Sets and Classes

David Pierce

March 2, 2007

Contents

Preface

These notes are for use in a course called Set Theory, given in the Mathematics Department of Middle East Technical University, Ankara, under the designation Math 320. The catalogue-description of the course is:

> Language and axioms of set theory. Ordered pairs, relations and functions. Order relation[s] and well ordered sets. Ordinal numbers, transfinite induction, arithmetic of ordinal numbers. Cardinality and arithmatic of cardinal numbers. Axiom of choice, generalized continuum hypothesis.

These notes cover these topics, and more.

I wrote a first version of these notes in 2001, for Fundamentals of Mathematics (Math 111); but a good part of the notes went beyond what that course had time for. I revised the notes for use in Math 320 as taught by Ayşe Berkman in the spring semester of the 2004/5 academic year. Now I have completely rewritten the notes from the 2005 edition (specifically, from Ayşe's copy of that edition, with her handwritten comments). Since many things have changed, these notes must still be considered as a rough draft.

In writing the present edition of these notes, I have placed more emphasis than in previous editions on the following picture of set-theory:

(i) There are things called *sets*, with certain properties. Sets compose a so-called *universal class* (denoted by $\mathbf{V}$ in these notes and elsewhere).

(ii) There is a logical language for talking about sets; the one non-logical symbol of this language is $\in$, to express *membership* of one set in another. (*Equality* of sets can be *defined* in terms of membership.)

(iii) In the language of sets, a formula φ with one free variable defines a *sub-class*—namely, $\{x\colon \varphi(x)\}$—of the universal class.

(iv) Sets are also classes. Most of things that one does with sets in mathematics—like taking unions or intersections or power-sets—can be done with classes.

(v) There is no reason to assume that all classes are sets. Indeed, there is a class of all sets that are not members of themselves, namely $\{x\colon x \notin x\}$,

but it is not a set. Thus, the so-called Russell Paradox is simply a basic theorem of set-theory (see ¶3.1.8 below).

This picture of set-theory can be seen in Levy [10]; I have found his book a useful reference, though it is dense with detail. Levy is also a good source for historical references. Other useful books (at different levels) have been Suppes [18], Kunen [9], and Moschovakis [11]; also Shoenfield [16] for Ch. 8. I have used Fraenkel *et al.* [7] for a review of the development of the individual axioms that generate what is called Zermelo–Fraenkel set-theory. Those axioms are listed on p. 9 below. In that list, a weaker form of the Union Axiom is included, along with the usual form. In the text, I try to introduce axioms only when they are needed for something interesting; the full Union Axiom is not needed for a long while; meanwhile, a weaker version suffices. (Possibly I have overlooked some earlier implicit use of Union or other axioms; I remind the reader of my comment that these notes are still a rough draft.)

Unlike most set-theory textbooks, I aim (in Ch. 2) to give a precise formal account of the logic behind set-theory. It may be pointed out that the Zermelo–Fraenkel axioms were worked out *before* the formality of the logic was worked out. So, the reader is perhaps not required to know the formal logic. But it seems useful to me to know that classes can be given a precise definition.

I try to work with the idea that most of the Zermelo–Fraenkel axioms amount to assertions that certain classes are sets. I have not seen the Infinity Axiom treated explicitly in this way, except in these notes. Here I take some trouble (in Ch. 4) to see how the *class* of natural numbers can be obtained, *without* the assumption that it already exists as a set.

These notes are not intended for use in isolation from the classroom. Points presented here in outline may be elaborated more fully in lectures. I take Euclid's *Elements* [6] as a model. Euclid simply presents propositions and proofs, with no explanation of *why* one would want to prove these propositions. Presumably the explanation is left to the living teacher.

Terms being **defined** in these notes are printed in boldface; technical terms being emphasized, but not properly defined, are *slanted*. There is an index of these terms at the back (pp. 104–107). Most chapters have exercises at the end; often the exercises ask the reader to supply the proofs of propositions (lemmas, theorems) given in the chapter.

The remainder of this preface is adapted from the 2005 edition of these notes.

* * * * *

Any text on axiomatic set-theory will introduce the set ω, which is the smallest set that contains $\varnothing$ and is closed under the operation $x \mapsto x \cup \{x\}$.

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) I prove (in Ch. 5) facts about the natural numbers *from the Peano axioms*, not just *in* ω.

(ii) I mention structures that are models of some, but not all, of the Peano axioms. (See for example Fig. 4.1.)

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, because it can bring out a distinction that is often ignored. If a given set contains all of the natural numbers, this can sometimes be proved by the technique of *(ordinary) induction*. If the domain of a certain function is the set of natural numbers, then the function can sometimes be defined by the technique of *(ordinary) recursion*. One writer—Vaught [20, ch. 2, § 4]—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 works *only* in models of the Peano axioms, while there are other structures in which *proof* by induction works.

There are *strong* (also called *trans-finite*) versions of induction and recursion. There is proof by strong induction, and there is definition by strong recursion. Admission of either of *these* is equivalent to admission of the other; the ordered sets that admit them are precisely the *well-ordered* sets. (See § 6.1.) Some basic undergraduate texts suggest confusion on this point. For example, in talking about the integers, one book[1] 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* (¶6.1.4), something that does not follow from strong induction.

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 [8] defines ω on the first page

[1] Namely, Epp [5, § 4.4, p. 213], used sometimes in Math 111 and 112 at METU.

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 [4] 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 [11] *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.

List of Figures

List of Axioms

Here, a and b are arbitrary sets, $\boldsymbol{C}$ is an arbitrary class, $\boldsymbol{F}$ is an arbitrary function, and ω and $\mathbf{WF}$ are the classes defined on pp. 53 and 94 respectively.

(i) [p. 27] Extension:

$$a = b \Leftrightarrow a \subseteq b \,\&\, b \subseteq a.$$

(ii) [p. 30] Comprehension-Scheme:

$$\exists x\ x = a \cap \boldsymbol{C}.$$

(iii) [p. 31] Pairing:

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

(iv) [p. 40] Replacement-scheme:

$$a \subseteq \mathrm{dom}(\boldsymbol{F}) \Rightarrow \exists x\ x = \boldsymbol{F}[\,a\,].$$

(v) [p. 45] Power-set:

$$\exists x\ x = \mathcal{P}(a).$$

(vi) [p. 48] Weak Union:

$$\exists x\ x = a \cup \{b\}.$$

(vii) [p. 72] Union:

$$\exists x\ x = \bigcup a.$$

(viii) [p. 75] Infinity:

$$\exists x\ x = \omega.$$

(ix) [p. 88] Choice:

Every set has a choice-function.

(x) [p. 95] Foundation:

$$\mathbf{V} = \mathbf{WF}.$$

Chapter 1

Introduction

1.1 What a set is

1.1.1. A **set** is a thing that **contains** other things. Those other things are called **members** or **elements** of the set. The set **comprises** its members. But the set cannot be separated from its elements the way a box can be emptied of its contents: the members **compose** the set. A set *is* its elements, considered as one thing. It is a multitude that is also a unity.

1.1.2. A flock of pigeons is a set; a pair of socks is a set; a deck of cards is a set; the number of days in a week is a set. Words like *flock, pair, deck,* and *number* are (or can be) **collective nouns.** In English, such nouns can be used as subjects of singular or plural verbs:

> *A flock of pigeons is attacking that crust of bread.*
>
> *A flock of pigeons are attacking that crust of bread.*

The concept of being describable by a collective noun has shown itself to be a fruitful subject for mathematical study. In this study, we use the words *set* and *class* as the most generally applicable collective nouns. It will turn out that we must make a distinction: every set is a class (¶3.1.4), but not every class is a set (¶3.1.8).

1.1.3. A set contains other things (¶1.1.1). As it happens, we may modify the expression *other things* in three ways: We might ignore the *other,* allowing a set to contain *itself.* Or, a set might contain, not other *things,* but just *one* other thing. Or possibly a set contains *nothing* at all.

1.2 Why study sets

1.2.1. Sets are a foundation for mathematics, in that the objects of mathematics can be understood as sets. For example, a function f can be understood as the set of *ordered pairs* (a, b) such that $f(a) = b$; and an ordered

pair itself can be understood as a set. Perhaps a and b here are *real numbers;* these can be understood as certain sets of *rational numbers;* rational numbers can be understood as certain sets of *integers;* integers can be understood as certain sets of *natural numbers;* these are sets that are built up from the one set that contains—nothing.

1.2.2. Sets are a foundation for logic, the science of reasoning. For example, there is a correspondence between the *union,* $a \cup b$, of two sets a and b and the *disjunction,* $P \vee Q$, of two propositions P and Q.

1.2.3. Set-theory is an example of an **axiomatic system.** In such a system, one **postulates** certain truths, called **axioms,** which are held to be self--evident; from the axioms, by means of logic, one derives other truths, which are not so evident. Thus, set-theory is a modern example of a method as old as Euclid [6] for organizing and developing a body of mathematical work.

1.3 Cardinals and ordinals

1.3.1. Set-theory is the origin of some amazing results, as for example concerning the numbers we start using early in life. As young children, we learn to chant a sequence of numbers: *one, two, three* in English; bir, iki, üç in Turkish. We learn to use these numbers as *cardinal numbers,* to indicate the *sizes* of sets. Later we learn that *zero*/sıfır is also a size.

1.3.2. But the sequence of numbers also has an *order,* like the order of letters in the alphabet. So we can use numbers as *ordinal numbers,* to indicate *position* of elements *within* a set. We also learn special words for ordinal numbers: *first, second, third* in English; birinci, ikinci, üçüncü in Turkish.

1.3.3. So numbers have two uses, as cardinals and as ordinals; these uses are completely different; and yet the *same* underlying numbers are used in each case. For example, if a book starts on page 1 and ends on page 108, then we know two things: the book has 108 pages, and the last page of the book is the 108th page of the book.

1.3.4. I once boarded an airplane and found my seat. Through my window, I could see that a valuable cargo was being loaded: perhaps it was banknotes. The cargo was in small cardboard boxes. Armed guards were standing by. The boxes were placed one by one on a moving ramp that carried them into the hold of the airplane. As the boxes ascended the ramp, a man wrote numerals on them: 1, 2, 3, and so on. He was counting the boxes: he was making sure that none was missing.

1.3.5. I did not see how the boxes were unloaded at the end of the flight. They may have been sent down the ramp from the hold so that the box labelled 1 came first, the box labelled 2 came second, and so forth. But this

procedure may have been difficult for the cargo handlers to accomplish; in any case, it would have been unnecessary. No matter the order with which the boxes came off the plane, they could have been counted by the same procedure used when they were loaded. If no box was missing, then the same number of boxes would have been found.

1.3.6. When we count a set, then the order of the elements in the set does not matter. Take the sixteen pawns from a chess-set and put them in a bag. Draw them out one by one, each time uttering the next number on the standard list. When you draw the last pawn, you will reach 16 on the list. Put the pawns back in the bag, shake the bag, and repeat; you will still reach 16. This is a fact so basic that we do not need to learn it in school. However, set-theory opens up a world in which our usual method of counting *fails*. This is the world of *infinite* sets.

1.3.7. What is a number? In set-theory, we identify certain sets that we call 0, 1, 2, 3, and so forth. These sets are (by our definition) the *natural numbers*. Each natural number n has a *successor,* which can be called $n+1$. We *assume* that all of the natural numbers compose a set, called ω. (This letter is not w, the so-called double u; it is the Greek minuscule *omega,* the last letter of the alphabet.[1]) We treat ω as a new number. Then we can form the new numbers $\omega + 1$, $\omega + 2$, and so forth. Beyond all of these new numbers, there is $\omega + \omega$ or $\omega \cdot 2$, then $\omega \cdot 2 + 1$, and so forth; then $\omega \cdot 3$ and so forth; then $\omega \cdot \omega$ or ω^2, &c., then ω^ω, ω^{ω^ω}, and so on. All of these new numbers are examples of *trans-finite ordinal numbers*. We may be able to use these trans-finite ordinals to count a set. Perhaps we can assign *each* natural number to some different element of the set, only to find that some elements of the set are left over. So we assign ω itself to one of these, and $\omega + 1$ to another, and so forth. Perhaps, in this way, all of the ordinals that come before $\omega^3 \cdot 7 + \omega^2 \cdot 4 + \omega \cdot 5$ are assigned to elements, but there is no element left to which we can assign $\omega^3 \cdot 7 + \omega^2 \cdot 4 + \omega \cdot 5$ itself. Then how many elements has the set? It will turn out that the set has the same number of elements that ω has: We can rearrange our numbering of the set so that *each* element is assigned a natural number.

1.3.8. There *are* sets larger than ω; these too can be 'counted' with ordinals, but with ambiguity as before. The size of ω is given the name $\aleph_0$; this is the first *trans-finite cardinal number*. (The letter $\aleph$ is *aleph,* the first letter of the Hebrew alphabet.) There is a next larger cardinal number, $\aleph_1$; then we have $\aleph_2$, $\aleph_3$, ..., $\aleph_\omega$, $\aleph_{\omega+1}$, ..., $\aleph_{\omega\cdot 2}$, and so on. Thus the trans-finite cardinal numbers are *indexed* by the ordinal numbers. But the size of an infinite set is not obtained or defined directly by a process of counting.

[1] Omega is a *large* or *long* (mega) o, to be contrasted with the small or short (micro) o, omicron. One might even call omega a double o, and indeed its written minuscule form seems to come from oo.

1.4 Sets as numbers

1.4.1. I suggested in ¶1.1.2 that *number* can be a collective noun. It is not always so. If I ask you,

> *Pick a number between one and ten,*

you will probably not think of a set or a number *of things;* you will just pick one abstract thing, called *five* perhaps, or *eight.* But if we observe,

> *A number of people are gathering in the street,*

then the emphasis is on the *people* as much as on how many there are. In this latter sentence, the sense of *number* would seem to be that of the Greek ἀριθμός. (See Table 1.1 on page 15 for the Greek alphabet.) The word ἀριθμός is the origin of the word *arithmetic,* and it is commonly translated as *number;* but note how Euclid of Alexandria defines it [6, Book VII, Definitions]:

> (i) Μονάς ἐστιν, καθ' ἣν ἕκαστον τῶν ὄντων ἓν λέγεται.
> A **unit** is that by virtue of which each thing is called *one* (ἕν).
>
> (ii) Ἀριθμὸς δὲ τὸ ἐκ μονάδων συγκείμενον πλῆθος.
> A **number** is a multitude (πλῆθος) composed of units.

This account of *number* bears some resemblance to the account of *set* in ¶1.1.1. However, Euclid does not allow the exceptional cases discussed in ¶1.1.3; in particular, for Euclid, *one* is not a number.[2]

1.4.2. If we take *number* seriously as a collective noun roughly equivalent to *set,* then a certain passage by Plato of Athens becomes an argument in favor of studying sets. Numbers and sets are worth studying, because they somehow combine opposites like *many* and *one, multiplicity* and *unity.* The Platonic passage is from the work commonly called the *Republic.* The *Republic* is written as if by Plato's teacher Socrates; in it, Socrates recounts a long conversation in which he describes an ideal city, as an analogy for the ideal person. Certain citizens of the ideal city will be *guardians;* Socrates describes their education. The following translation from Book VII (524d–525b) is mine, but depends on the translations of Shorey [14] and Waterfield [15]. I

[2] This can be inferred from some other definitions in Book VII of the *Elements:* 'A **prime number** is that which is measured by a unit alone. A **composite number** is that which is measured by some other number.'

have inserted some of the original Greek words, especially[3] those that are origins of English words.

> 'So this is what I was just trying to explain: Some things are *thought-provoking* (παρακλητικὰ τῆς διανοίας), and some are not. Those things are 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

[3] I have also included certain derivatives of the present participle ὄντ- corresponding to the English *being* and the Turkish olan or olur. Addition of the abstract-noun suffix -ία to the feminine form of ὄντ- 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.

order to take hold of being-ness, or else he will never *become a calculator* (λογιστικῷ γενέσθαι).'

'Just so' he said.

'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[4] and an easy turning of the soul itself from becoming towards truth and being-ness.'

'You speak superbly' he said.

Table 1.1: The Greek alphabet

A α	**a**lpha	H η	**ē**ta	N ν	**n**u	T τ	**t**au
B β	**b**eta	Θ ϑ	**th**eta	Ξ ξ	**x**i	Υ υ	**u**psilon
Γ γ	**g**amma	I ι	**i**ota	O ο	**o**micron	Φ φ	**ph**i
Δ δ	**d**elta	K κ	**k**appa	Π π	**p**i	X χ	**ch**i
E ε	**e**psilon	Λ λ	**l**ambda	P ρ	**r**ho	Ψ ψ	**ps**i
Z ζ	**z**eta	M μ	**m**u	Σ σ/ς	**s**igma	Ω ω	**ō**mega

Each Greek letter has a name, written here in Latin letters; the first letter or two of the Latin name provides a transliteration for the Greek letter. The Greek vowels are α, ε, η, ι, ο, υ, ω. In texts, an initial vowel takes a rough-breathing mark (as in ἁ) or a smooth-breathing mark (ἀ); the former mark corresponds to a preceding h; the latter can be ignored. Likewise, ῥ is transliterated as rh. Some vowels may be given tonal accents (ά, ᾶ, ὰ). A terminal ω may have an iota subscript (ῳ). Of the two forms of minuscule sigma, the ς appears at the ends of words; elsewhere, σ appears.

[4]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.

Chapter 2

Logic

2.1 Propositional logic

2.1.1. In all formality, set-theory is a *theory* in a *first-order logic*. Such a logic is based on a **propositional logic,** which consists of:

(i) an *indefinite* number of **(propositional) variables:** P, Q, R, ... (see ¶2.1.5);

(ii) a **signature,** which is a definite number of **(propositional) connectives;**

(iii) brackets: (and).

From the variables, connectives, and brackets, *formulas* are built up *recursively.recursively* Our official signature for propositional logic will comprise two connectives, $\neg$ and $\Rightarrow$. In our logic then, **(propositional) formulas** are given by the following definition:

(i) Every propositional variable is a propositional formula;

(ii) if F is a propositional formula, then so is $\neg F$: this is the **negation** of F;

(iii) If F and G are propositional formulas, then so is $(F \Rightarrow G)$: this is the **implication** of G by[1] F.

The definition is formulas is called **recursive,** because Parts (ii) and (iii) show how to obtain formulas *from other formulas.* We can *use* these parts of the definition, because Part (i) gives us some formulas to start with. The process of constructing a formula can be shown in a *tree* (Figure 2.1). Each

[1] One might read the formula also as the implication of G *in* F. One normally refers to a formula $(F \Rightarrow G)$ merely as an *implication,* without specifying how the sub-formulas F and G are involved in the implication. But we may read the formula as 'F implies G.' The verb *imply* is from the Latin for *fold in;* so the formula $(F \Rightarrow G)$ suggests that G is 'folded into' F, so that, when one 'has' F, then one also has G.

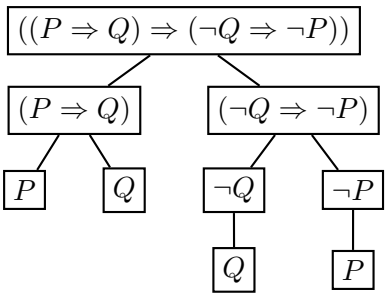

Figure 2.1: Analysis of a propositional formula

node of the tree is a formula, coming from a formula or formulas below it by (ii) or (iii) in the definition; if nothing is below the node, then it is obtained from (i).

2.1.2. In writing formulas, we may follow the convention whereby:

(i) outer parentheses may be removed;

(ii) $(F \Rightarrow G \Rightarrow H)$ means $(F \Rightarrow (G \Rightarrow H))$.

So, for example, the official formula $((P \Rightarrow Q) \Rightarrow (\neg Q \Rightarrow \neg P))$ can be abbreviated as $(P \Rightarrow Q) \Rightarrow \neg Q \Rightarrow \neg P$.

2.1.3. To each variable in a propositional formula, we may assign one of the two **truth-values,** namely **true** and **false.** Then the whole formula becomes true or false under this **truth-assignment,** according to the following rules:

(i) If a formula is a variable, then it takes the truth-value assigned to that variable.

