\omega^2\omega$ as ω^3, and so on; and then we can write ω^ω for the least ordinal that includes the ordinals ω^n with n in ω. Formally, we have the following, by trans-finite induction:

24.9 Theorem. *For any ordinal α, there is a unique function $\beta \mapsto \alpha^\beta$ on* **ON** *such that:*

(*) $\alpha^0 = 1$;

(†) $\alpha^{\beta+1} = \alpha^\beta\alpha$ *for all β in* **ON**;

(‡) $\alpha^\delta = \{\gamma : \exists\beta(\beta \in \delta \land \gamma \in \alpha^\beta)\}$, *for all limit ordinals δ.*

24.10 Lemma. *Applied to natural numbers, the definition in the theorem agrees with the definition given in § 3.*

24.11 Exercise. Prove the lemma.

We can now continue or initial segment of **ON** from where we left off:

$$\{\dots; \omega^2, \omega^2+1, \dots; \omega^2+\omega, \dots; \omega^2 2, \dots; \omega^3, \dots; \omega^\omega, \dots; \\ \omega^{\omega 2}, \dots; \omega^{\omega^2}, \dots; \omega^{\omega^\omega}, \dots; \omega^{\omega^{\omega^\omega}}, \dots\}.$$

So we have named a lot of infinite ordinals. Still, we have only just begun, since all of them are countable.

25 Cardinal arithmetic

Now let κ, μ and ν be cardinals. We can refer to the cardinal κ^+ as the *successor* of κ, but we must be clear that we mean the successor of κ *as a cardinal.* In general, κ^+ is not $\kappa+1$ unless κ is finite.

We can define addition and multiplication of cardinals, though again we must distinguish these from the corresponding operations on ordinals.

25.1 Definition. Cardinal addition and multiplication are thus:

- $\kappa + \mu = |(\kappa \times \{0\}) \cup (\mu \times \{1\})|$;
- $\kappa\mu = |\kappa \times \mu|$.

We can also define powers of cardinals with cardinal exponents, but the definition is more divergent from the ordinal case. First note the following consequence in that case:

25.2 Theorem. $\alpha^{\beta+\gamma} = \alpha^\beta\alpha^\gamma$.

Proof. The claim is true when $\gamma = 0$. If it is true when $\gamma = \zeta$, then

$$\alpha^{\beta+(\zeta+1)} = \alpha^{(\beta+\zeta)+1} = \alpha^{\beta+\zeta}\alpha = \alpha^\beta\alpha^\zeta\alpha = \alpha^\beta\alpha^{\zeta+1},$$

so the claim is true when $\gamma = \zeta + 1$. Finally, if it is true when $\gamma < \delta$, and δ is a limit ordinal, then, since $\delta = \bigcup\{\gamma : \gamma < \delta\}$, we have

$$\begin{aligned}
\alpha^{\beta+\delta} &= \alpha^{\bigcup\{\beta+\gamma:\gamma<\delta\}} \\
&= \bigcup\{\alpha^{\beta+\gamma} : \gamma < \delta\} \\
&= \bigcup\{\alpha^\beta\alpha^\gamma : \gamma < \delta\} \\
&= \alpha^\beta\left(\bigcup\{\alpha^\gamma : \gamma < \delta\}\right) \\
&= \alpha^\beta\alpha^\delta.
\end{aligned}$$

The claim follows by trans-finite induction. □

So that the corresponding theorem will hold in the cardinal case, we make the following.

25.3 Definition. $\kappa^\mu = |{}^\mu\kappa|$. (See Definition 17.1.)

25.4 Theorem. $\kappa^{\mu+\nu} = \kappa^\mu\kappa^\nu$.

Proof. We exhibit a one-to-one correspondence between the set of functions from

$$(\mu \times \{0\}) \cup (\nu \times \{1\})$$

to κ, and the set ${}^\mu\kappa \times {}^\nu\kappa$. If f is in the former, then define (g, h) in the latter by $g(x) = f(x, 0)$ and $h(x) = f(x, 1)$. □

Because of Lemma 17.2, we have $\mathfrak{c} = 2^{\aleph_0}$, and $\beth_{\alpha+1} = 2^{\beth_\alpha}$.

References

[1] Stanley N. Burris. *Logic for Mathematics and Computer Science*. Prentice Hall, Upper Saddle River, New Jersey, USA, 1998.

[2] Krzysztof Ciesielski. *Set theory for the working mathematician*, volume 39 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1997.

[3] Richard Dedekind. *Essays on the theory of numbers. I:Continuity and irrational numbers. II:The nature and meaning of numbers*. Dover Publications Inc., New York, 1963.

[4] René Descartes. *Meditationes*. Paris, 1641.

[5] René Descartes. *Meditations on First Philosophy*. Hackett Publishing Co., Indianapolis, Indiana, USA, 1979. Translated from the Latin by Donald A. Cress.

[6] Keith Devlin. *The joy of sets*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, second edition, 1993. Fundamentals of contemporary set theory.

[7] Susanna S. Epp. *Discrete Mathematics with Applications*. PWS Publishing Company, Boston, Massachusetts, USA, 1995. 2nd edition.

[8] Euclid. *The thirteen books of Euclid's Elements translated from the text of Heiberg. Vol. I: Introduction and Books I, II. Vol. II: Books III–IX. Vol. III: Books X–XIII and Appendix*. Dover Publications Inc., New York, 1956. Translated with introduction and commentary by Thomas L. Heath, 2nd ed.

[9] András Hajnal and Peter Hamburger. *Set theory*, volume 48 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1999. Translated from the 1983 Hungarian original by Attila Máté.

[10] Kenneth Kunen. *Set theory*, volume 102 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., Amsterdam, 1983. An introduction to independence proofs, Reprint of the 1980 original.

[11] Edmund Landau. *Foundations of Analysis. The Arithmetic of Whole, Rational, Irrational and Complex Numbers*. Chelsea Publishing Company, New York, N.Y., 1951. Translated by F. Steinhardt.

[12] Azriel Lévy. *Basic set theory*. Springer-Verlag, Berlin, 1979.

[13] Yiannis N. Moschovakis. *Notes on set theory*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1994.

[14] Plato. *Republic*. Loeb Classical Library. Harvard University Press, Cambridge, Massachusetts, USA, 1980. With an English Translation by Paul Shorey. In two volumes.

[15] Plato. *Republic*. Oxford University Press, Translated with an Introduction and Notes by Robin Waterfield 1998.

[16] Bruno Poizat. *A course in model theory.* Universitext. Springer-Verlag, New York, 2000. An introduction to contemporary mathematical logic, Translated from the French by Moses Klein and revised by the author.

[17] Robert R. Stoll. *Set theory and logic.* Dover Publications Inc., New York, 1979. Corrected reprint of the 1963 edition.

[18] Robert L. Vaught. *Set theory.* Birkhäuser Boston Inc., Boston, MA, second edition, 1995. An introduction.

Contents