(ii) The formula $\neg F$ is false just in case F is true.

(iii) The formula $F \Rightarrow G$ is false just in case F is true and G is false.

These rules can be expressed by **truth-tables;** here, 0 means false, and 1 means true[2]; the value of a formula is written under its connective:

$\neg$	F
1	0
0	1

F	$\Rightarrow$	G
0	1	0
1	0	0
0	1	1
1	1	1

The truth-table of a complex formula is filled out in stages, as in Table 2.1. The formula in that table is a **tautology:** it is true under every truth--assignment.

[2]Some writers, as Stoll [17, Ch. 4, Exercise 3.7], use 0 and 1 in the opposite sense.

P	$\Rightarrow$	$\neg$	Q	$\Rightarrow$	$\neg$	$(P$	$\Rightarrow$	$Q)$
0			0			0		0
1			0			1		0
0			1			0		1
1			1			1		1
0		1	0			0	1	0
1		1	0			1	0	0
0		0	1			0	1	1
1		0	1			1	1	1
0		1	0		0	0	1	0
1		1	0		1	1	0	0
0		0	1		0	0	1	1
1		0	1		1	1	1	1
0		1	0	0	0	0	1	0
1		1	0	1	1	1	0	0
0		0	1	1	0	0	1	1
1		0	1	1	1	1	1	1
0	1	1	0	0	0	0	1	0
1	1	1	0	1	1	1	0	0
0	1	0	1	1	0	0	1	1
1	1	0	1	1	1	1	1	1

Table 2.1: The filling-out of a truth-table

2.1.4. Two propositional formulas F and G are **equivalent** if, under every truth-assignment, F and G take the same truth-value; in this case, we may write

$$F \sim G.$$

We could show that our signature is **adequate** in the sense that every formula in every signature is equivalent to a formula in our signature. Nonetheless, we may like to use formulas from a larger signature, as abbreviations for formulas of our official signature. In particular:

$$\begin{aligned}(F \vee G) &\sim (\neg F \Rightarrow G);\\ (F \,\&\, G) &\sim \neg(\neg F \vee \neg G);\\ (F \Leftrightarrow G) &\sim (F \Rightarrow G) \,\&\, (G \Rightarrow F).\end{aligned}$$

Then two formulas F and G are equivalent if and only if the formula $F \Leftrightarrow G$ is a tautology. In writing, we may follow the convention whereby $\&$ and $\vee$ are applied before $\Rightarrow$ and $\Leftrightarrow$, so that, for example, $F \Rightarrow G \,\&\, G \Rightarrow F$ stands for $F \Rightarrow (G \,\&\, G) \Rightarrow F$ (which stands for $F \Rightarrow ((G \,\&\, G) \Rightarrow F)$).

2.1.5. I say in ¶2.1.1 that we have an *indefinite* number of propositional variables. Normally we would say that our variables compose an infinite set,

which we might write as $\{P_0, P_1, P_2, \dots\}$ or $\{P_n : n \in \omega\}$ (¶1.3.7). However, we do not *have* ω officially yet. Nor do we have a good way to explain what comes after 2 in the sequence 0, 1, 2, ... In any case, we never need infinitely many variables at once. We could give a formal recursive definition of an infinite set of variables: we might say that P is a variable, and if Q is a variable, then so is Q'. Then our variables are P, P', P'', P''', and so on. The point is that we can always obtain as many variables as we need. Similar considerations will apply in ¶2.2.1.

2.1.6. When we say for example that F is a propositional formula (as in ¶2.1.1), we do not mean that the *letter* F is itself a formula. The letter itself merely *stands for* a formula; the letter is thus a kind of variable. It is not a variable of our propositional logic; it is a variable of our ordinary language, which we are using to talk *about* the logic. If one wants to give it a name, such a variable can be called a **syntactical variable** [1, §08]. Then Q in ¶2.1.5 is also a syntactical variable. I shall not worry further about identifying syntactical variables as such.

2.2 First-order logic

2.2.1. Our logic for talking about sets will be a so-called *first-order* logic. Its symbols will be:

(i) $\in$, the sign of **membership** in a set;[3]

(ii) **(individual) variables:** z, y, x, ...;

(iii) **(individual) constants:** a, b, c, ...;

(iv) the propositional connectives $\neg$ and $\Rightarrow$;

(v) the **existential quantifier,** $\exists$;

(vi) brackets.

A **term** in our logic is an individual variable or constant. In talking about terms, we may symbolize them with letters like t and s. An **atomic formula** is an expression of the form

$$s \in t$$

(where s and t are terms). The **formulas** in general are defined recursively, as in ¶2.1.1, but with an additional possibility:

(i) Atomic formulas are formulas;

(ii) if φ is a formula, then so is $\neg\varphi$;

[3]The $\in$ can be understood as a form of ε (epsilon), standing for the Latin ELEMENTVM.

(iii) if φ and ψ are formulas, then so is $(\varphi \Rightarrow \psi)$;

(iv) if φ is a formula, and x is a variable, then $\exists x\, \varphi$ is a formula.

We may use additional connectives, as in ¶2.1.4.

2.2.2. Every formula has **sub-formulas:**

(i) Every formula is a sub-formula of itself;

(ii) φ is a sub-formula of $\neg\varphi$ and of $\exists x\, \varphi$;

(iii) φ and ψ are sub-formulas of $(\varphi \Rightarrow \psi)$;

(iv) if φ is a sub-formula of ψ, and ψ is a sub-formula of χ, then φ is a sub-formula of χ.

In a formula, some **occurrences** of a variable are **free;** the other occurrences are **bound.** The rules are that every occurrence of x in $\exists x\, \varphi$ is bound, and that those occurrences remain bound when $\exists x\, \varphi$ is used as a sub-formula of another formula. The **free variables** of a formula are those that have free occurrences in a formula (even though they might also have bound occurrences). So x is a free variable of $(\exists x\, x \in x) \Rightarrow x \in y$, although, in this formula, x has three bound occurrences, but only one free occurrence.

2.2.3. If φ is a formula, x is a variable, and t is a term, then we let the expression

$$\varphi_t^x$$

denote the result of replacing each *free* occurrence of x in φ with t. For example, if φ is $x \in y \Rightarrow \exists x\, x \in y$, then φ_a^x is $a \in y \Rightarrow \exists x\, x \in y$. Suppose no other variable than x is free in φ. We may indicate this by writing φ as

$$\varphi(x),$$

and we may call φ a **singulary** or **unary** formula.[4] Also, we can write φ_t^x as

$$\varphi(t).$$

[4]Following Quine, Church [1, § 02, p. 12, n. 29] suggests *singulary* as a more etymologically correct word than *unary.* Indeed, whereas the first five Latin cardinal numbers are UN-, DU-, TRI-, QUATTUOR, QUINQUE, the first five Latin *distributive* numbers—corresponding to the Turkish birer, ikişer, üçer, dörder, beşer [13]—are SINGUL-, BIN-, TERN-, QUATERN-, QUIN-. The latter sequence that gives us *binary, ternary, quaternary,* and *quinary.* So *singulary* appears to be a better word than *unary.* In fact, *singulary* does not appear in the original *Oxford English Dictionary* [12]. The word *unary does* appear in this dictionary, but it is considered obsolete: only one use of the word, from 1576, was discovered in English literature. There, *unary* meant *unit*, although the word *unit* was not actually invented until 1570, when it was introduced by [John] Dee to correspond to the Greek μοναδ-.

2.2.4. A formula with no free variables is a **sentence.** A sentence with no constants is either true or false; a sentence *with* constants *becomes* true or false when those constants are **interpreted** as particular sets. The rules are as follows.

(i) If a and b are constants, then the sentence

$$a \in b$$

is true whenever a and b are understood to denote sets such that the set (denoted by) b contains the set (denoted by) a.

(ii) If σ and τ are sentences, then the truth-values of $\neg\sigma$ and $(\sigma \Rightarrow \tau)$ follow from those of σ and τ according to the rules of propositional logic in ¶2.1.3.

(iii) Suppose $\exists x\ \varphi$ is a sentence, and the constant a does not appear in φ. Then $\exists x\ \varphi$ true just in case the sentence $\varphi(a)$ is true under *some* interpretation of the constant a (as a set).

The qualification about a in (iii) is needed to guard against examples like the following. If φ is $x \in a$, then $\varphi(a)$ is $a \in a$; this may never be true, depending on the axioms that we ultimately choose for our set-theory. But $\exists x\ \varphi$ *is* true, unless a has no members (that are sets) at all. If a does have a set as a member, then we can call it b, so that $\varphi(b)$ is true.

2.2.5. As implied in ¶2.2.4, in our logic, variables and constants refer *only* to sets. This means that the only elements of sets that we can talk about are other sets. Indeed, we shall restrict our attention to the sets whose *only* members are other sets. We shall see that this is not a real limitation, mathematically speaking.

2.2.6. We may write

$$s \notin t$$

instead of $\neg\, s \in t$. Also, we may write

$$\forall x\ \varphi$$

instead of $\neg\exists x\ \neg\varphi$. If this is a *sentence,* then it is true just in case $\varphi(a)$ is true under *every* interpretation of a (as a set), assuming that a does not already appear in φ. Hence, as an alternative to asserting $\forall x\ \varphi$, we may assert $\varphi(a)$ simply, when it is understood that a may be any set. I shall often follow this convention.

2.2.7. A formula $\varphi(x)$ in (at most) one free variable x **defines** a **class** (assuming that any constants in φ have been interpreted). This class **comprises**

those sets a such that $\varphi(a)$ is true; such sets are **members** or **elements** of the class; the class can be denoted

$$\{x\colon \varphi(x)\}$$

(or $\{y\colon \varphi(y)\}$, &c.). Thus a class is something like a set. However, we are defining classes *in terms of* sets. In particular, there is a **universal class** of all sets: that class can be written as

$$\{x\colon x \in x \Rightarrow x \in x\},$$

which we may denote by

$$\mathbf{V}.$$

In our conception, the sets come first; then we recognize that some of them belong to classes. It may be that some classes are already members of $\mathbf{V}$; but we should not expect *all* classes to have this property.

2.2.8. We should not expect all classes to be sets. Yet this mistake was made at the beginning of the study of sets in the nineteenth century; rather, no possibility of a distinction between classes and sets was recognized. This led to problems, discussed below. Hence an *axiomatic* treatment of sets was pursued, as described in ¶1.2.3, in an attempt to avoid the problems. These notes present set-theory as a full-blown axiomatic system; however, its development as such spans several decades of (almost) living memory. Euclid's *Elements* is the classical presentation of an axiomatic system, but it too is given to us full-blown; we do not have any earlier texts to tell us how the idea of erecting a mathematical theory on axioms was discovered.

2.2.9. Two classes are considered **equal** or the **same** if they have the same members, regardless of whether they are defined by the same formula. In a word, classes are equal when they have the same **extension.** For example, $\mathbf{V}$ is also the class $\{x\colon x \in x \vee x \notin x\}$.

2.3 Proof

2.3.1. We have *defined* when a sentence is true, in ¶2.2.4; but how do we *establish* that a particular sentence is true? We have noted, in ¶2.2.8, that perhaps not every class $\{x\colon \varphi(x)\}$ is a set; this class is a set, if the sentence

$$\exists x\ (y \in x \Leftrightarrow \varphi(y))$$

is true; but how can we use the *definition* of truth to tell whether the sentence is true? It was suggested in the beginning (¶1.2.3) that we would use an axiomatic system. Now we shall be able to say what this means precisely.

2.3.2. In the most precise, formal sense, an **axiomatic system** consists of:

(i) **axioms,** which are just certain sentences (in a particular logic);

(ii) **rules of inference,** which are formal ways of obtaining new sentences from given sentences.

The rules of inference can be applied recursively to the axioms; all of the resulting sentences are **theorems** of the system. A **proof** or **deduction** that a particular sentence is a theorem is a tree like Figure 2.1, although one may write down the proof also in a more conventional linear fashion: see ¶2.3.6 below. A useful axiomatic system will at least be **sound:** that is, all of its theorems will be true. The *ideal* sound axiomatic system will also be **complete,** in that *every* true sentence of the relevant logic will be a theorem of the system.

2.3.3. Some sentences are true for logical reasons. For example, suppose we take a tautology of propositional logic, and replace its variables with sentences of our first-order logic. The resulting sentence is also called a **tautology.** A tautology in this sense is true under every interpretation of its constants; it is true, even if $\in$ is understood to indicate something other than membership of one set in another. That is, a tautology is true under an *arbitrary* assignment of truth-values to the atomic sentences.

2.3.4. In general, if a sentence is true under every truth-assignment to the atomic sentences, then the sentence is called a **validity.** So tautologies are validities. But there are validities that are not tautologies: for example, $\forall x\ (x \in x \Rightarrow x \in x)$, or $\forall x\ (\varphi \Rightarrow \psi) \Rightarrow \forall x\ \varphi \Rightarrow \forall x\ \psi$, where φ and ψ are arbitrary formulas (in at most one free variable, x).

2.3.5. A result known as **Gödel's Completeness Theorem** is that the validities are precisely the **theorems** of a certain axiomatic system. The axioms of this system are of three kinds, as follows; here, φ is a singulary formula, and σ is a sentence:

(i) the tautologies;

(ii) $(\varphi(a) \Rightarrow \sigma) \Rightarrow \exists x\ \varphi(x) \Rightarrow \sigma$, where a does not appear in σ;

(iii) $\varphi(a) \Rightarrow \exists x\ \varphi(x)$ (where a is allowed to appear in σ).

The system has one rule of inference, called ***Modus Ponens*** in Latin and **Detachment** in English; the rule allows to obtain the sentence τ from the two sentences σ and $(\sigma \Rightarrow \tau)$. Hence the theorems of the system have the following recursive definition:

(i) the axioms are theorems;

(ii) if σ and $(\sigma \Rightarrow \tau)$ are theorems, then so is τ.

2.3.6. The sentence $\forall x\ (x \in x \Rightarrow x \in x)$ is really an abbreviation for $\neg\exists x\ \neg(x \in x \Rightarrow x \in x)$. Now let $\varphi(x)$ stand for $\neg(x \in x \Rightarrow x \in x)$. In the axiomatic system of ¶2.3.5, the sentence $\neg\exists x\ \varphi(x)$ is a theorem by the proof given in Figure 2.2. Every node follows from the nodes *above* it by

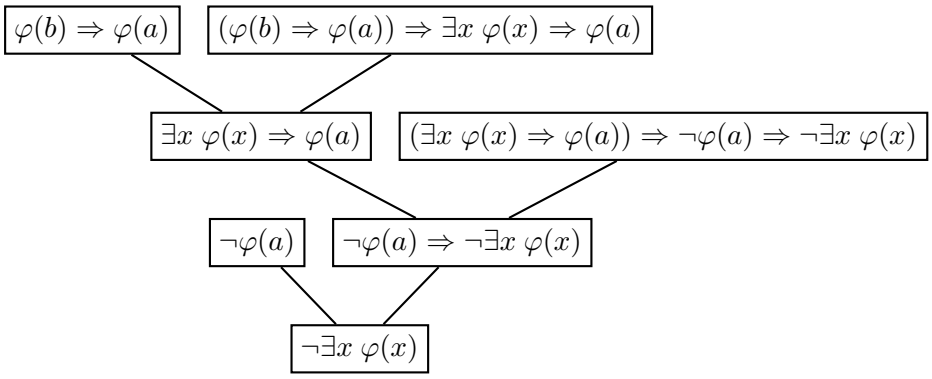

Figure 2.2: A formal proof

Detachment; or if there are no nodes above, it is an axiom. The proof can also be written out line by line, as follows:

(i) $\neg\varphi(a)$ [tautology]

(ii) $\varphi(b) \Rightarrow \varphi(a)$ [tautology]

(iii) $(\varphi(b) \Rightarrow \varphi(a)) \Rightarrow \exists x\ \varphi(x) \Rightarrow \varphi(a)$ [axiom]

(iv) $\exists x\ \varphi(x) \Rightarrow \varphi(a)$ [detachment, lines (ii) and (iii)]

(v) $(\exists x\ \varphi(x) \Rightarrow \varphi(a)) \Rightarrow \neg\varphi(a) \Rightarrow \neg\exists x\ \varphi(x)$ [tautology]

(vi) $\neg\varphi(a) \Rightarrow \neg\exists x\ \varphi(x)$ [detachment, lines (iv) and (v)]

(vii) $\neg\exists x\ \varphi(x)$ [detachment, lines (i) and (vi)]

2.3.7. We are not going to try to prove Gödel's Completeness Theorem. Indeed, the proof would involve some set-theoretical ideas that we have yet to develop. But the Theorem is an *if-and-only-if:* a sentence is a validity if and only if it is a theorem. The *if*-part follows, by an **inductive** proof, made possible by the recursive definition of theorems: The axioms are validities, and if σ and $(\sigma \Rightarrow \tau)$ are validities, then so is τ by ¶2.2.4; therefore every theorem is a validity.

2.3.8. We are looking for more than the validities: we are looking for truths *about sets.* In principle, we can do this by means of an axiomatic system enlarging that of ¶2.3.5. Being validities, the axioms in ¶2.3.5 can be called **logical axioms;** now we shall introduce new axioms that are not validities,

but that express what we understand about sets. From these axioms, along with the logical axioms and the Rule of Detachment, we shall deduce various interesting theorems. We shall not try to write out formal deductions as in ¶2.3.6; but in principle, we could do it.

Chapter 3

Foundations of mathematics

3.1 Equality of sets

3.1.1. The sign of equality (=) is usually included as a logical symbol in first-order logics. Then one has the formulas

$$s = t \tag{3.11}$$

among the atomic formulas, and the sentences

$$\forall x\ x = x,$$
$$\forall x\ \forall y\ (x = y \Rightarrow \varphi(x) \Rightarrow \varphi(y))$$

are taken to be validities, for every singulary atomic formula φ. In particular, in the logic of sets, one has the sentences

$$\forall x\ \forall y\ (x = y \Rightarrow x \in a \Rightarrow y \in a), \tag{3.12}$$
$$\forall x\ \forall y\ (x = y \Rightarrow a \in x \Rightarrow a \in y) \tag{3.13}$$

as validities.

3.1.2. Alternatively, we can reach (3.12) and (3.13) by another route. First, we introduce the formula $s = t$ as an *abbreviation* for

$$\forall x\ (s \in x \Leftrightarrow t \in x) \tag{3.14}$$

where x is neither s nor t. So, by *definition,* two sets are equal if (and only if) they are members of the same sets. Then (3.12) is a consequence of this definition. Indeed, if two sets are such that no set contains one of them without the other, then the two sets would appear to be indistinguishable, set-theoretically speaking; so they might as well be counted as the same. From this definition, (3.12) follows. But then we must rule out the possibility that (3.13) is violated—that two sets with different elements are still equal. This we do with the following—our first axiom of set-theory.

3.1.3 Axiom (Extension)**.** *Two sets are equal if and only if they have the same elements:*

$$a = b \Leftrightarrow \forall x\ (x \in a \Leftrightarrow x \in b). \tag{3.15}$$

3.1.4. So the Extension Axiom is that sets are determined by their elements, as suggested in ¶1.1.1. Another way to say this is that a set is determined by its *extension* (¶2.2.9). Equivalently, all sets are classes; in particular, every set a is equal to the class $\{x\colon x \in a\}$.

3.1.5. We may denote classes by boldface capital letters.[1] (We started doing this in ¶2.2.7 with the universal class, $\mathbf{V}$. The handwritten version of a boldface letter is the letter with a wavy line underneath; so $\mathbf{V}$ can be written by hand as $\underset{\sim}{\mathrm{V}}$.) If we let $\boldsymbol{C}$ denote a class $\{x\colon \varphi(x)\}$, then, in any formula, we may use the expression

$$t \in \boldsymbol{C} \tag{3.16}$$

instead of $\varphi(t)$. Then we can use the sentence

$$\boldsymbol{C} = \boldsymbol{D} \tag{3.17}$$

to stand for the sentence $\forall x\ (x \in \boldsymbol{C} \Leftrightarrow x \in \boldsymbol{D})$. For $\neg\, \boldsymbol{C} = \boldsymbol{D}$, we can write

$$\boldsymbol{C} \neq \boldsymbol{D}.$$

3.1.6. A class $\boldsymbol{C}$ is a **sub-class** of a class $\boldsymbol{D}$ if $\boldsymbol{D}$ contains all members of $\boldsymbol{C}$. In that case, we may say also that $\boldsymbol{D}$ **includes** $\boldsymbol{C}$, or that $\boldsymbol{C}$ is **included in** $\boldsymbol{D}$; also, we write

$$\boldsymbol{C} \subseteq \boldsymbol{D};$$

so this is an abbreviation for $\forall x\ (x \in \boldsymbol{C} \Rightarrow x \in \boldsymbol{D})$. Instead of $\neg\, \boldsymbol{C} \subseteq \boldsymbol{D}$, we may write

$$\boldsymbol{C} \not\subseteq \boldsymbol{D}.$$

The class $\boldsymbol{C}$ is a **proper sub-class** of $\boldsymbol{D}$ if $\boldsymbol{C} \subseteq \boldsymbol{D}$, but $\boldsymbol{C} \neq \boldsymbol{D}$; in that case, we write

$$\boldsymbol{C} \subset \boldsymbol{D}.$$

Since sets are classes by the Extension Axiom (¶3.1.3), we may use the same notation for sets. A sub-class that is also a set can be called a **subset.**

3.1.7. We can rewrite (3.15) as

$$a = b \Leftrightarrow a \subseteq b\ \&\ b \subseteq a.$$

Two sets are equal just in case each one is a subset of the other. (This holds for classes in general, by ¶2.2.9.)

[1] This convention is followed by Kunen [9, Ch. 1, §9], for example, though not by Moschovakis [11, **3.19**]; Levy [10, I.4.1] uses plainface capital letters for classes.

3.1.8 Theorem (Russell Paradox)**.** *Not all classes are sets; in particular, the class*

$$\{x\colon x \notin x\}$$

is not a set: symbolically, $\neg\exists y\, \forall x\, (x \in y \Leftrightarrow x \notin x)$.

Proof. Let $\boldsymbol{R}$ be the given class. It suffices by ¶3.1.7 to show that no subset of $\boldsymbol{R}$ is equal to $\boldsymbol{R}$. Suppose $r \subseteq \boldsymbol{R}$, so that

$$\forall x\, (x \in r \Rightarrow x \in \boldsymbol{R}).$$

Then, in particular, if $r \in r$, then $r \in \boldsymbol{R}$, so $r \notin r$ by definition of $\boldsymbol{R}$. Therefore, logically,

$$r \notin r, \tag{3.18}$$

which, by definition of $\boldsymbol{R}$, means

$$r \in \boldsymbol{R}. \tag{3.19}$$

The last two conclusions—(3.18) and (3.19)—imply $\boldsymbol{R} \not\subseteq r$, so $r \neq \boldsymbol{R}$. □

3.1.9. Often Theorem 3.1.8 is proved by **contradiction** as follows:

> Suppose $\boldsymbol{R}$ is a set. Then $\boldsymbol{R} \in \boldsymbol{R} \Rightarrow \boldsymbol{R} \notin \boldsymbol{R}$ and $\boldsymbol{R} \notin \boldsymbol{R} \Rightarrow \boldsymbol{R} \in \boldsymbol{R}$, so $\boldsymbol{R} \in \boldsymbol{R} \Leftrightarrow \boldsymbol{R} \notin \boldsymbol{R}$, which is absurd. Therefore $\boldsymbol{R}$ is not a set.

This is a valid argument. However, I prefer to avoid proofs by contradiction, for reasons of style. In a proof of $P \Rightarrow Q$ by contradiction, one assumes P and $\neg Q$, and proves an absurdity like $P \,\&\, \neg P$. Often in such proofs, however, one does not need the assumption of P; one really just proves $\neg Q \Rightarrow \neg P$, the **contrapositive** of $P \Rightarrow Q$. Then the needless assumption of P simply prevents anything in the proof from having independent value. By contrast, in the proof given in ¶3.1.8, we happen to learn something more than the truth of the theorem: namely that no subset of $\boldsymbol{R}$ is a member of itself.[2]

3.2 New classes from old

3.2.1. Several **operations** on classes correspond to logical operations on formulas:

(i) The **complement** of $\boldsymbol{C}$ is $\{x\colon x \notin \boldsymbol{C}\}$, denoted by

$$\boldsymbol{C}^{\mathrm{c}}.$$

[2]The argument for this—$r \subseteq \boldsymbol{R}\ \&\ r \in r \Rightarrow r \notin r$—can be understood as using the method of contradiction. However, we still prove $r \notin r$ directly; we do not have to go back and say that our original assumption that $r \in r$ is wrong; we simply use the tautology $(P \Rightarrow \neg P) \Rightarrow \neg P$.

(ii) The **union** of $\boldsymbol{C}$ and $\boldsymbol{D}$ is $\{x\colon x \in \boldsymbol{C} \vee x \in \boldsymbol{D}\}$, denoted by

$$\boldsymbol{C} \cup \boldsymbol{D}.$$

(iii) The **intersection** of $\boldsymbol{C}$ and $\boldsymbol{D}$ is $\{x\colon x \in \boldsymbol{C} \,\&\, x \in \boldsymbol{D}\}$, denoted by

$$\boldsymbol{C} \cap \boldsymbol{D}.$$

(iv) The **empty class** is $\{x\colon x \neq x\}$, denoted by

$$\varnothing.$$

(v) The **difference** of $\boldsymbol{C}$ from $\boldsymbol{D}$ is $\{x\colon x \in \boldsymbol{C} \,\&\, x \notin \boldsymbol{D}\}$ or $\boldsymbol{C} \cap \boldsymbol{D}^{\mathrm{c}}$, denoted also by

$$\boldsymbol{C} \smallsetminus \boldsymbol{D}.$$

(vi) The **symmetric difference** of $\boldsymbol{C}$ and $\boldsymbol{D}$ is $\{x\colon x \in \boldsymbol{C} \Leftrightarrow x \notin \boldsymbol{D}\}$ or $(\boldsymbol{C} \smallsetminus \boldsymbol{D}) \cup (\boldsymbol{D} \smallsetminus \boldsymbol{C})$, denoted also by

$$\boldsymbol{C} \vartriangle \boldsymbol{D}.$$

Hence, for example,

$$\varnothing^{\mathrm{c}} = \mathbf{V},$$
$$\boldsymbol{C} = \boldsymbol{D} \Leftrightarrow \boldsymbol{C} \vartriangle \boldsymbol{D} = \varnothing.$$

Since sets are classes (¶3.1.4), operations on classes can be performed on sets in particular.

3.2.2. Two other operations on classes correspond to *infinitary* logical operations (such logical operations are used in some logics, though not in ours):

(i) The **union** of a single class $\boldsymbol{C}$ is the class $\{x\colon \exists y\, (x \in y \,\&\, y \in \boldsymbol{C})\}$ of elements of the elements of $\boldsymbol{C}$; it is denoted by

$$\bigcup \boldsymbol{C}.$$

(ii) The **intersection** of a class $\boldsymbol{C}$ is the class $\{x\colon \forall y\, (y \in \boldsymbol{C} \Rightarrow x \in y)\}$ of elements common to the elements of $\boldsymbol{C}$; it is denoted by

$$\bigcap \boldsymbol{C}.$$

There is also a class of *subsets* of a class $\boldsymbol{C}$, namely $\{x\colon \forall y\, (y \in x \Rightarrow y \in \boldsymbol{C})\}$ or $\{x\colon x \subseteq \boldsymbol{C}\}$; we may call this the **subset-class** of $\boldsymbol{C}$, and denote it by

$$\mathcal{P}(\boldsymbol{C}).$$

(The subset-class of a *set* will later be called the *power-set* of the set; the power-set will *be* a set, by ¶4.1.8; but for now, it is simply a class.)

3.2.3. We can put two *sets* a and b into a class, namely the class

$$\{x\colon x = a \vee x = b\};$$

this class is commonly denoted

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

This class is going to be a set (¶3.3.6); but without using this, we can still observe:

$$a \cup b = \bigcup\{a, b\}; \tag{3.20}$$

$$a \cap b = \bigcap\{a, b\} \tag{3.21}$$

(Exercise 4). We do not have such equations for classes in general, since we do not have a way to put classes, as such, into other classes. However, $\bigcup a$ is a set; but we do not need to use this until after ¶¶ 4.3.4 and 6.2.3. Meanwhile, if $\boldsymbol{C} \subseteq \mathcal{P}(b)$, then $\bigcup \boldsymbol{C} \subseteq b$, so $\bigcup \boldsymbol{C}$ will be a set by ¶3.3.2 (Exercise 5).

3.3 New sets from old

3.3.1. Classes that are not sets are called **proper classes.** The proof of the Russell Paradox (3.1.8) suggests that proper classes are *too big* to be sets. In the belief that size is the only bar to being a set, we postulate the following *scheme* of axioms: it is a **scheme,** because it comprises one axiom for each singulary formula:

3.3.2 Axiom Scheme (Comprehension)**.** *Every sub-class of a set is a set: For every singulary formula φ,*

$$\exists x\, \forall y\, (y \in x \Leftrightarrow y \in a \,\&\, \varphi(y)). \tag{3.22}$$

3.3.3. The set whose existence is expressed by (3.22) is the intersection

$$a \cap \{x\colon \varphi(x)\};$$

this can be denoted by

$$\{x \in a\colon \varphi(x)\}.$$

We may refer to this as *the* set guaranteed by the sentence, because of the Extension Axiom (¶3.1.3). As a first consequence of the Comprehension-Scheme, we have that, if $a \in \boldsymbol{C}$, then $\bigcap \boldsymbol{C}$ is the *set* $\{x \in a\colon x \in \bigcap \boldsymbol{C}\}$ (Exercise 3). Likewise, $a \smallsetminus \boldsymbol{D}$ is the set $\{x \in a\colon x \notin \boldsymbol{D}\}$.

3.3.4. There is an assumption so basic that we do not bother to state it formally as an axiom.[3] This assumption is that there *are* sets. Hence, by the Comprehension-Scheme, the empty class $\varnothing$ is a set, called the **empty set.** (There is only one empty set, by the Extension Axiom.)

3.3.5. We observed (¶3.2.3) that, from two sets a and b, we can form the class denoted $\{a, b\}$. But if any class is a set, surely this class is:

3.3.6 Axiom (Pairing)**.** *Any two sets are contained in a third:*

$$\exists x\ (a \in x\ \&\ b \in x).$$

3.3.7. As stated, the axiom is merely that some set contains a and b; additional members of the set are not excluded. So the class $\{a, b\}$ is a sub-class of some set. By the Comprehension-Scheme, $\{a, b\}$ is itself a set. This set is an **(unordered) pair.** In case $a = b$, the set is a **singleton,** denoted

$$\{a\}$$

or $\{b\}$. We may use the equation

$$u = \{s, t\}$$

to stand for $\forall x\ (x \in u \Leftrightarrow x = s \vee x = t)$; then

$$\varphi(\{s, t\})$$

stands for $\exists x\ (\varphi(x)\ \&\ x = \{s, t\})$ (which stands for $\exists x\ (\varphi(x)\ \&\ \forall y\ (y \in x \Leftrightarrow y = s \vee y = t))$).

3.4 Relations

3.4.1. A formula with (at most) two free variables is a **binary** formula. Suppose φ is such, with its free variables among x and y. Then we can write φ as

$$\varphi(x, y).$$

In this case, $\varphi(s, t)$ means

$$\varphi^{x\ y}_{s\ t};$$

this is the result of *simultaneously* replacing free occurrences of x with s, and y with t. So it is the same formula as $\varphi^{y\ x}_{t\ s}$, but it need not be the same as $(\varphi^x_s)^y_t$. For example, if φ is $x \in y$, then $\varphi^{x\ y}_{y\ x}$ is $y \in x$, while $(\varphi^x_y)^y_x$ is $x \in x$. Also, $\varphi(x, y)$ will usually be different from $\varphi(y, x)$.

[3] Some writers do, as Kunen [9, I 5, p. 10].

3.4.2. A binary formula determines a class of pairs. Indeed, $\varphi(x, y)$ determines the class

$$\{z\colon \exists x\ \exists y\ (z = \{x, y\}\ \&\ \varphi(x, y))\}.$$

However, this is also the class determined in the same way by $\varphi(x, y) \vee \varphi(y, x)$ (Exercise 6).

3.4.3. For more control, we want to combine a and b into an **ordered pair,** denoted

$$(a, b),$$

such that

$$(a, b) = (c, d) \Leftrightarrow a = c\ \&\ b = d. \tag{3.23}$$

One way (but not the only way) to achieve this is by defining (a, b) as

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

(See Exercises 7 and 8.) Then the class $\{z\colon \exists x\ \exists y\ (z = (x, y)\ \&\ \varphi(x, y))\}$ can be denoted by

$$\{(x, y)\colon \varphi(x, y)\};$$

this is the class **defined by** $\varphi(x, y)$.

3.4.4. In particular, we now we have a new operation on classes: the **Cartesian product** of $\boldsymbol{C}$ and $\boldsymbol{D}$ is

$$\{(x, y)\colon x \in \boldsymbol{C}\ \&\ y \in \boldsymbol{D}\};$$

this class is denoted

$$\boldsymbol{C} \times \boldsymbol{D},$$

and it is a sub-class of $\mathcal{P}(\mathcal{P}(\boldsymbol{C} \cup \boldsymbol{D}))$ (Exercise 9). The class $\{(x, y)\colon \varphi(x, y)\}$ is a sub-class of $\mathbf{V} \times \mathbf{V}$.

3.4.5. A sub-class of $\mathbf{V} \times \mathbf{V}$ is a **(binary) relation.** If $\boldsymbol{R}$ is a binary relation, then we usually write

$$s\ \boldsymbol{R}\ t$$

instead of $(s, t) \in \boldsymbol{R}$. The **domain** of $\boldsymbol{R}$ is the class $\{x\colon \exists y\ x\ \boldsymbol{R}\ y\}$; this can be denoted by

$$\mathrm{dom}(\boldsymbol{R}).$$

The **range** of $\boldsymbol{R}$ is the class $\{y\colon \exists x\ x\ \boldsymbol{R}\ y\}$; this can be denoted by

$$\mathrm{rng}(\boldsymbol{R}).$$

The union of the domain and range is the **field:**

$$\mathrm{dom}(\boldsymbol{R}) \cup \mathrm{rng}(\boldsymbol{R}) = \mathrm{fld}(\boldsymbol{R}).$$

3.5 Kinds of relations

3.5.1. A binary relation $\boldsymbol{R}$ has a **converse,** $\{(x, y)\colon y\ \boldsymbol{R}\ x\}$, which can be denoted by

$$\breve{\boldsymbol{R}}.$$

Then

$$\mathrm{dom}(\breve{\boldsymbol{R}}) = \mathrm{rng}(\boldsymbol{R}), \tag{3.24}$$

$$\mathrm{rng}(\breve{\boldsymbol{R}}) = \mathrm{dom}(\boldsymbol{R}), \tag{3.25}$$

$$\breve{\breve{\boldsymbol{R}}} = \boldsymbol{R} \tag{3.26}$$

(Exercise 11). If $\breve{\boldsymbol{R}} = \boldsymbol{R}$, then $\boldsymbol{R}$ is called **symmetric;** this means

$$a\ \boldsymbol{R}\ b \Leftrightarrow b\ \boldsymbol{R}\ a.$$

Then $\boldsymbol{R}$ is **symmetric on** $\boldsymbol{C}$ if $\boldsymbol{R} \cap (\boldsymbol{C} \times \boldsymbol{C})$ is symmetric.

3.5.2. If $\boldsymbol{R}$ and $\boldsymbol{S}$ are both binary relations, then the relation

$$\{(x, z)\colon \exists y\ (x\ \boldsymbol{R}\ y\ \&\ y\ \boldsymbol{S}\ z)\}$$

is the **composite** of $\boldsymbol{R}$ and $\boldsymbol{S}$, denoted[4]

$$\boldsymbol{R}/\boldsymbol{S}.$$

Logically, if $\boldsymbol{T}$ is also a relation, then

$$(\boldsymbol{R}/\boldsymbol{S})/\boldsymbol{T} = \boldsymbol{R}/(\boldsymbol{S}/\boldsymbol{T}). \tag{3.27}$$

If $\boldsymbol{R}/\boldsymbol{R} \subseteq \boldsymbol{R}$, then $\boldsymbol{R}$ is called **transitive;** this means

$$a\ \boldsymbol{R}\ b\ \&\ b\ \boldsymbol{R}\ c \Rightarrow a\ \boldsymbol{R}\ c. \tag{3.28}$$

We may then write

$$a\ \boldsymbol{R}\ b\ \boldsymbol{R}\ c$$

instead of $a\ \boldsymbol{R}\ b\ \&\ b\ \boldsymbol{R}\ c$. More generally, if $\boldsymbol{R} \cap (\boldsymbol{C} \times \boldsymbol{C})$ is transitive, then $\boldsymbol{R}$ is **transitive on** $\boldsymbol{C}$.

3.5.3. For any class $\boldsymbol{C}$, the relation $\{(x, y)\colon x = y\ \&\ x \in \boldsymbol{C}\}$ is the **diagonal** on $\boldsymbol{C}$, denoted

$$\Delta_{\boldsymbol{C}}.$$

Then

$$\Delta_{\mathrm{dom}(\boldsymbol{R})} \subseteq \boldsymbol{R}/\breve{\boldsymbol{R}}, \tag{3.29}$$

$$\Delta_{\mathrm{rng}(\boldsymbol{R})} \subseteq \breve{\boldsymbol{R}}/\boldsymbol{R}, \tag{3.30}$$

for every binary relation $\boldsymbol{R}$ (Exercise 12). Also, $\boldsymbol{R}$ is called:

[4]Tarski [19, § 28, p. 92] uses the notation $\boldsymbol{R}/\boldsymbol{S}$ and refers to the indicated class as the *relative product* of $\boldsymbol{R}$ and $\boldsymbol{S}$. Suppes [18, § 3.1, Definition 7, p. 63] also uses the notation.

(i) **reflexive,** if $\Delta_{\mathbf{V}} \subseteq \boldsymbol{R}$, that is, $a \boldsymbol{R} a$;

(ii) **irreflexive,** if $\boldsymbol{R} \cap \Delta_{\mathbf{V}} = \varnothing$, that is, $\neg\, a \boldsymbol{R} a$;

(iii) **anti-symmetric,** if $\boldsymbol{R} \cap \breve{\boldsymbol{R}} \subseteq \Delta_{\mathbf{V}}$, that is, $a \boldsymbol{R} b \,\&\, b \boldsymbol{R} a \Rightarrow a = b$.

There are relative versions. The relation $\boldsymbol{R}$ is:

(i) **reflexive on $\boldsymbol{C}$**, if $\Delta_{\boldsymbol{C}} \subseteq \boldsymbol{R}$;

(ii) **irreflexive on $\boldsymbol{C}$**, if $\boldsymbol{R} \cap \Delta_{\boldsymbol{C}} = \varnothing$;

(iii) **anti-symmetric on $\boldsymbol{C}$**, if $\boldsymbol{R} \cap \breve{\boldsymbol{R}} \cap (\boldsymbol{C} \times \boldsymbol{C}) \subseteq \Delta_{\boldsymbol{C}}$.

3.5.4. A relation is an **equivalence-relation on $\boldsymbol{C}$** if it is reflexive, symmetric and transitive on $\boldsymbol{C}$. Then **equality,** understood as $\Delta_{\mathbf{V}}$, is an equivalence-relation on every class (Exercise 13). The diagonal $\Delta_{\boldsymbol{C}}$ is an equivalence-relation on every sub-class of $\boldsymbol{C}$.

3.5.5. The relation $\boldsymbol{R}$ is:

(i) an **ordering of $\boldsymbol{C}$**, if it is anti-symmetric, transitive, and either reflexive or irreflexive, on $\boldsymbol{C}$;

(ii) a **total ordering of $\boldsymbol{C}$**, if it is an ordering of $\boldsymbol{C}$ and

$$\boldsymbol{R} \cup \breve{\boldsymbol{R}} \cup \Delta_{\boldsymbol{C}} = \boldsymbol{C} \times \boldsymbol{C}.$$

An ordering in the present sense is often called a *partial ordering*, even though it might be total. An *irreflexive* ordering is also called a **strict ordering.** The converse of an ordering is an ordering. If $\boldsymbol{R}$ is a reflexive ordering of $\boldsymbol{C}$, then there is a corresponding strict ordering of $\boldsymbol{C}$, namely $\boldsymbol{R} \smallsetminus \Delta_{\boldsymbol{C}}$ (or $\boldsymbol{R} \smallsetminus \Delta_{\mathbf{V}}$, for example; it doesn't matter what $\boldsymbol{R} \smallsetminus (\boldsymbol{C} \times \boldsymbol{C})^{\mathrm{c}}$ is). A strict ordering $\boldsymbol{S}$ of $\boldsymbol{C}$ has the corresponding reflexive ordering $\boldsymbol{S} \cup \Delta_{\boldsymbol{C}}$.

3.5.6. There are standard examples:

(i) **Inclusion**—the class $\{(x, y) \colon x \subseteq y\}$, which can be denoted by $\subseteq$ alone—is a reflexive ordering of every class.

(ii) The converse of inclusion is usually denoted by $\supseteq$; it is a reflexive ordering, by the comment in ¶3.5.3.

(iii) **Proper inclusion**—the class $\{(x, y) \colon x \subset y\}$, or $\subset$—is a strict ordering.

It will take more work to define a good example of a total ordering. Often a reflexive ordering is symbolized by $\leqslant$; then the corresponding strict ordering is denoted by $<$. The converse of $\leqslant$ is $\geqslant$; the converse of $<$ is $>$. **Containment**—$\{(x, y) \colon x \in y\}$ or $\in$—is not yet an example of anything in particular; but it will be. (See ¶¶ 3.5.9 and 4.5.6.)

3.5.7. Suppose $<$ is a strict ordering of $\boldsymbol{C}$. An element a of $\boldsymbol{C}$ is **minimal** with respect to $<$, or **$<$-minimal,** if

$$b \in \boldsymbol{C} \Rightarrow \ \neg\, b < a.$$

A $>$-minimal element is **$<$-maximal.** An element a of $\boldsymbol{C}$ is **least** with respect to $<$, or **$<$-least,** if

$$b \in \boldsymbol{C} \Rightarrow a \leqslant b.$$

A $>$-least element is **$<$-greatest.** Least elements are minimal elements. Least elements are unique when they exist; but they need not exist. With respect to a total ordering, a minimal element is a least element.

3.5.8. An **initial segment** of $\boldsymbol{C}$ (with respect to the strict ordering $<$) is a sub-class $\boldsymbol{D}$ of $\boldsymbol{C}$ such that

$$a \in \boldsymbol{D} \ \&\ b \in \boldsymbol{C} \ \&\ b < a \Rightarrow b \in \boldsymbol{D}.$$

We may say that initial segments are **closed under** $<$. A **proper initial segment** is an initial segment that is a proper sub-class.

3.5.9. The class $\boldsymbol{C}$ is **well-ordered** by $<$ if:

(i) $<$ is a strict total ordering of $\boldsymbol{C}$;

(ii) every proper initial segment of $\boldsymbol{C}$ with respect to $<$ is a set;

(iii) every non-empty subset of $\boldsymbol{C}$ has a $<$-least element.

In this case, every non-empty sub-*class* of $\boldsymbol{C}$ has a least element: Indeed, if $\boldsymbol{D}$ is such a sub-class, and it contains only a, then a is its least element; but if $\boldsymbol{D}$ also has another element, then we may assume that this is greater than a, so that $\{x\colon x \in \boldsymbol{C} \ \&\ x \leqslant a\}$ is a set (being a proper initial segment). Then $\{x\colon x \in \boldsymbol{D} \ \&\ x \leqslant a\}$ is a non-empty subset of $\boldsymbol{C}$, and its least element is the least element of $\boldsymbol{D}$.

3.6 Functions

3.6.1. The various operations on classes defined in ¶¶ 3.2.1 and 3.2.2 are examples of **functions.** The union-operation is the function by which the class $\boldsymbol{C} \cup \boldsymbol{D}$ is obtained from the classes $\boldsymbol{C}$ and $\boldsymbol{D}$; the subset-class operation is a function converting $\boldsymbol{C}$ into $\mathcal{P}(\boldsymbol{C})$. In this sense, a function is not a set or a class; it is a feature of our logic.

3.6.2. Binary relations always determine certain functions. Suppose $\boldsymbol{R}$ is a binary relation. Then we can make the definitions:

$$a\boldsymbol{R} = \{x\colon a\ \boldsymbol{R}\ x\},$$
$$\boldsymbol{R}a = \{x\colon x\ \boldsymbol{R}\ a\}.$$

The classes $a\boldsymbol{R}$ and $\boldsymbol{R}a$ are functions of the set a. Some kinds of relations can be understood in terms of these functions. For example, the relation $\boldsymbol{R}$ is symmetric if and only if $\forall x\ x\boldsymbol{R} = \boldsymbol{R}x$ (Exercise 14). If $\boldsymbol{E} \subseteq \boldsymbol{C} \times \boldsymbol{C}$ and is an equivalence-relation on $\boldsymbol{C}$, and $a \in \boldsymbol{C}$, then $a\boldsymbol{E}$ is the **equivalence-class** of a with respect to $\boldsymbol{E}$, or the **$\boldsymbol{E}$-class** of a. Then a is a **representative** of this class; every other member of the class is also a representative. There is something denoted

$$\boldsymbol{C}/\boldsymbol{E},$$

whose members are the equivalence-classes of the elements of $\boldsymbol{C}$. Then there is a function that converts every set in $\boldsymbol{C}$ into its equivalence-class, which is in $\boldsymbol{C}/\boldsymbol{E}$. However, since the elements of $\boldsymbol{C}/\boldsymbol{E}$ may be proper classes, we do not necessarily have $\boldsymbol{C}/\boldsymbol{E}$ even as a class. We can still work with it though; we may call it a **virtual class.** See § 8.2.

3.6.3. Often it is *sets* that are functions of other sets. A binary relation $\boldsymbol{F}$ is **functional** if

$$\breve{\boldsymbol{F}}/\boldsymbol{F} \subseteq \Delta_{\mathbf{V}},$$

that is,

$$a\ \boldsymbol{F}\ b\ \&\ a\ \boldsymbol{F}\ c \Rightarrow b = c. \tag{3.31}$$

This is equivalent to $\breve{\boldsymbol{F}}/\boldsymbol{F} = \Delta_{\mathrm{rng}(\boldsymbol{F})}$ (Exercise 15). In this case, $\boldsymbol{F}$ is a **function on** its domain. Suppose that domain is $\boldsymbol{C}$, and $\mathrm{rng}(\boldsymbol{F}) \subseteq \boldsymbol{D}$. Then we may write either of

$$\boldsymbol{F}\colon \boldsymbol{C} \to \boldsymbol{D},$$
$$\boldsymbol{C} \xrightarrow{\boldsymbol{F}} \boldsymbol{D};$$

these are abbreviations of the sentence

$$\forall x\ ((\exists y\ x\ \boldsymbol{F}\ y \Rightarrow x \in \boldsymbol{C})\ \&\ (x \in \boldsymbol{C} \Rightarrow \exists y\ (x\ \boldsymbol{F}\ y\ \&\ y \in \boldsymbol{D}\ \&\ \forall z\ (x\ \boldsymbol{F}\ z \Rightarrow y = z)))).$$

We may also say that $\boldsymbol{F}$ is a function **from $\boldsymbol{C}$ to $\boldsymbol{D}$**. If $a\ \boldsymbol{F}\ b$, then we usually write

$$\boldsymbol{F}(a) = b.$$

This is consistent with our definition of equality of sets, by which equality is an equivalence-relation (¶3.5.4): now the implication (3.31) becomes

$$\boldsymbol{F}(a) = b\ \&\ \boldsymbol{F}(a) = c \Rightarrow b = c.$$

As an alternative notation for $\boldsymbol{F}$ itself, we may write

$$x \mapsto \boldsymbol{F}(x).$$

3.6.4. We can produce a few examples of functions in the sense of ¶3.6.3:

(i) Most basic is the **identity function,** which is $x \mapsto x$ or $\Delta_{\mathbf{V}}$; considered as a function, this can be denoted by

$$\mathrm{id}_{\mathbf{V}}\,.$$

The **identity on** $\boldsymbol{C}$ is $\Delta_{\boldsymbol{C}}$, usually denoted by

$$\mathrm{id}_{\boldsymbol{C}}\,.$$

(ii) If $\boldsymbol{F}$ is a function, then so is $x \mapsto (x, \boldsymbol{F}(x))$: its domain is $\mathrm{dom}(\boldsymbol{F})$, and its range is $\boldsymbol{F}$.

(iii) If a is a set, then $x \mapsto a$ is a **constant function,** with domain $\mathbf{V}$.

(iv) If $\boldsymbol{F}\colon \boldsymbol{C} \to \boldsymbol{D}$, and $\boldsymbol{E} \subseteq \boldsymbol{C}$, then $\boldsymbol{F} \cap (\boldsymbol{E} \times \mathbf{V})$ (which is $\boldsymbol{F} \cap (\boldsymbol{E} \times \boldsymbol{D})$) is a function with domain $\boldsymbol{E}$ called the **restriction** of $\boldsymbol{F}$ to $\boldsymbol{E}$ and denoted

$$\boldsymbol{F} \restriction \boldsymbol{E}.$$

In particular, $\mathrm{id}_{\mathbf{V}} \restriction \boldsymbol{C}$ is $\mathrm{id}_{\boldsymbol{C}}$.

(v) By (a special case of) the Pairing Axiom, we have a function $x \mapsto \{x\}$ on $\mathbf{V}$.

(vi) By ¶3.4.3, there is a function $\boldsymbol{F}$ on $\mathbf{V} \times \mathbf{V}$ such that $\boldsymbol{F}(b) = a \Leftrightarrow \exists x\; b = (a, x)$; this function can be denoted

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

it is **projection** onto the first coordinate. Likewise, there is $(x, y) \mapsto y$.

(vii) Hence, by Pairing again, we have a function $(x, y) \mapsto \{x, y\}$.

3.6.5. If $\boldsymbol{F}$ and $\boldsymbol{G}$ are functional relations, and $\mathrm{rng}(\boldsymbol{F}) \subseteq \mathrm{dom}(\boldsymbol{G})$, then the composite $\boldsymbol{F}/\boldsymbol{G}$ is a function on $\mathrm{dom}(\boldsymbol{F})$; usually this function is denoted by either of

$$(\boldsymbol{G} \circ \boldsymbol{F}),$$
$$x \mapsto \boldsymbol{G}(\boldsymbol{F}(x)).$$

If also $\boldsymbol{H}$ is functional, and $\mathrm{rng}(\boldsymbol{G}) \subseteq \mathrm{dom}(\boldsymbol{H})$, then, as we have (3.27), so

$$\boldsymbol{H} \circ (\boldsymbol{G} \circ \boldsymbol{F}) = (\boldsymbol{H} \circ \boldsymbol{G}) \circ \boldsymbol{F}. \tag{3.32}$$

3.6.6. Suppose $\boldsymbol{F}\colon \boldsymbol{C} \to \boldsymbol{D}$, and $\boldsymbol{F}/\breve{\boldsymbol{F}} \subseteq \Delta_{\mathbf{V}}$, equivalently, $\boldsymbol{F}/\breve{\boldsymbol{F}} = \Delta_{\mathrm{dom}(\boldsymbol{F})}$ (Exercise 16). This means

$$\boldsymbol{F}(a) = \boldsymbol{F}(b) \Rightarrow a = b.$$

Then $\boldsymbol{F}$ is called an **injective** function, or an **injection** from $\boldsymbol{C}$ (in)to $\boldsymbol{D}$, or an **embedding** of $\boldsymbol{C}$ in $\boldsymbol{D}$; we may write

$$\boldsymbol{F}\colon \boldsymbol{C} \rightarrowtail \boldsymbol{D}.$$

(So the tail of the arrow indicates injectivity.) The converse of an injective function is also a function (Exercise 17), called the **inverse** of the function; when it exists, the inverse of $\boldsymbol{F}$ is denoted

$$\boldsymbol{F}^{-1}.$$

3.6.7. Suppose again $\boldsymbol{F}\colon \boldsymbol{C} \to \boldsymbol{D}$. If $\boldsymbol{D} = \mathrm{rng}(\boldsymbol{F})$, then $\boldsymbol{F}$ is **surjective onto $\boldsymbol{D}$** (or a **surjection onto $\boldsymbol{D}$**); we may write

$$\boldsymbol{F}\colon \boldsymbol{C} \twoheadrightarrow \boldsymbol{D}.$$

(So the second head of the arrow indicates surjectivity.) Note well that a function cannot be surjective simply; it is only surjective with respect to the set that the function is surjective *onto* (namely its *range*). If $\boldsymbol{F}$ is injective, and surjective onto $\boldsymbol{D}$, then $\boldsymbol{F}$ is a **bijection** from $\boldsymbol{C}$ to $\boldsymbol{D}$, and we may write

$$\boldsymbol{F}\colon \boldsymbol{C} \rightarrowtail\!\!\!\!\twoheadrightarrow \boldsymbol{D}.$$

3.6.8 Theorem. *Suppose $\boldsymbol{F}\colon \boldsymbol{C} \to \boldsymbol{D}$.*

(i) $\boldsymbol{F}$ is injective if and only if $\boldsymbol{C}$ is empty or there is a function $\boldsymbol{G}$ from $\boldsymbol{D}$ to $\boldsymbol{C}$ such that $\boldsymbol{G} \circ \boldsymbol{F} = \mathrm{id}_{\boldsymbol{C}}$.

(ii) $\boldsymbol{F}$ is a bijection from $\boldsymbol{C}$ to $\boldsymbol{D}$ if and only if there is a function $\boldsymbol{G}$ from $\boldsymbol{D}$ to $\boldsymbol{C}$ such that $\boldsymbol{G} \circ \boldsymbol{F} = \mathrm{id}_{\boldsymbol{C}}$ and $\boldsymbol{F} \circ \boldsymbol{G} = \mathrm{id}_{\boldsymbol{D}}$.

Proof. Exercise 18. □

3.7 Functions from functions

3.7.1. A function $\boldsymbol{F}$ induces two functions on classes:

(i) If $\boldsymbol{E}$ is a sub-class of $\mathrm{dom}(\boldsymbol{F})$, then the class

$$\{y\colon \exists x\ (x \in \boldsymbol{E}\ \&\ \boldsymbol{F}(x) = y)\} \tag{3.33}$$

is the **image** of $\boldsymbol{E}$ under $\boldsymbol{F}$ and can be denoted by either of

$$\boldsymbol{F}[\,\boldsymbol{E}\,],$$
$$\{\boldsymbol{F}(x)\colon x \in \boldsymbol{E}\}.$$

Then $\mathrm{rng}(\boldsymbol{F}) = \boldsymbol{F}[\,\mathrm{dom}(\boldsymbol{F})\,]$.

(ii) The class

$$\{x\colon x \in \mathrm{dom}(\boldsymbol{F}) \mathbin{\&} \boldsymbol{F}(x) \in \boldsymbol{E}\} \tag{3.34}$$

is the **pre-image** of $\boldsymbol{E}$ under $\boldsymbol{F}$ and can be denoted by

$$\boldsymbol{F}^{-1}[\,\boldsymbol{E}\,].$$

Pre-images of all classes exist, regardless of whether the function $\boldsymbol{F}$ itself has an inverse (¶3.6.6). In particular, $\mathrm{dom}(\boldsymbol{F}) = \boldsymbol{F}^{-1}[\,\boldsymbol{E}\,]$ whenever $\mathrm{rng}(\boldsymbol{F}) \subseteq \boldsymbol{E}$.

Note the great difference in form between (3.33) and (3.34). The difference is reflected in the following.

3.7.2 Theorem. *Suppose $\boldsymbol{F}$ is a function. Then*

$$\boldsymbol{F}[\,\boldsymbol{C} \cup \boldsymbol{D}\,] = \boldsymbol{F}[\,\boldsymbol{C}\,] \cup \boldsymbol{F}[\,\boldsymbol{D}\,], \tag{3.35}$$

$$\boldsymbol{F}[\,\boldsymbol{C} \cap \boldsymbol{D}\,] \subseteq \boldsymbol{F}[\,\boldsymbol{C}\,] \cap \boldsymbol{F}[\,\boldsymbol{D}\,], \tag{3.36}$$

$$\boldsymbol{F}[\,\mathrm{dom}(\boldsymbol{F}) \setminus \boldsymbol{C}\,] \supseteq \mathrm{rng}(\boldsymbol{F}) \setminus \boldsymbol{F}[\,\boldsymbol{C}\,] \tag{3.37}$$

for all sub-classes $\boldsymbol{C}$ and $\boldsymbol{D}$ of $\mathrm{dom}(\boldsymbol{F})$*, and*

$$\boldsymbol{F}^{-1}[\,\boldsymbol{C} \cup \boldsymbol{D}\,] = \boldsymbol{F}^{-1}[\,\boldsymbol{C}\,] \cup \boldsymbol{F}^{-1}[\,\boldsymbol{D}\,], \tag{3.38}$$

$$\boldsymbol{F}^{-1}[\,\boldsymbol{C} \cap \boldsymbol{D}\,] = \boldsymbol{F}^{-1}[\,\boldsymbol{C}\,] \cap \boldsymbol{F}^{-1}[\,\boldsymbol{D}\,], \tag{3.39}$$

$$\boldsymbol{F}^{-1}[\,\mathrm{rng}(\boldsymbol{F}) \setminus \boldsymbol{C}\,] = \mathrm{dom}(\boldsymbol{F}) \setminus \boldsymbol{F}^{-1}[\,\boldsymbol{C}\,] \tag{3.40}$$

for all classes $\boldsymbol{C}$ and $\boldsymbol{D}$. Moreover, the following statements are equivalent:

(i) $\boldsymbol{F}$ *is injective;*

(ii) $\boldsymbol{C} \cup \boldsymbol{D} \subseteq \mathrm{dom}(\boldsymbol{F}) \Rightarrow \boldsymbol{F}[\,\boldsymbol{C} \cap \boldsymbol{D}\,] = \boldsymbol{F}[\,\boldsymbol{C}\,] \cap \boldsymbol{F}[\,\boldsymbol{D}\,]$ *for all classes $\boldsymbol{C}$ and $\boldsymbol{D}$;*

(iii) $\boldsymbol{C} \subseteq \mathrm{dom}(\boldsymbol{F}) \Rightarrow \boldsymbol{F}[\,\mathrm{dom}(\boldsymbol{F}) \setminus \boldsymbol{C}\,] = \mathrm{rng}(\boldsymbol{F}) \setminus \boldsymbol{F}[\,\boldsymbol{C}\,]$ *for all classes $\boldsymbol{C}$;*

(iv) $a \cup b \subseteq \mathrm{dom}(\boldsymbol{F}) \Rightarrow \boldsymbol{F}[\,a \cap b\,] = \boldsymbol{F}[\,a\,] \cap \boldsymbol{F}[\,b\,]$ *for all sets a and b;*

(v) $a \subseteq \mathrm{dom}(\boldsymbol{F}) \Rightarrow \boldsymbol{F}[\,\mathrm{dom}(\boldsymbol{F}) \setminus a\,] = \mathrm{rng}(\boldsymbol{F}) \setminus \boldsymbol{F}[\,a\,]$ *for all sets a.*

Proof. Exercise 20. □

3.7.3. If $\mathrm{dom}(\boldsymbol{F})$ is a set, then $\boldsymbol{F}^{-1}[\,\boldsymbol{E}\,]$ is a set, by the Comprehension-Scheme (¶3.3.2). Likewise, if $\mathrm{rng}(\boldsymbol{F})$ is a set, and $\boldsymbol{E} \subseteq \mathrm{dom}(\boldsymbol{F})$, then $\boldsymbol{F}[\,\boldsymbol{E}\,]$ is a set. However, possibly $\mathrm{rng}(\boldsymbol{F})$ is a set, while $\mathrm{dom}(\boldsymbol{F})$ is a proper class: consider a constant function. However, when we noted in ¶3.3.1 that some classes are too big to be sets, we also suggested that, if a class is *not* too big to be a set, then it *is* a set. Presumably the range of a function is not bigger than its domain, since the range contains no more than one element for each element of the domain; so we postulate the following.

3.7.4 Axiom Scheme (Replacement)**.** *The image of a set under a function is a set:*

$$a \subseteq \operatorname{dom}(\boldsymbol{F}) \Rightarrow \exists x\, x = \boldsymbol{F}[\,a\,]$$

for all functions $\boldsymbol{F}$*, so that the class* $\{\boldsymbol{F}(x)\colon x \in a\}$ *is a set whenever* a *is a subset of* $\operatorname{dom}(\boldsymbol{F})$*.*

3.7.5 Lemma.

(i) A function whose domain is a set is a set.

(ii) The domain of an injection into a set is a set.

(iii) Suppose $f\colon a \to b$*. Then* $x \mapsto f[\,x\,]$ *is a (well-defined) function from* $\mathcal{P}(a)$ *to* $\mathcal{P}(b)$*, and* $y \mapsto f^{-1}[\,y\,]$ *is a well-defined function from* $\mathcal{P}(b)$ *to* $\mathcal{P}(a)$*. Moreover,*

$$f[\bigcup c] = \bigcup\{f[\,x\,]\colon x \in c\},$$
$$f[\bigcap c] \subseteq \bigcap\{f[\,x\,]\colon x \in c\}$$

for all subsets c *of* $\mathcal{P}(a)$*; and*

$$f^{-1}[\bigcup c] = \bigcup\{f^{-1}[\,x\,]\colon x \in c\},$$
$$f^{-1}[\bigcap c] = \bigcap\{f^{-1}[\,x\,]\colon x \in c\}$$

for all subsets c *of* $\mathcal{P}(b)$*.*

Proof. Exercise 22. □

Exercises

(1) Prove carefully that $\{a\} = \{b\} \Leftrightarrow a = b$.

(2) What are $\bigcap \varnothing$ and $\bigcup \varnothing$?

(3) Show that the intersection of a non-empty class is a set (¶3.3.3).

(4) Show (3.21) and (3.20).

(5) Show that, if $\boldsymbol{C} \subseteq \mathcal{P}(a)$, then $\bigcup \boldsymbol{C}$ is a set.

(6) Show that the class determined by $\varphi(x, y)$ in the sense of ¶3.4.2 is the same as the class determined by $\varphi(x, y) \vee \varphi(x, y)$.

(7) Show the definition of ordered pair in ¶3.4.3 causes (3.23) to be true.

(8) Show that the same is true if (a, b) is defined as $\{\{\varnothing, \{a\}\}, \{\{b\}\}\}$.

(9) Show that $\boldsymbol{C} \times \boldsymbol{D} \subseteq \mathcal{P}(\mathcal{P}(\boldsymbol{C} \cup \boldsymbol{D}))$.

(10) For each of the listed classes, using only the symbols listed in ¶2.2.1, write out a formula that defines the same class. Treat $\boldsymbol{C}$ as $\{x\colon \varphi(x)\}$, and $\boldsymbol{D}$ as $\{x\colon \psi(x)\}$.

(a) $\varnothing$

(b) $\boldsymbol{C} \vartriangle \boldsymbol{D}$

(c) $\boldsymbol{C} \times \boldsymbol{D}$

(11) Verify (3.24), (3.25), and (3.26).

(12) Verify (3.29) and (3.30).

(13) Prove that $\Delta_{\mathbf{V}}$ is an equivalence-relation using only the definition of equality in ¶3.1.2 (and not for example the Extension Axiom).

(14) Prove the claim in ¶3.6.2 that the relation $\boldsymbol{R}$ is symmetric if and only if $\forall x\ x\boldsymbol{R} = \boldsymbol{R}x$.

(15) Prove that a binary relation $\boldsymbol{F}$ is functional if and only if $\breve{\boldsymbol{F}}/\boldsymbol{F} = \Delta_{\mathrm{rng}(\boldsymbol{F})}$.

(16) If $\boldsymbol{R}$ is a binary relation, prove that $\boldsymbol{R}/\breve{\boldsymbol{R}} \subseteq \Delta_{\mathbf{V}}$ if and only if $\boldsymbol{R}/\breve{\boldsymbol{R}} = \Delta_{\mathrm{dom}(\boldsymbol{R})}$.

(17) Show that the converse of an injective function is a function.

(18) Prove Theorem 3.6.8.

(19) Suppose $\boldsymbol{F}\colon \boldsymbol{C} \twoheadrightarrow \boldsymbol{D}$. Is there a function $\boldsymbol{G}$ from $\boldsymbol{D}$ to $\boldsymbol{C}$ such that $\boldsymbol{F} \circ \boldsymbol{G} = \mathrm{id}_{\boldsymbol{D}}$?

(20) Prove Theorem 3.7.2.

(21) Write out the Replacement-Scheme (¶3.7.4) using only the symbols of ¶2.2.1, and using an *arbitrary* binary formula $\varphi(x, y)$ instead of $\boldsymbol{F}$ (so your sentence will have to express the condition that $\varphi(x, y)$ defines a functional relation).

(22) Prove Lemma 3.7.5.

Chapter 4

Size and order

4.1 Cardinality

4.1.1. Two classes $\boldsymbol{C}$ and $\boldsymbol{D}$ have the **same cardinality** (or the **same size**) if there is a bijection between them; in that case, we may write

$$\boldsymbol{C} \approx \boldsymbol{D};$$

we may also say that $\boldsymbol{C}$ and $\boldsymbol{D}$ are **equipollent.** If one of these classes is a set, then so is the other, by the Replacement-Scheme (¶3.7.4). If $a \approx b$, then we may write also

$$\operatorname{card}(a) = \operatorname{card}(b). \tag{4.1}$$

Restricted to sets, equipollence is an equivalence-relation. (That is, the class $\{(x, y) \colon x \approx y\}$ is an equivalence-relation.) If we understand $\operatorname{card}(a)$ as the equivalence-class $\{x \colon x \approx a\}$, then (4.1) is indeed a consequence of $a \approx b$. (See Exercise 1.) However, we shall ultimately (¶7.2.1) find a definition of $\operatorname{card}(a)$ as a *set* that is equipollent with a. Meanwhile, we may write

$$\boldsymbol{C} \not\approx \boldsymbol{D}$$

if $\boldsymbol{C}$ and $\boldsymbol{D}$ are not equipollent.

4.1.2. A class $\boldsymbol{D}$ is **larger than** or **bigger than,** or **has greater cardinality than,** a class $\boldsymbol{C}$, if there is an injection from $\boldsymbol{C}$ into $\boldsymbol{D}$, but no bijection; then $\boldsymbol{C}$ is **smaller than,** or **has lesser cardinality than,** the class $\boldsymbol{D}$; we may write

$$\boldsymbol{C} \prec \boldsymbol{D}.$$

If there is an injection from $\boldsymbol{C}$ to $\boldsymbol{D}$, then we write

$$\boldsymbol{C} \preccurlyeq \boldsymbol{D}.$$

Therefore

$$\boldsymbol{C} \preccurlyeq \boldsymbol{D} \Leftrightarrow \boldsymbol{C} \prec \boldsymbol{D} \vee \boldsymbol{C} \approx \boldsymbol{D}.$$

If $\boldsymbol{C} \preccurlyeq \boldsymbol{D}$, and $\boldsymbol{D}$ is a set, then so is $\boldsymbol{C}$, by Lemma 3.7.5. If $a \preccurlyeq b$, then we may write also

$$\operatorname{card}(a) \leqslant \operatorname{card}(b).$$

On sets, the relations $\preccurlyeq$ and $\prec$ are transitive; the former is reflexive, but the latter is irreflexive (Exercise 2). Loosely speaking, by the following theorem, both relations are anti-symmetric on cardinalities; so they are a reflexive and a strict ordering, respectively.

4.1.3 Theorem (Schroeder–Bernstein[1])**.** $a \preccurlyeq b \mathbin{\&} b \preccurlyeq a \Rightarrow a \approx b$.

Proof. Suppose $f\colon a \rightarrowtail b$ and $g\colon b \rightarrowtail a$. Let $\boldsymbol{C}$ be the class

$$\{x\colon x \subseteq b \mathbin{\&} (b \smallsetminus f[\,a\,]) \cup (f \circ g)[\,x\,] \subseteq x\}.$$

Then $\boldsymbol{C}$ contains b. In particular, $\boldsymbol{C}$ is non-empty, so (by ¶3.3.3) its intersection is a set d. Then $d \in \boldsymbol{C}$, and

$$(b \smallsetminus f[\,a\,]) \cup (f \circ g)[\,d\,] = d \tag{4.2}$$

(Exercise 3). Hence

$$f[\,a\,] \smallsetminus (f \circ g)[\,d\,] = b \smallsetminus d$$

(Exercise 4). This means

$$f[\,a \smallsetminus g[\,d\,]\,] = f[\,a\,] \smallsetminus (f \circ g)[\,d\,] = b \smallsetminus d = b \smallsetminus g^{-1}[\,g[\,d\,]\,].$$

Now define h from a to b by

$$h(x) = \begin{cases} f(x), & \text{if } x \in a \smallsetminus g[\,d\,]; \\ g^{-1}(x), & \text{if } x \in g[\,d\,]. \end{cases} \tag{4.3}$$

Then $h\colon a \twoheadrightarrow b$. □

4.1.4. Where did this proof of the Schroeder–Bernstein Theorem come from? Perhaps more importantly, if we know that the Theorem is correct, but have forgotten the proof, how can we recover it? To define the bijection h from a to b, all we have to work with are f and g^{-1}. The domain of g^{-1} is only $g[\,b\,]$. So h and f will have to agree on $a \smallsetminus g[\,b\,]$, and perhaps on a larger set. Hence there will be a subset d of b such that h is as in (4.3). Since h will be a bijection onto b, and f and g^{-1} are already injections, we need to ensure

(i) $f[\,a \smallsetminus g[\,d\,]\,] \cap g^{-1}[\,g[\,d\,]\,] = \varnothing$, for injectivity of h;

[1] The theorem is also called the Cantor–Bernstein Theorem, as for example by Levy [10, III.2.8, p. 85], who nonetheless observes that Dedekind gave the first proof in 1887. The proof given here is due to Zermelo, according to Moschovakis [11, ch. 4, exercises 4.26 and 4.27, pp. 50 f.].

(ii) $f[\,a \smallsetminus g[\,d\,]\,] \cup g^{-1}[\,g[\,d\,]\,] = b$, for surjectivity of h onto b.

These conditions can be expressed by the single equation

$$b \smallsetminus g^{-1}[\,g[\,d\,]\,] = f[\,a \smallsetminus g[\,d\,]\,],$$

that is, $b \smallsetminus d = f[\,a\,] \smallsetminus (f \circ g)[\,d\,]$, which can be written as

$$(b \smallsetminus f[\,a\,]) \cup (f \circ g)[\,d\,] = d.$$

Now we can obtain d as an intersection, as in the proof.

4.1.5. The Schroeder–Bernstein Theorem is often proved in the following way. Assuming $f\colon a \rightarrowtail b$ and $g\colon b \rightarrowtail a$, we make the following definitions:

$$\begin{aligned} a_1 &= g[\,b\,], & b_1 &= f[\,a\,], \\ a_2 &= g[\,b_1\,], & b_2 &= f[\,a_1\,], \\ a_3 &= g[\,b_2\,], & b_3 &= f[\,a_2\,], \\ a_4 &= g[\,b_3\,], & b_4 &= f[\,a_3\,], \end{aligned}$$

and so on. (We are not ready to be precise about what *and so on* means here; this is why the proof above does not follow the present lines.) Then $a \supseteq a_1 \supseteq a_2 \supseteq \cdots$, and $b \supseteq b_1 \supseteq b_2 \supseteq \cdots$. Also, f determines a bijection from $a \smallsetminus a_1$ to $b_1 \smallsetminus b_2$, from $a_2 \smallsetminus a_3$ to $b_3 \smallsetminus b_4$, and so on, while g^{-1} determines a bijection from $a_1 \smallsetminus a_2$ to $b \smallsetminus b_1$, from $a_3 \smallsetminus a_4$ to $b_2 \smallsetminus b_3$, and so on. Then there is a bijection h from a to b that agrees with these, and agrees with f at the elements of a that are not yet accounted for. In fact, this will be the same h found in the proof above.

4.1.6 Theorem (Cantor)**.** *The subset-class of a set is larger than the set itself:*

$$a \prec \mathcal{P}(a).$$

Proof. We have $x \mapsto \{x\}\colon a \rightarrowtail \mathcal{P}(a)$, so $a \preccurlyeq \mathcal{P}(a)$. Suppose $f\colon a \rightarrowtail \mathcal{P}(a)$. Let b be the set $\{x \in a\colon x \notin f(x)\}$. Then

$$c \in b \Rightarrow c \in b \smallsetminus f(c), \tag{4·4}$$

$$c \in a \smallsetminus b \Rightarrow c \in f(c) \smallsetminus b \tag{4·5}$$

(Exercise 6). Thus, if $c \in a$, then $f(c) \neq b$. So $b \notin \operatorname{rng}(f)$. Therefore, there is no bijection from a to $\mathcal{P}(a)$; so $a \prec \mathcal{P}(a)$. □

4.1.7. Note well how the preceding proof requires b to be a set (Exercise 7). If a were a proper class, then the proof would fail. In particular, we cannot yet address the cardinality of $\mathcal{P}(\mathcal{P}(a))$, since we have not established that $\mathcal{P}(a)$ is a set. This is what the following axiom is for.

4.1.8 Axiom (Power-set)**.** *The subset-class of a set is a set: that is,*

$$\exists x\ x = \mathcal{P}(a).$$

4.1.9. As foretold in ¶3.2.2, we may now refer to the subset-class of a set as its **power-set.** We can form chains:

$$a \prec \mathcal{P}(a) \prec \mathcal{P}(\mathcal{P}(a)) \prec \mathcal{P}(\mathcal{P}(\mathcal{P}(a))) \prec \cdots .$$

In particular, letting a be the empty set, we have

$$\varnothing \prec \{\varnothing\} \prec \{\varnothing, \{\varnothing\}\} \prec \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\varnothing, \{\varnothing\}\}\} \prec \cdots ;$$

but this is hardly surprising. Cantor's Theorem becomes remarkable when we have *infinite* sets.

4.2 Ordinary induction and recursion

4.2.1. By the formal definition that we shall use, a class is **infinite** if it is equipollent with a proper sub-class of itself. (Later, in Ch. 7, we shall refer to such sets as **Dedekind-infinite.**) Hence for example the universal class $\mathbf{V}$ is infinite. To see this precisely, first note that, by the Power-Set Axiom (¶4.1.8), we have a function $x \mapsto \mathcal{P}(x)$ from $\mathbf{V}$ to itself. This function is injective, since, if $a \neq b$, then we may assume that $a \smallsetminus b$ is non-empty, so that $a \smallsetminus b \in \mathcal{P}(a) \smallsetminus \mathcal{P}(b)$. Also, every power-set contains $\varnothing$, but $\varnothing$ contains nothing, so $\varnothing$ is not a power-set; thus, $x \mapsto \mathcal{P}(x)$ is not surjective onto $\mathbf{V}$. So $\mathbf{V}$ is infinite, by definition. However, by the Russell Paradox (¶3.1.8), $\mathbf{V}$ has a sub-class that is not a set; so $\mathbf{V}$ itself is not a set, by the Comprehension-Scheme (¶3.3.2). Is there an infinite *set?* It is generally assumed that there is, and we shall make the assumption in ¶6.3.7; but let us first see how far we can go without it.

4.2.2. Suppose $\boldsymbol{C}$ is a class, and $\boldsymbol{F}$ is a function, such that

(i) $\boldsymbol{C} \subseteq \mathrm{dom}(\boldsymbol{F})$;

(ii) $\boldsymbol{F}[\,\boldsymbol{C}\,] \subseteq \boldsymbol{C}$.

Then $\boldsymbol{C}$ can be said to be **closed under** $\boldsymbol{F}$. Suppose also $i \in \boldsymbol{C}$. Then $\boldsymbol{C}$ is not just a class anymore: it has additional 'structure' provided by $\boldsymbol{F}$ and i. We can think of $\boldsymbol{C}$ and $\boldsymbol{F}$ and i together as something called a *structure,* denoted by

$$(\boldsymbol{C}, \boldsymbol{F}, i).$$

Since $\mathbf{V}$ itself, together with $\in$, will another example of a structure—to be denoted $(\mathbf{V}, \in)$—, let us distinguish $(\boldsymbol{C}, \boldsymbol{F}, i)$ as an **iterative structure.**[2] (See § 5.1.) There are two properties that such a structure might have:

[2] I could not think of a better term.

(i) If $\boldsymbol{C}$ is a sub-class of every class that contains i and is closed under $\boldsymbol{F}$, then $(\boldsymbol{C}, \boldsymbol{F}, i)$ is **recursive.**

(ii) If $\boldsymbol{C}$ has no proper sub-class that contains i and is closed under $\boldsymbol{F}$, then $(\boldsymbol{C}, \boldsymbol{F}, i)$ **admits (proof by ordinary) induction.**

If $(\boldsymbol{C}, \boldsymbol{F}, i)$ is recursive, then $\boldsymbol{C}$ can be understood as being **defined recursively** by the rules

(i) $i \in \boldsymbol{C}$;

(ii) $a \in \boldsymbol{C} \Rightarrow \boldsymbol{F}(a) \in \boldsymbol{C}$.

Here, $\boldsymbol{C}$ contains what the rules require it to contain, *and nothing else.* If $(\boldsymbol{C}, \boldsymbol{F}, i)$ admits induction, this means that sub-classes of $\boldsymbol{C}$ can be proved to be equal to $\boldsymbol{C}$ **by (ordinary) induction:** Suppose $\boldsymbol{D} \subseteq \boldsymbol{C}$. If

(i) $i \in \boldsymbol{D}$ (the **base** of the induction) and

(ii) $a \in \boldsymbol{D} \Rightarrow \boldsymbol{F}(a) \in \boldsymbol{D}$ (that is, $\boldsymbol{F}(a) \in \boldsymbol{D}$ follows from the **inductive hypothesis** $a \in \boldsymbol{D}$),

then $\boldsymbol{D} = \boldsymbol{C}$.

4.2.3 Lemma. *An iterative structure is recursive if and only if it admits induction.*

Proof. Let $(\boldsymbol{C}, \boldsymbol{F}, i)$ be an iterative structure.

(i) Suppose $(\boldsymbol{C}, \boldsymbol{F}, i)$ does not admit induction. Then $\boldsymbol{C}$ has a proper sub-class $\boldsymbol{D}$ that contains i and is closed under $\boldsymbol{F}$. Then $\boldsymbol{C}$ is not a sub-class of $\boldsymbol{D}$, so $(\boldsymbol{C}, \boldsymbol{F}, i)$ is not recursive.

(ii) Suppose conversely that $(\boldsymbol{C}, \boldsymbol{F}, i)$ is not recursive. Then there is a class $\boldsymbol{D}$ containing i and closed under $\boldsymbol{F}$ of which $\boldsymbol{C}$ is not a sub-class. Then $\boldsymbol{C} \cap \boldsymbol{D}$ is a proper sub-class of $\boldsymbol{C}$ that contains i and is closed under $\boldsymbol{F}$ (by Theorem 3.7.2). Hence $(\boldsymbol{C}, \boldsymbol{F}, i)$ does not admit induction. □

4.2.4. Suppose the iterative structure $(\boldsymbol{C}, \boldsymbol{F}, i)$ admits induction. Then it can be denoted suggestively by

$$\{i, \boldsymbol{F}(i), \boldsymbol{F}(\boldsymbol{F}(i)), \dots\}.$$

See Fig. 4.1. Suppose $(\boldsymbol{D}, \boldsymbol{G}, j)$ is another iterative structure (not necessarily admitting induction.) Then there is at most one function $\boldsymbol{H}$ from $\boldsymbol{C}$ to $\boldsymbol{D}$ such that

(i) $\boldsymbol{H}(i) = j$;

(ii) $a \in \boldsymbol{C} \Rightarrow \boldsymbol{H}(\boldsymbol{F}(a)) = \boldsymbol{G}(\boldsymbol{H}(a))$, that is, $\boldsymbol{H} \circ \boldsymbol{F} = \boldsymbol{G} \circ \boldsymbol{H}$.

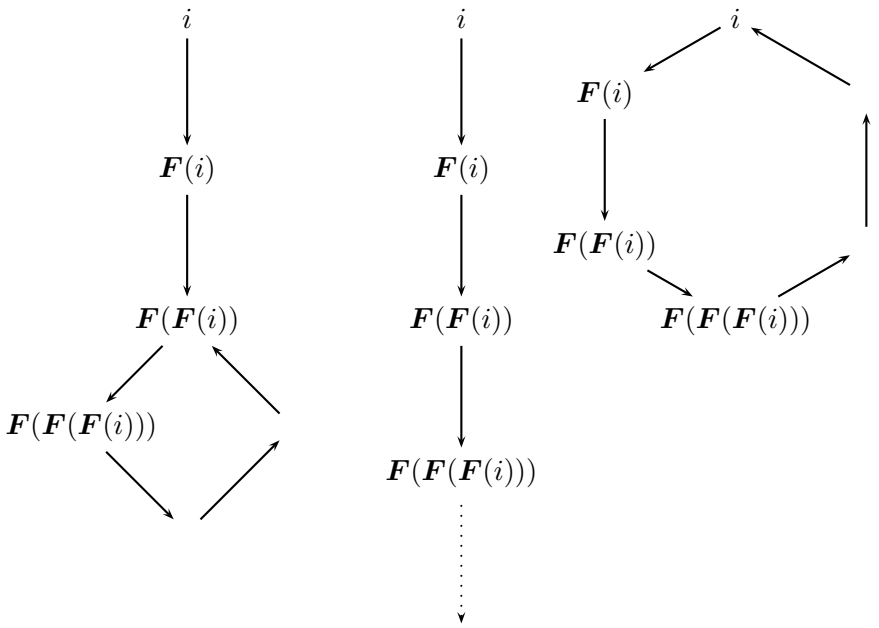

Figure 4.1: Some iterative structures admitting induction

Indeed, suppose $\boldsymbol{H}_0$ and $\boldsymbol{H}_1$ are two such functions. Let

$$\boldsymbol{C}_1 = \{x\colon x \in \boldsymbol{C} \mathbin{\&} \boldsymbol{H}_0(x) = \boldsymbol{H}_1(x)\}.$$

Since $\boldsymbol{H}_0(i) = j = \boldsymbol{H}_1(i)$, we have $i \in \boldsymbol{C}_1$. Suppose $a \in \boldsymbol{C}_1$, so that $\boldsymbol{H}_0(a) = \boldsymbol{H}_1(a)$. Then

$$\boldsymbol{H}_0(\boldsymbol{F}(a)) = \boldsymbol{G}(\boldsymbol{H}_0(a)) = \boldsymbol{G}(\boldsymbol{H}_1(a)) = \boldsymbol{H}_1(\boldsymbol{F}(a)),$$

so $\boldsymbol{F}(a) \in \boldsymbol{C}_1$. Therefore, by induction, $\boldsymbol{C}_1 = \boldsymbol{C}$. So there is at most one function $\boldsymbol{H}$; if $\boldsymbol{H}$ does exist at all, then it is said to be defined **recursively** by the two rules above. However, in some cases, $\boldsymbol{H}$ does not exist (Exercise 8).

4.3 Countably infinite classes

4.3.1. Suppose $\boldsymbol{C}$ is infinite. This means:

(i) there is an element i of $\boldsymbol{C}$, and

(ii) there is a function $\boldsymbol{F}$ from $\boldsymbol{C}$ under which $\boldsymbol{C}$ is closed, such that

(iii) $i \notin \boldsymbol{F}[\,\boldsymbol{C}\,]$, and

(iv) $\boldsymbol{F} \restriction \boldsymbol{C}$ is injective.

(In particular, $(\boldsymbol{C}, \boldsymbol{F}, i)$ is an iterative structure.) Suppose also

(v) $(\boldsymbol{C}, \boldsymbol{F}, i)$ admits induction.

Then $\boldsymbol{C}$ is called **countably infinite.** I propose to refer to $(\boldsymbol{C}, \boldsymbol{F}, i)$ as an **arithmetic structure.** The five numbered conditions here are sometimes referred to as the **Peano axioms.**

4.3.2 Lemma. *Every infinite set has a countably infinite subset.*

Proof. Say a is infinite. Then there is an element i of a and an injection f from a into $a \setminus \{i\}$. Let $\boldsymbol{C}$ be the class

$$\{x\colon x \subseteq a \mathbin{\&} i \in x \mathbin{\&} f[\,x\,] \subseteq x\};$$

this is the class of subsets of a that contain i and are closed under f. Then $\bigcap \boldsymbol{C}$ also belongs to $\boldsymbol{C}$ (Exercise 9) and is a set a_0. Then (a_0, f, i) is recursive, so a_0 is countably infinite. □

4.3.3. If we have two sets a and b, then surely we can form a set that *includes* a and *contains* b; that is, surely the class $a \cup \{b\}$ is a set. The Pairing Axiom is a special case of this observation; the full Union Axiom (¶6.2.3) will be a generalization.

4.3.4 Axiom (Weak Union)**.** *The union of a set and a singleton is a set:*

$$\exists x\ x = a \cup \{b\}.$$

4.3.5 Theorem (Recursion)**.** *Suppose $(\boldsymbol{C}, \boldsymbol{F}, i)$ is an arithmetic structure, and $(\boldsymbol{D}, \boldsymbol{G}, j)$ is an iterative structure. Then there is (uniquely, by ¶4.2.4) a function $\boldsymbol{H}$ from $\boldsymbol{C}$ into $\boldsymbol{D}$ defined recursively by*

(i) $\boldsymbol{H}(i) = j$*;*

(ii) $a \in \boldsymbol{C} \Rightarrow \boldsymbol{H}(\boldsymbol{F}(a)) = \boldsymbol{G}(\boldsymbol{H}(a))$*, that is,* $\boldsymbol{H} \circ \boldsymbol{F} = \boldsymbol{G} \circ \boldsymbol{H}$*.*

Proof. Let $\boldsymbol{E}$ be the class

$$\{x\colon \forall y\ (y \in x \Rightarrow y = (i, j) \vee \\ \vee \exists u\ \exists v\ ((u, v) \in x \cap (\boldsymbol{C} \times \boldsymbol{D}) \mathbin{\&} (\boldsymbol{F}(u), \boldsymbol{G}(v)) = y))\},$$

which is a sub-class of $\mathcal{P}(\boldsymbol{C} \times \boldsymbol{D})$. Let $\boldsymbol{H} = \bigcup \boldsymbol{E}$. Then:

(i) $\{(i, j)\} \in \boldsymbol{E}$, so $i\ \boldsymbol{H}\ j$;

(ii) if $a\ \boldsymbol{H}\ b$, then (a, b) is an element of a member c of $\boldsymbol{E}$; but then $c \cup \{(\boldsymbol{F}(a), \boldsymbol{G}(b))\} \in \boldsymbol{E}$; so $\boldsymbol{F}(a)\ \boldsymbol{H}\ \boldsymbol{G}(b)$.

By induction then, $\mathrm{dom}(\boldsymbol{H}) = \boldsymbol{C}$. We have not yet used the particular properties of $\boldsymbol{F}$; but now we need them to prove that $\boldsymbol{H}$ is functional:

(i) If $i \boldsymbol{H} k$, then (i, k) belongs to an element of $\boldsymbol{E}$; since $i \notin \mathrm{rng}(\boldsymbol{F})$, this means, by the definition of $\boldsymbol{E}$, that $k = j$.

(ii) Suppose b is unique such that $a \boldsymbol{H} b$. We have shown $\boldsymbol{F}(a) \boldsymbol{H} \boldsymbol{G}(b)$. Suppose also $\boldsymbol{F}(a) \boldsymbol{H} c$. Then (by definition of $\boldsymbol{E}$ and $\boldsymbol{H}$, since $i \notin \mathrm{rng}(\boldsymbol{F})$) we have $a_1 \boldsymbol{H} b_1$ for some a_1 and b_1 such that $\boldsymbol{F}(a_1) = \boldsymbol{F}(a)$ and $\boldsymbol{G}(b_1) = c$. Since $\boldsymbol{F}$ is injective, this means $a_1 = a$; so $b_1 = b$, by the uniqueness of b; hence $c = \boldsymbol{G}(b)$.

By induction then, $\boldsymbol{H}$ is functional. Since we have $a \boldsymbol{H} c \Rightarrow \boldsymbol{F}(a) \boldsymbol{H} \boldsymbol{G}(c)$, we can now write this as $\boldsymbol{H}(a) = c \Rightarrow \boldsymbol{H}(\boldsymbol{F}(a)) = \boldsymbol{G}(c)$. Equivalently, $a \in \boldsymbol{C} \Rightarrow \boldsymbol{H}(\boldsymbol{F}(a)) = \boldsymbol{G}(\boldsymbol{H}(a))$. Therefore $\boldsymbol{H}$ is as desired. □

4.3.6 Theorem. *All countably infinite classes are equipollent.*

Proof. Suppose $(\boldsymbol{C}, \boldsymbol{F}, i)$ and $(\boldsymbol{D}, \boldsymbol{G}, j)$ are arithmetic structures. By the Recursion Theorem, there are functions $\boldsymbol{H}$ from $\boldsymbol{C}$ to $\boldsymbol{D}$, and $\boldsymbol{K}$ from $\boldsymbol{D}$ to $\boldsymbol{C}$, such that

(i) $\boldsymbol{H}(i) = j$ and $\boldsymbol{K}(j) = i$;

(ii) $\boldsymbol{H} \circ \boldsymbol{F} = \boldsymbol{G} \circ \boldsymbol{H}$ and $\boldsymbol{K} \circ \boldsymbol{G} = \boldsymbol{F} \circ \boldsymbol{K}$.

Then also

(i) $(\boldsymbol{K} \circ \boldsymbol{H})(i) = i$;

(ii) $(\boldsymbol{K} \circ \boldsymbol{H}) \circ \boldsymbol{F} = \boldsymbol{F} \circ (\boldsymbol{K} \circ \boldsymbol{H})$ since, by (3.32), we can compute $(\boldsymbol{K} \circ \boldsymbol{H}) \circ \boldsymbol{F} = \boldsymbol{K} \circ (\boldsymbol{H} \circ \boldsymbol{F}) = \boldsymbol{K} \circ (\boldsymbol{G} \circ \boldsymbol{H}) = (\boldsymbol{K} \circ \boldsymbol{G}) \circ \boldsymbol{H} = (\boldsymbol{F} \circ \boldsymbol{K}) \circ \boldsymbol{H} = \boldsymbol{F} \circ (\boldsymbol{K} \circ \boldsymbol{H})$.

Thus $\boldsymbol{K} \circ \boldsymbol{H}$ is a recursively defined function from $\boldsymbol{C}$ to itself. But $\mathrm{id}_{\boldsymbol{C}}$ satisfies the same definition. Therefore (¶4.2.4) $\boldsymbol{K} \circ \boldsymbol{H} = \mathrm{id}_{\boldsymbol{C}}$. Likewise $\boldsymbol{H} \circ \boldsymbol{K} = \mathrm{id}_{\boldsymbol{D}}$. Therefore, by Theorem 3.6.8, $\boldsymbol{H}$ is a bijection from $\boldsymbol{C}$ to $\boldsymbol{D}$. □

4.4 Infinite sets

4.4.1. We have shown (¶4.2.1) that there are infinite classes. We have defined (¶4.2.4) countably infinite classes, and we have shown (¶4.3.6) that all of them have the same cardinality. We have shown (¶4.3.2) that, if there is an infinite set, then there is a countably infinite set. We have *not* shown that a countably infinite *class* exists, much less a set. If we believe that there is an infinite set, then we can *postulate* its existence. Indeed, suppose $(\boldsymbol{C}, \boldsymbol{F}, i)$ is an iterative structure, and $\boldsymbol{F}$ is injective on $\boldsymbol{C}$, and $i \notin \boldsymbol{F}[\boldsymbol{C}]$. For example, the structure might be $(\mathbf{V}, x \mapsto \mathcal{P}(x), \varnothing)$. Let $\boldsymbol{D}$ be the class

$$\{x \colon x \subseteq \boldsymbol{C} \mathbin{\&} i \in \boldsymbol{C} \mathbin{\&} \boldsymbol{F}[x] \subseteq x\}.$$

(Compare with the proof of Lemma 4.3.2.) All elements of $\boldsymbol{D}$ are infinite sets. If there *are* elements, then $\bigcap \boldsymbol{D}$ is a countably infinite set. If $\boldsymbol{D}$ is empty, then $\bigcap \boldsymbol{D} = \mathbf{V}$ (Exercise 2), which is not a set (¶4.2.1). So, by postulating that $\bigcap \boldsymbol{D}$ is a set, we ensure that $\boldsymbol{D}$ is not empty, so that there are infinite sets. But this seems like cheating, since, before making such a postulate, we cannot exhibit an element of $\boldsymbol{C}$. Can we obtain a countably infinite *class* without knowing whether there are any infinite sets?

4.4.2. We can try to define a countably infinite class as in the proof of the Recursion Theorem. Suppose $(\boldsymbol{C}, \boldsymbol{F}, i)$ is an iterative structure, and let

$$\boldsymbol{D} = \{x\colon \forall y\, (y \in x \Rightarrow y = i \vee \exists z\, (z \in x \cap \boldsymbol{C}\ \&\ \boldsymbol{F}(z) = y))\}. \tag{4.6}$$

Then $\boldsymbol{D} \subseteq \mathcal{P}(\boldsymbol{C})$, and $\{i\} \in \boldsymbol{D}$; also, if $b \in \boldsymbol{D}$, and $a \in b$, then $b \cup \{\boldsymbol{F}(a)\} \in \boldsymbol{D}$. Thus, $i \in \bigcup \boldsymbol{D}$; and if $a \in \bigcup \boldsymbol{D}$, then $\boldsymbol{F}(a) \in \bigcup \boldsymbol{D}$. So $(\bigcup \boldsymbol{D}, \boldsymbol{F}, i)$ is an iterative structure. However, it might not be arithmetic, even if $\boldsymbol{F}\colon \boldsymbol{C} \rightarrowtail \boldsymbol{C} \smallsetminus \{i\}$. For example, suppose $\boldsymbol{F}$ is the injective function $x \mapsto \{x\}$ on $\mathbf{V}$, and i is $\varnothing$. If there is a set a that is equal to the set $\{a\}$, then $\boldsymbol{F}(a) = a$, so that $\{a\} \in \boldsymbol{D}$, and then $a \in \bigcup \boldsymbol{D}$. So $\bigcup \boldsymbol{D}$ properly includes $\bigcup \boldsymbol{D} \smallsetminus \{a\}$, although the latter class still contains i and is closed under $\boldsymbol{F}$.

4.4.3 Theorem. *Suppose $(\boldsymbol{C}, \boldsymbol{F}, i)$ is an iterative structure, and let $\boldsymbol{D}$ be as in (4.6). If $\boldsymbol{C}$ is well-ordered by a relation $<$ such that*

$$a \in \boldsymbol{C} \Rightarrow a < \boldsymbol{F}(a), \tag{4.7}$$

then $(\bigcup \boldsymbol{D}, \boldsymbol{F}, i)$ is an arithmetic structure, so $\bigcup \boldsymbol{D}$ is countably infinite.

Proof. We showed in ¶4.4.2 that $(\bigcup \boldsymbol{D}, \boldsymbol{F}, i)$ is an iterative structure. Suppose $\boldsymbol{E}$ is a class of which $\bigcup \boldsymbol{D}$ is *not* a sub-class. Then $\bigcup \boldsymbol{D} \smallsetminus \boldsymbol{E}$ is non-empty. If $\boldsymbol{C}$ is well-ordered, then $\bigcup \boldsymbol{D} \smallsetminus \boldsymbol{E}$ has a least element, a. By definition of $\boldsymbol{D}$, either $a = i$, or $a = \boldsymbol{F}(b)$ for some b in $\bigcup \boldsymbol{D}$. If $a = i$, then $i \notin \boldsymbol{E}$. If $a = \boldsymbol{F}(b)$, then $b < a$, so $b \in \boldsymbol{E}$, but $\boldsymbol{F}(b) \notin \boldsymbol{E}$, and so $\boldsymbol{E}$ is not closed under $\boldsymbol{F}$. The contrapositive of this conclusion is that $\bigcup \boldsymbol{D}$ is a sub-class of every class that contains i and is closed under $\boldsymbol{F}$. So $(\bigcup \boldsymbol{D}, \boldsymbol{F}, i)$ is recursive, hence it admits induction.

It remains to show $\boldsymbol{F} \restriction \bigcup \boldsymbol{D}\colon \bigcup \boldsymbol{D} \rightarrowtail \bigcup \boldsymbol{D} \smallsetminus \{i\}$. By induction, i is the least element of $\bigcup \boldsymbol{D}$. Indeed, let $\boldsymbol{E} = \{x\colon x \in \bigcup \boldsymbol{D}\ \&\ i \leqslant x\}$. Then $i \in \boldsymbol{E}$, and if $a \in \boldsymbol{E}$, then $i \leqslant a < \boldsymbol{F}(a)$ by (4.7), so $\boldsymbol{F}(a) \in \boldsymbol{E}$. Therefore $\boldsymbol{E} = \bigcup \boldsymbol{D}$. It now follows from (4.7) that $i \notin \boldsymbol{F}[\bigcup \boldsymbol{D}]$.

Finally, $\boldsymbol{F}$ is injective on $\bigcup \boldsymbol{D}$. Indeed, suppose if possible that $\bigcup \boldsymbol{D}$ has elements a and b such that $a < b$, but $\boldsymbol{F}(a) = \boldsymbol{F}(b)$. Then

$$a < b < \boldsymbol{F}(a). \tag{4.8}$$

In particular, $b \neq i$, so $b = \boldsymbol{F}(c)$ for some c in $\bigcup \boldsymbol{D}$. Then $c < b$. We can let a be least such that b exists as in (4.8); then we can let b be least. In particular,

by the minimality of b, since $c < b$, we have $c \leqslant a$. But $\boldsymbol{F}(c) = b \neq \boldsymbol{F}(a)$, so $c \neq a$. Hence $c < a$. By minimality of a, we then have $\boldsymbol{F}(c) \leqslant a$, so $b \leqslant a$, contrary to (4.8). Therefore $\boldsymbol{F}$ is injective on $\bigcup \boldsymbol{D}$, and so $(\bigcup \boldsymbol{D}, \boldsymbol{F}, i)$ is an arithmetic structure. □

4.4.4. We shall show in the next section that the conditions of Theorem 4.4.3 are met when $\boldsymbol{C}$ is a certain class called **ON**, and $\boldsymbol{F}$ is $x \mapsto x \cup \{x\}$, and $i = \varnothing$. How might this be discovered? The only strict ordering that we know is $\subset$. Also, to obtain a set that properly includes a set a, the easiest thing to do is to take $a \cup \{a\}$. (However, this does not work if $a = \{a\}$.) So we should try letting $\boldsymbol{F}$ be $x \mapsto x \cup \{x\}$. The simplest set is $\varnothing$, so we should try letting i be this. Also, the class $\boldsymbol{D}$ given by (4.6) contains $\varnothing$, and $\{i\}$, and $\{i, \boldsymbol{F}(i)\}$, and $\{i, \boldsymbol{F}(i), \boldsymbol{F}(\boldsymbol{F}(i))\}$, and so forth. Once we do obtain $\bigcap \boldsymbol{D}$ as $\{i, \boldsymbol{F}(i), \boldsymbol{F}(\boldsymbol{F}(i)), \dots\}$, then we might be able to define a function $\boldsymbol{H}$ from $\bigcap \boldsymbol{D}$ into $\boldsymbol{D}$ so that

$$\begin{aligned} \boldsymbol{H}(i) &= \varnothing, \\ \boldsymbol{H}(\boldsymbol{F}(i)) &= \{i\}, \\ \boldsymbol{H}(\boldsymbol{F}(\boldsymbol{F}(i))) &= \{i, \boldsymbol{F}(i)\}, \\ \boldsymbol{H}(\boldsymbol{F}(\boldsymbol{F}(\boldsymbol{F}(i)))) &= \{i, \boldsymbol{F}(i), \boldsymbol{F}(\boldsymbol{F}(i))\}, \dots \end{aligned}$$

We should be able to obtain this by the Recursion Theorem if there is a function $\boldsymbol{G}$ on $\boldsymbol{D}$ such that

$$\begin{aligned} \boldsymbol{G}(\varnothing) &= \{i\}, \\ \boldsymbol{G}(\{i\}) &= \{i, \boldsymbol{F}(i)\}, \\ \boldsymbol{G}(\{i, \boldsymbol{F}(i)\}) &= \{i, \boldsymbol{F}(i), \boldsymbol{F}(\boldsymbol{F}(i))\}, \dots \end{aligned}$$

We seem to have this if $i = \varnothing$ and $\boldsymbol{G}$ is $x \mapsto x \cup \{x\}$. Since $a \in a \cup \{a\}$, the fundamental relation of containment will be involved in the construction.

4.5 Ordinals

4.5.1. A class $\boldsymbol{C}$ is called **transitive** if it *includes* each of its elements, that is, $a \in \boldsymbol{C} \Rightarrow a \subseteq \boldsymbol{C}$, or equivalently

$$b \in a \;\&\; a \in \boldsymbol{C} \Rightarrow b \in \boldsymbol{C}.$$

Compare with (3.28); however, transitivity of *classes* must be distinguished from transitivity of *relations* on classes (¶3.5.2). We shall be particularly interested in the relation of containment, $\in$:

(i) $\Big\{\varnothing, \{\varnothing\}, \big\{\{\varnothing\}\big\}\Big\}$ is a transitive set, but containment is not transitive on this set.

(ii) On the set $\Big\{\{\varnothing\},\{\{\varnothing\}\},\Big\{\{\varnothing\},\{\{\varnothing\}\}\Big\}\Big\}$, containment is transitive, but the set itself is not transitive (it does not include its member $\{\varnothing\}$).

For every set a, let $a\cup\{a\}$ be called the **successor** of a; this can be denoted by

$$a'.$$

Then (Exercise 11)

$$a' = \bigcap\{x\colon a\in x \mathbin{\&} a\subseteq x\}. \tag{4.9}$$

4.5.2 Lemma. *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 $b\in a$, then $\{b\}\subseteq a$, but also $b\subseteq a$ by transitivity of a, so that $b'\subseteq a$. If $c\in a'$, then either $c=a$ or $c\in a$; in either case, $c\subseteq a'$; thus a' is transitive. □

4.5.3. A set is an **ordinal (number)** if it is transitive and also well-ordered by containment. Trivially, $\varnothing$ is an ordinal. We shall generally denote ordinals by letters from the beginning of the Greek alphabet: α, β, γ, δ, and so on. The class of ordinals is denoted by **ON**.

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

(i) $b\in\alpha$;

(ii) $b\subset\alpha$, *and* b *is transitive;*

(iii) b *is a proper initial segment of* α *with respect to containment.*

Hence every element of α is an ordinal. Moreover, α' is also an ordinal.

Proof. First, there are three implications to prove:

(i) Suppose $b\in\alpha$. Then $b\neq\alpha$, since α is strictly well-ordered by containment (see Exercise 12); and $b\subseteq\alpha$, by transitivity of α. Hence $b\subset\alpha$. Say $c\in b$. Then $c\in\alpha$. Say also $d\in c$. Then $d\in\alpha$ by transitivity of α. Thus, $d\in c \mathbin{\&} c\in b$, and b, c, and d are in α; so $d\in b$ by transitivity of $\in$ on α. Thus $c\subseteq b$. Therefore b is transitive.

(ii) Suppose $b\subset\alpha$, and b is transitive. Say $c\in b$ and $d\in c$. Then $d\in b$ by transitivity of b. Hence b is a proper initial segment of α.

(iii) Suppose b is a proper initial segment of α with respect to $\in$. Let c be the $\in$-least element of $\alpha\smallsetminus b$. In particular, $c\notin b$; but $c\in\alpha$, so $c\subseteq\alpha$. We shall show $c=b$. We first show $b\subseteq c$. Say $d\in\alpha$, but $d\notin c$. Then $d=c$ or $c\in d$, since α is totally ordered by $\in$. If $d=c$, then $d\in\alpha\smallsetminus b$,

so $d \notin b$. If $c \in d$, then $d \notin b$, since b is an initial segment of α, and $c \notin b$. In either case, $d \notin b$. Thus, $b \subseteq c$. Conversely, if $d \in c$, then $d \in \alpha$; so $d \in b$, since c is $\in$-least in $\alpha \setminus b$. Hence $c \subseteq b$, so $b = c$, and $b \in \alpha$.

Every subset of a well-ordered set is well-ordered; so a transitive subset of α is an ordinal; but we have just shown that the elements of α are transitive subsets. So the elements of α are ordinals.

To see that α' is an ordinal, note first that elements of α are *proper* subsets and are hence distinct from α. So $\alpha \notin \alpha$. Hence, since $\in$ is irreflexive on α, it remains so on $\alpha \cup \{\alpha\}$, that is, α'. Also, if $b \in \alpha$, then $b \subseteq \alpha$, so $\alpha \notin b$ (since $\alpha \notin \alpha$). So, since $\in$ is anti-symmetric on α, it remains so on α'. Since α is transitive, and $\in$ is transitive on α, we have that $\in$ is transitive on α'. So $\in$ is a strict total ordering of α', and α is $\in$-greatest in α'. The $\in$-least element of a non-empty subset b of α' is the least element of $b \cap \alpha$, if this set is non-empty; otherwise, it is α itself. In particular, α' is well-ordered by $\in$. So α' is an ordinal. □

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

Proof. Since $\alpha \notin \beta$, we know from Lemma 4.5.4 that α is not a proper subset of β. Since also we assume $\alpha \neq \beta$, we have that α is not a subset of β, so $\alpha \setminus \beta$ is non-empty. Let γ be its $\in$-least element. Then γ is an initial segment of α, again by Lemma 4.5.4; so its elements are in β, and $\gamma \subseteq \beta$. But $\gamma \notin \beta$. Since γ is an ordinal, γ must not be a proper subset of β. Hence $\gamma = \beta$. □

4.5.6 Theorem. ON *is transitive and well-ordered by containment.*

Proof. Exercise 13. □

4.5.7 Corollary (Burali-Forti Paradox)**. ON** *is not a set, so it is not an ordinal.*

Proof. Exercise 14. □

4.5.8. By Theorem 4.4.3, there is a class $\{\varnothing, \varnothing', \varnothing'', \dots\}$; it is the class of **natural numbers,** and it can be denoted by

$$\omega.$$

There are **numerals** for denoting natural numbers:

$$\begin{aligned} 0 &= \varnothing, & 5 &= 4', \\ 1 &= 0', & 6 &= 5', \\ 2 &= 1', & 7 &= 6', \\ 3 &= 2', & 8 &= 7', \\ 4 &= 3', & 9 &= 8'. \end{aligned}$$

If ω is a proper class, then it is just **ON** (Exercise 15). We shall ultimately (¶6.3.7) say that ω *is* a set; but we need not do this for a while.

Exercises

(1) Show that equipollence of sets is an equivalence-relation, and hence that it is justifiable to define the cardinality of a set a as $\{x\colon x \approx a\}$ (as in ¶4.1.1).

(2) Show that $\preccurlyeq$ is transitive and reflexive, but $\prec$ is transitive and irreflexive.

(3) Establish (4.2) in the proof of the Schroeder–Bernstein Theorem.

(4) Assuming $\boldsymbol{A} \subseteq \boldsymbol{B} \subseteq \boldsymbol{D}$ and $\boldsymbol{C} \subseteq \boldsymbol{D}$, show that $\boldsymbol{B} \setminus \boldsymbol{A} = \boldsymbol{D} \setminus \boldsymbol{C} \Leftrightarrow \boldsymbol{C} = (\boldsymbol{D} \setminus \boldsymbol{B}) \cup \boldsymbol{A}$. Show also that the condition $\boldsymbol{A} \subseteq \boldsymbol{B}$ is necessary.

(5) Is the Schroeder–Bernstein Theorem true for classes in general?

(6) Establish the implications (4.4) and (4.5) in the proof of Cantor's Theorem.

(7) Why does the proof of Cantor's Theorem (¶4.1.6) require $\{x \in a\colon x \notin f(x)\}$ to be a set?

(8) Find an example of a structure $(\boldsymbol{C}, \boldsymbol{F}, i)$ admitting induction and an iterative structure $(\boldsymbol{D}, \boldsymbol{G}, j)$ such that there is *no* function $\boldsymbol{H}$ from $\boldsymbol{C}$ to $\boldsymbol{D}$ such that $\boldsymbol{H}(i) = j$ and $\boldsymbol{H} \circ \boldsymbol{F} = \boldsymbol{G} \circ \boldsymbol{H}$.

(9) Prove the claim in the proof of Lemma 4.3.2.

(10) How is the Weak Union Axiom (¶4.3.4) used in the proof of the Recursion Theorem (¶4.3.5)?

(11) Prove (4.9).

(12) Show that, if $a = \{a\}$, then a is not an ordinal.

(13) Prove Theorem 4.5.6.

(14) Prove the Burali-Forti Paradox (¶4.5.7).

(15) Show that, if ω is a proper class, then $\omega = \mathbf{ON}$.

Chapter 5

The natural numbers

5.1 Structures

5.1.1. The class of functions from a set a into a class $\boldsymbol{C}$ can be denoted

$$^{a}\boldsymbol{C}.$$

We shall be interested especially in the classes $^{n}\boldsymbol{C}$, where $n \in \omega$ and $\boldsymbol{C} \neq \varnothing$. In this case, we may write

$$\boldsymbol{C}^{n}$$

instead of $^{n}\boldsymbol{C}$, although some cases of this notation will get a different interpretation in ¶7.4.1. Meanwhile have

$$\boldsymbol{C}^{0} = \{0\} = 1, \tag{5.1}$$

as well as

$$\boldsymbol{F} \mapsto \boldsymbol{F}(0) \colon \boldsymbol{C}^{1} \rightarrowtail \boldsymbol{C}, \tag{5.2}$$

$$\boldsymbol{F} \mapsto (\boldsymbol{F}(0), \boldsymbol{F}(1)) \colon \boldsymbol{C}^{2} \rightarrowtail \boldsymbol{C} \times \boldsymbol{C}. \tag{5.3}$$

A notation like $\boldsymbol{C} \times \boldsymbol{C} \times \boldsymbol{C}$ is ambiguous; it could be $(\boldsymbol{C} \times \boldsymbol{C}) \times \boldsymbol{C}$ or $\boldsymbol{C} \times (\boldsymbol{C} \times \boldsymbol{C})$; but the distinction is usually unimportant, since the two classes are in bijection with each other and with $\boldsymbol{C}^{3}$ (Exercise 1).

5.1.2. An arbitrary element of $\boldsymbol{C}^{n}$ can be written as

$$(a_i \colon i \in n),$$

or as $(a_0, \ldots, a_{n-1})$, or just as

$$\vec{a}.$$

If the free variables of a formula φ are the variables x_i, where $i \in n$, then we can write φ as $\varphi(x_0, \ldots, x_{n-1})$ or $\varphi(\vec{x})$. Such a formula is called n-ary, and the result of replacing each free occurrence of x_i with a_i, for each i in

n, is $\varphi(a_0, \dots, a_{n-1})$ or $\varphi(\vec{a})$. In general, a sub-class of $\boldsymbol{C}^n$ can be called an **n-ary relation on $\boldsymbol{C}$**: such a relation is

$$\{\vec{x} \in \boldsymbol{C}^n : \varphi(\vec{x})\}$$

for some n-ary formula φ. By (5.2), a 1-ary relation on $\boldsymbol{C}$ can be considered as a sub-class of $\boldsymbol{C}$; by (5.3), a 2-ary relation can be considered as a binary relation. A function from $\boldsymbol{C}^n$ into $\boldsymbol{C}$ is an **n-ary operation on $\boldsymbol{C}$**. By (5.1) and (5.2), a 0-ary operation can be considered as an element of $\boldsymbol{C}$. A singulary (1-ary) operation $\boldsymbol{F}$ is sometimes written as

$$x \mapsto x^{\boldsymbol{F}}$$

(as in the case of the successor-operation $x \mapsto x'$) instead of $x \mapsto \boldsymbol{F}(x)$; a binary (2-ary) operation $\boldsymbol{G}$ is often written as

$$(x, y) \mapsto x \, \boldsymbol{G} \, y$$

instead of $(x, y) \mapsto \boldsymbol{G}(x, y)$.

5.1.3. A **structure** is a non-empty class equipped with some operations and relations. The class is then the **universe** of the structure. Examples include:

(i) the iterative structures of ¶4.2.2, such as $(\omega, x \mapsto x', 0)$;

(ii) $(\boldsymbol{C}, \boldsymbol{R})$, where $\boldsymbol{R}$ is an ordering of $\boldsymbol{C}$: this structure is an **ordered class** or just an **order;**

(iii) $(\mathbf{V}, \in)$;

(iv) $(\mathcal{P}(\boldsymbol{C}), \cap, \cup, x \mapsto x^{\mathrm{c}}, \varnothing, \boldsymbol{C})$.

We can understand a structure formally as a class (Exercise 2), but it is not necessary to do so. When we speak in general terms, we may denote a structure by the Fraktur form of the letter denoting its universe. So the structure $(\boldsymbol{C}, \dots)$ may be denoted by $\mathfrak{C}$, and $(a, \dots)$ by $\mathfrak{a}$. The Fraktur letters can be written by hand in the Sütterlin script (Fig. 5.1).[1]

5.1.4. A structure $\mathfrak{C}$ has a **signature,** which comprises:

(i) an **n-ary function-symbol** for each n-ary operation of $\mathfrak{C}$;

(ii) an **n-ary predicate** for each n-ary relation of $\mathfrak{C}$.

[1] If you learn to write these letters, you will be among the few people who know how to. I took the figure from `www.suetterlinschrift.de/Englisch/Sutterlin.htm`.

Figure 5.1: The Sütterlin script

If s is one of these symbols, then the corresponding operation or relation can be denoted

$$s^{\mathfrak{C}}.$$

We may need this notation when another structure $\mathfrak{D}$ has the same signature as $\mathfrak{C}$. Then $s^{\mathfrak{C}}$ and $s^{\mathfrak{D}}$ are different operations, although they are represented by the same symbol s.

5.1.5. Suppose $\mathfrak{C}$ and $\mathfrak{D}$ have a common signature. If also there is a function $\boldsymbol{H}$ from $\boldsymbol{C}$ to $\boldsymbol{D}$ such that

(i) $\boldsymbol{H}(F^{\mathfrak{C}}(a_i\colon i \in n)) = F^{\mathfrak{D}}(\boldsymbol{H}(a_i) : i \in n)$ when $n \in \omega$ and F is an n-ary function-symbol of the signature, and

(ii) $(a_i\colon i \in n) \in R^{\mathfrak{C}} \cap \boldsymbol{C}^n \Rightarrow (\boldsymbol{H}(a_i)\colon i \in n) \in R^{\mathfrak{D}}$ when $n \in \omega$ and R is an n-ary predicate of the signature,

then $\boldsymbol{H}$ is a **homomorphism** from $\mathfrak{C}$ to $\mathfrak{D}$, and we may write

$$\boldsymbol{H}\colon \mathfrak{C} \to \mathfrak{D}.$$

We can understand all iterative structures as having the same signature. Then the Recursion Theorem is that there is a unique homomorphism from an (arbitrary) arithmetic structure into an iterative structure. By ¶4.2.4, there is at most one homomorphism from any structure admitting induction into an iterative structure.

5.1.6. Again suppose $\mathfrak{C}$ and $\mathfrak{D}$ have a common signature. A function $\boldsymbol{H}$ from $\boldsymbol{C}$ into $\boldsymbol{D}$ is an **embedding** of $\mathfrak{C}$ in $\mathfrak{D}$ if $\boldsymbol{H}$ is an injective homomorphism and also

$$(\boldsymbol{H}(a_i)\colon i \in n) \in R^{\mathfrak{D}} \Rightarrow (a_i\colon i \in n) \in R^{\mathfrak{C}}$$

when R is an n-ary predicate of the signature. If $\boldsymbol{C} \subseteq \boldsymbol{D}$, and $\mathrm{id}_{\boldsymbol{C}}$ is an embedding, then $\mathfrak{C}$ is a **sub-structure** of $\boldsymbol{D}$, and we may write

$$\mathfrak{C} \subseteq \mathfrak{D}.$$

Suppose in particular that $\mathfrak{D}$ is an iterative structure. Then $\mathfrak{D}$ admits induction if and only if, whenever $\mathfrak{C}$ is an iterative structure, and $\boldsymbol{H}$ is an embedding of $\mathfrak{C}$ in $\mathfrak{D}$, then $\boldsymbol{H}$ is surjective onto $\boldsymbol{D}$ (Exercise 3).

5.1.7. If $\boldsymbol{H}$ is a bijection from $\boldsymbol{C}$ to $\boldsymbol{D}$, and $\boldsymbol{H}$ is a homomorphism from $\mathfrak{C}$ to $\mathfrak{D}$, and $\boldsymbol{H}^{-1}$ is a homomorphism from $\mathfrak{D}$ to $\mathfrak{C}$, then $\boldsymbol{H}$ is an **isomorphism** from $\mathfrak{C}$ to $\mathfrak{D}$, and we may write

$$\boldsymbol{H}\colon \mathfrak{C} \xrightarrow{\cong} \mathfrak{D}$$

or simply

$$\mathfrak{C} \cong \mathfrak{D}.$$

By the proof of Theorem 4.3.6, all arithmetic structures are isomorphic. We know of an arithmetic structure, namely $(\omega, x \mapsto x', \varnothing)$. However, throughout this chapter, let

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

denote an arbitrary arithmetic structure. We shall think of $\mathbb{N}$ as the class of natural numbers, when the understanding of natural numbers as ordinals in the sense of ¶4.5.3 is unimportant. To simplify writing, we may use $\mathbb{N}$ and ω by themselves to refer to structures of which they are universes.

5.2 Addition on structures admitting induction

5.2.1. We shall see that some statements about arithmetic structures are actually true about all structures that admit ordinary induction. In this section and the next, let us denote by

$$(\boldsymbol{A}, \boldsymbol{S}, i)$$

an arbitrary structure admitting induction.

5.2.2 Theorem and Definition. *On $\boldsymbol{A}$, there is a unique binary operation, called **addition** and denoted by*

$$(x, y) \mapsto x + y,$$

such that (for all a and b in $\boldsymbol{A}$)

(i) $a + i = a$,

(ii) $a + \boldsymbol{S}(b) = \boldsymbol{S}(a + b)$.

Proof. We first prove uniqueness. Suppose at least one such operation $+$ exists. For each a in $\boldsymbol{A}$, let the operation $y \mapsto a + y$ be denoted by $\boldsymbol{F}_a$. Then

(i) $\boldsymbol{F}_a(i) = a$,

(ii) $\boldsymbol{F}_a(\boldsymbol{S}(b)) = \boldsymbol{S}(\boldsymbol{F}_a(b))$.

Then $\boldsymbol{F}_a$ is recursively defined, so by ¶4.2.4 it is is unique. Therefore $+$ is unique.

Let $\boldsymbol{G}$ be the homomorphism from $(\mathbb{N}, x \mapsto x^+, 0)$ into $(\boldsymbol{A}, \boldsymbol{S}, i)$ (¶5.1.5). Then let

$$\boldsymbol{R} = \bigcup \Big\{ v \colon \forall u \, \Big(u \in v \Rightarrow \exists y \, \big(y \in \boldsymbol{A} \;\&\; u = (0, i, y, y) \big) \vee \\ \vee \exists x \, \exists y \, \exists z \, \big((x, y, z) \in \mathbb{N} \times \boldsymbol{A} \times \boldsymbol{A} \;\& \\ \& \; (x, \boldsymbol{G}(x), y, z) \in v \;\&\; u = (x^+, \boldsymbol{S}(\boldsymbol{G}(x)), y, \boldsymbol{S}(z)) \big) \Big) \Big\},$$

a sub-class of $\mathbb{N} \times \boldsymbol{A} \times \boldsymbol{A} \times \boldsymbol{A}$. If $d \in \mathbb{N}$, let

$$\boldsymbol{R}_d = \{ (y, z) \colon (d, \boldsymbol{G}(d), y, z) \in \boldsymbol{R} \}.$$

Then $\boldsymbol{R}_0$ is an operation on $\boldsymbol{A}$, and $\boldsymbol{R}_0 = \mathrm{id}_{\boldsymbol{A}}$; also, if $\boldsymbol{R}_d$ is an operation on $\boldsymbol{A}$, then so is $\boldsymbol{R}_{d^+}$, and $\boldsymbol{R}_{d^+} = \boldsymbol{S} \circ \boldsymbol{R}_d$ (Exercise 4). By induction, each $\boldsymbol{R}_d$ is an operation on $\boldsymbol{A}$. We can also show by induction that

(i) $\boldsymbol{R}_d(i) = \boldsymbol{G}(d)$,

(ii) $\boldsymbol{R}_d(\boldsymbol{S}(b)) = \boldsymbol{S}(\boldsymbol{R}_d(b))$.

Indeed, the claims are easily true when $d = 0$, since $\boldsymbol{R}_0 = \mathrm{id}_{\boldsymbol{A}}$. If the claims are true when $d = e$, then, since $\boldsymbol{R}_{e^+} = \boldsymbol{S} \circ \boldsymbol{R}_e$, we have

(i) $\boldsymbol{R}_{e^+}(i) = \boldsymbol{S}(\boldsymbol{R}_e(i)) = \boldsymbol{S}(\boldsymbol{G}(e)) = \boldsymbol{G}(e^+)$,

(ii) $\boldsymbol{R}_{e^+}(\boldsymbol{S}(b)) = \boldsymbol{S}(\boldsymbol{R}_e(\boldsymbol{S}(b))) = \boldsymbol{S}(\boldsymbol{S}(\boldsymbol{R}_e(b))) = \boldsymbol{S}(\boldsymbol{R}_{e^+}(b))$,

so the claims are true when $d = e^+$. This completes the induction.

Finally, by the uniqueness argument above, if $\boldsymbol{G}(d) = \boldsymbol{G}(e)$, then $\boldsymbol{R}_d = \boldsymbol{R}_e$. So let

$$+ = \{ (x, y, z) \colon \exists u \, (u, x, y, z) \in \boldsymbol{R} \}.$$

Then the classes $\{ (y, z) \colon (a, y, z) \in \boldsymbol{R} \}$ are the functions $\boldsymbol{F}_a$ given above, and $+$ is as desired. □

5.2.3. It is sometimes argued that the existence of $+$ follows immediately by induction:

> We define $x + i$ as x, for all x in $\boldsymbol{A}$. If $x + b$ has been defined for all x in $\boldsymbol{A}$, then we let $x + \boldsymbol{S}(b) = \boldsymbol{S}(x + b)$. Therefore, by induction, $x + y$ is defined for all x and y in $\boldsymbol{A}$.

However, induction is not a method of definition; it is a method of proving that one class is equal to another. So the proposed proof here would have to start out in the following fashion:

> Let $\boldsymbol{A}_0$ be the sub-class of $\boldsymbol{A}$ comprising all y for which the operation $x \mapsto x + y$ is defined on $\boldsymbol{A}$ so that $\forall x\ x + i = x$ and $\forall x\ \forall z\ x + \boldsymbol{S}(z) = \boldsymbol{S}(x + z)$.

But this definition of $\boldsymbol{A}_0$ is invalid: the conditions imposed on the singulary operation $x \mapsto x + y$ make no sense here.

5.2.4. If we know that $\boldsymbol{A}$ is a *set,* then we can prove Theorem 5.2.2 as follows: As we showed, for each a in $\boldsymbol{A}$, there is at most one operation f_a on $\boldsymbol{A}$ such that

(i) $f_a(i) = a$,

(ii) $\forall y\ f_a(\boldsymbol{S}(y)) = \boldsymbol{S}(f_a(y))$.

Since the functions f_a will be sets, we can prove by induction that the set of a for which f_a exists as desired is $\boldsymbol{A}$ itself. Then we have a function $x \mapsto f_x$ on $\boldsymbol{A}$, and we can let

$$+ = \bigcup \{\{x\} \times f_x \colon x \in \boldsymbol{A}\}.$$

If we do not know that $\boldsymbol{A}$ is a set, then we have to follow a more roundabout route, as above.

5.2.5 Lemma. *For all a and b in $\boldsymbol{A}$,*

(i) $i + a = a$,

(ii) $\boldsymbol{S}(b) + a = \boldsymbol{S}(b + a)$.

Proof. By definition of $+$, we have $i + i = i$. Suppose $i + b = b$. Then also by definition of $+$, we have $i + \boldsymbol{S}(b) = \boldsymbol{S}(i + b) = \boldsymbol{S}(b)$. By induction, $\forall x\ i + x = x$.

Let $\boldsymbol{A}_0 = \{x \colon \forall y\ \boldsymbol{S}(y) + x = \boldsymbol{S}(y + x)\}$. Since $\boldsymbol{S}(b) + i = \boldsymbol{S}(b) = \boldsymbol{S}(b + i)$, we have $i \in \boldsymbol{A}_0$. Suppose $a \in \boldsymbol{A}_0$, so that $\boldsymbol{S}(b) + a = \boldsymbol{S}(b + a)$. Then $\boldsymbol{S}(b) + \boldsymbol{S}(a) = \boldsymbol{S}(\boldsymbol{S}(b) + a) = \boldsymbol{S}(\boldsymbol{S}(b + a)) = \boldsymbol{S}(b + \boldsymbol{S}(a))$, so $\boldsymbol{S}(a) \in \boldsymbol{A}_0$. □

5.2.6. In proving the last lemma, we cannot establish $\boldsymbol{S}(b) + a = \boldsymbol{S}(b + a)$ by induction if we use the class $\{y \colon \forall x\ \boldsymbol{S}(y) + x = \boldsymbol{S}(y + x)\}$ instead of $\boldsymbol{A}_0$.

5.2.7 Theorem. *For all a, b, and c in $\boldsymbol{A}$,*

(i) $\boldsymbol{S}(a) = a + \boldsymbol{S}(i)$;

(ii) $a + b = b + a$, *that is,* $+$ *is* ***commutative;***

(iii) $(a + b) + c = a + (b + c)$, *that is,* $+$ *is* ***associative;***

(iv) $a + c = b + c \Rightarrow a = b$, *that is,* $+$ *admits* ***cancellation.***

Proof. Exercise 5. □

5.3 Multiplication and exponentiation

5.3.1. We again let $(\boldsymbol{A}, \boldsymbol{S}, i)$ be an arbitrary structure admitting ordinary induction.

5.3.2 Theorem and Definition. *On* $\boldsymbol{A}$*, there is a unique binary operation, called* ***multiplication*** *and denoted by*

$$(x, y) \mapsto x \cdot y$$

or $(x, y) \mapsto xy$*, such that (for all* a *and* b *in* $\boldsymbol{A}$*)*

(i) $a \cdot i = i$,

(ii) $a \cdot \boldsymbol{S}(b) = a \cdot b + a$.

Proof. Exercise 6. □

5.3.3 Lemma. *For all* a *and* b *in* $\boldsymbol{A}$*,*

(i) $i \cdot a = i$,

(ii) $\boldsymbol{S}(b) \cdot a = b \cdot a + a$.

Proof. Exercise 7. □

5.3.4 Theorem. *For all* a*,* b*, and* c *in* $\boldsymbol{A}$*,*

(i) $\boldsymbol{S}(i) \cdot a = a$;

(ii) $a \cdot b = b \cdot a$, *that is, multiplication is* ***commutative;***

(iii) $(a+b)\cdot c = a\cdot c+b\cdot c$, *that is, multiplication* ***distributes*** *over addition;*

(iv) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$, *that is, multiplication is* ***associative.***

(v) $a \cdot c^+ = b \cdot c^+ \Rightarrow a = b$, *that is,* $\cdot$ *admits* ***cancellation.***

Proof. Exercise 8. □

5.3.5 Theorem and Definition. *There is a unique function from* $\boldsymbol{A} \times \mathbb{N}$ *into* $\boldsymbol{A}$*, called* ***exponentiation*** *and denoted by*

$$(x, y) \mapsto x^y,$$

such that, for all a *in* $\boldsymbol{A}$ *and* m *in* $\mathbb{N}$*,*

(i) $a^0 = \boldsymbol{S}(i)$*,*

(ii) $a^{m^+} = a^m \cdot a$*.*

Proof. By the Recursion Theorem (¶4.3.5), for each a in $\boldsymbol{A}$, there is a unique function $\boldsymbol{F}_a$ from $\mathbb{N}$ into $\boldsymbol{A}$ such that

(i) $\boldsymbol{F}_a(0) = \boldsymbol{S}(i)$,

(ii) $\boldsymbol{F}_a(b^+) = \boldsymbol{F}_a(b) \cdot a$.

By the *proof* of the Recursion Theorem, there is a binary formula $\varphi(x, y)$ such that $\boldsymbol{F}_a = \{y\colon \varphi(a, y)\}$. Hence we can define $(x, y) \mapsto x^y$ as $\{(x, y)\colon x \in \boldsymbol{A} \mathbin{\&} \varphi(x, y)\}$. Since the functions $\boldsymbol{F}_a$ are unique, so is$(x, y) \mapsto x^y$. □

5.3.6. An **endomorphism** of a structure is a homomorphism from the structure into itself. The endomorphisms of $\mathfrak{C}$ compose a virtual class (¶3.6.2), which we may denote by

$$\mathrm{End}(\mathfrak{C});$$

this is closed under composition, so we have a (virtual) structure $(\mathrm{End}(\mathfrak{C}), \circ)$.

5.3.7 Theorem. .

(i) $a^{m+n} = a^m \cdot a^n$*, that is,* $x \mapsto a^x$ *is a homomorphism from* $(\mathbb{N}, +)$ *into* $(\boldsymbol{A}, \cdot)$*;*

(ii) $(a \cdot b)^m = a^m \cdot b^m$*, that is,* $x \mapsto x^a$ *is an endomorphism of* $(\boldsymbol{A}, \cdot)$*;*

(iii) $(a^m)^n = a^{m \cdot n}$*, that is,* $x \mapsto (y \mapsto y^x)$ *is a homomorphism from* $(\mathbb{N}, \cdot)$ *into* $(\mathrm{End}(\boldsymbol{A}, \cdot), \circ)$*.*

Proof. Exercise 10. □

5.4 The ordering of the natural numbers

5.4.1. There is a unique isomorphism $\boldsymbol{H}$ from $\mathbb{N}$ to ω. As ω is well-ordered by $\in$, so $\mathbb{N}$ is well-ordered by

$$\{(x, y)\colon \boldsymbol{H}(x) \in \boldsymbol{H}(y)\};$$

we denote this ordering by $<$. The elements of the initial segment

$$\{x \in \mathbb{N}\colon x < a\}$$

are the **predecessors** of a. In particular,

(i) $\{x \in \mathbb{N}\colon x < 0\} = \varnothing$,

(ii) $\{x \in N\colon x < a^+\} = \{x \in \mathbb{N}\colon x < a\} \cup \{a\}$.

This can be understood as a recursive definition of the function $x \mapsto \{y \in \mathbb{N}\colon y < x\}$, though the definition is not quite recursive in the sense of ¶4.2.4. The definition is justified by the following.

5.4.2 Theorem (Recursion with Parameter)**.** *Suppose $i \in \boldsymbol{C}$ and $\boldsymbol{F}\colon \mathbb{N} \times \boldsymbol{C} \to \boldsymbol{C}$. Then there is a unique function $\boldsymbol{G}$ from $\mathbb{N}$ into $\boldsymbol{C}$ such that*

(i) $\boldsymbol{G}(0) = i$,

(ii) $\boldsymbol{G}(a^+) = \boldsymbol{F}(a, \boldsymbol{G}(a))$.

Proof. Let $\boldsymbol{F}_1$ be the function $(x, y) \mapsto (x^+, \boldsymbol{F}(x, y))$ on $\mathbb{N} \times \boldsymbol{C}$. Then $(\mathbb{N} \times \boldsymbol{C}, \boldsymbol{F}_1, (0, i))$ is an iterative structure, so by the Recursion Theorem, there is a unique function $\boldsymbol{H}$ from $\mathbb{N}$ into $\mathbb{N} \times \boldsymbol{C}$ such that

(i) $\boldsymbol{H}(0) = (0, i)$,

(ii) $\boldsymbol{H}(a^+) = \boldsymbol{F}_1(\boldsymbol{H}(a))$.

Let π_0 be the function $(x, y) \mapsto x$, and let π_1 be $(x, y) \mapsto y$. Now define $\boldsymbol{G}$ as $\pi_1 \circ \boldsymbol{H}$, so that $\boldsymbol{G}(0) = i$. By induction, we can prove

$$\pi_0(\boldsymbol{H}(a) = a.$$

Indeed, the claim is true when $a = 0$. Suppose it is true when $a = b$; that is, assume

$$\boldsymbol{H}(b) = (b, \boldsymbol{G}(b)). \tag{5.4}$$

Then

$$\begin{aligned} \boldsymbol{H}(b^+) &= \boldsymbol{F}_1(\boldsymbol{H}(b)) \\ &= \boldsymbol{F}_1(b, \boldsymbol{G}(b)) \\ &= (b^+, \boldsymbol{F}(b, \boldsymbol{G}(b))). \end{aligned}$$

Hence $\pi_0(\boldsymbol{H}(b^+)) = b^+$. This computation establishes the claim. In particular, the inductive hypothesis (5.4) has been proved correct. Hence the computation also establishes that $\boldsymbol{G}(b^+) = \pi_1(\boldsymbol{H}(b^+)) = \boldsymbol{F}(b, \boldsymbol{G}(b))$. So $\boldsymbol{G}$ exists as desired.

The uniqueness of $\boldsymbol{G}$ follows from the uniqueness of $\boldsymbol{H}$, since each of these functions is a function of the other: $\boldsymbol{G} = \pi_1 \circ \boldsymbol{H}$, and also $\boldsymbol{H} = ((x, y) \mapsto (x, \boldsymbol{G}(x, y))$. □

5.4.3. Now we can define a function $x \mapsto \mathrm{pred}(x)$ on $\mathbb{N}$ by the rules[2]

[2]In using the notation $\mathrm{pred}(x)$ I follow Kunen [9, III 5.3]. The notation stands for *predecessor,* although some writers (As Levy [10, II.3.30]) use this term for what I would call an *immediate* predecessor.

(i) $\operatorname{pred}(0) = \varnothing$,

(ii) $\operatorname{pred}(a^+) = \operatorname{pred}(a) \cup \{a\}$.

Then

$$a < b \Leftrightarrow \operatorname{pred}(a) \subset \operatorname{pred}(b) :$$

this follows from the definition of $<$ in ¶5.4.1; it can also be used as a definition of $<$. A third possibility for a definition of $<$ on $\mathbb{N}$ is given by part (iv) of the following.

5.4.4 Theorem. *In* $\mathbb{N}$,

(i) $0 \leqslant a$,

(ii) $a < b \Leftrightarrow a + c < b + c$,

(iii) $a < b \Leftrightarrow a \cdot c^+ < b \cdot c^+$,

(iv) $a \leqslant b \Leftrightarrow \exists x\ a + x = b$.

Proof. Exercise 13. □

5.4.5. Theorem 5.4.2 allows to make some standard definitions:

(i) Letting $\boldsymbol{C} = \mathbb{N}$ and $i = 1$ and $\boldsymbol{F} = ((x, y) \mapsto x^+ \cdot y)$, we have the operation $x \mapsto x!$ on $\mathbb{N}$, defined recursively by

$$0! = 1 \ \&\ (n+1)! = n! \cdot (n+1).$$

(ii) Suppose $\boldsymbol{F} \colon \mathbb{N} \to \boldsymbol{C}$, where $\boldsymbol{C}$ is the universe of a structure equipped with addition and multiplication. Then the sum $\sum_{k=0}^{n} \boldsymbol{F}(k)$ and the product $\prod_{k=0}^{n} \boldsymbol{F}(k)$ are defined recursively as follows:

(a) $\sum_{k=0}^{0} \boldsymbol{F}(k) = \boldsymbol{F}(0)$ and $\sum_{k=0}^{n^+} \boldsymbol{F}(k) = \sum_{k=0}^{n} \boldsymbol{F}(k) + \boldsymbol{F}(n^+)$;

(b) $\prod_{k=0}^{0} \boldsymbol{F}(k) = \boldsymbol{F}(0)$ and $\prod_{k=0}^{n^+} \boldsymbol{F}(k) = \prod_{k=0}^{n} \boldsymbol{F}(k) \cdot \boldsymbol{F}(n^+)$.

5.5 The integers and the rational numbers

5.5.1. In $\mathbb{N}$, if $a \leqslant b$, then (Theorem 5.4.4) there is a solution to the equation

$$a + x = b;$$

this solution is unique (Theorem 5.2.7 (iv)) and can be denoted

$$b - a.$$

If $a \leqslant b$ and $c \leqslant d$, then $b - a = d - c \Leftrightarrow a + d = b + c$; if $a + d = b + c$ and $a \leqslant b$, then $c \leqslant d$ (Exercise 16).

5.5.2 Theorem and Definition. *On $\mathbb{N} \times \mathbb{N}$, let $\boldsymbol{E}$ be the binary relation given by*

$$(a, b)\ \boldsymbol{E}\ (c, d) \Leftrightarrow a + d = b + c.$$

Then $\boldsymbol{E}$ is an equivalence-relation. The virtual class $(\mathbb{N} \times \mathbb{N})/\boldsymbol{E}$ is denoted by

$$\mathbb{Z};$$

*its elements are the **integers.** The class*

$$\{(x, y)\colon (x, y) \in \mathbb{N} \times \mathbb{N}\ \&\ (x = 0 \vee y = 0)\}$$

contains exactly one representative for each $\boldsymbol{E}$-class, and nothing else; so we can treat $\mathbb{Z}$ as this class. Let the $\boldsymbol{E}$-class $(a, b)\boldsymbol{E}$ be denoted

$$b - a.$$

*Then addition and **additive inversion** and multiplication of $\boldsymbol{E}$-classes, and a total ordering of them, can be defined by the following rules:*

$$\begin{aligned}(b - a) + (d - c) &= (b + d) - (a + c),\\ -(b - a) &= a - b,\\ (b - a) \cdot (d - c) &= (b \cdot d + a \cdot c) - (b \cdot c + a \cdot d),\\ b - a < d - c &\Leftrightarrow b + c < a + d.\end{aligned}$$

*In $\mathbb{Z}$, let $0 - 0$ be denoted by 0, and let $1 - 0$ be denoted by 1. Then $(\mathbb{Z}, +, -, \cdot, 0, 1, <)$ is an **ordered ring:***

(i) $+$ *and* $\cdot$ *are commutative and associative;*

(ii) $\cdot$ *distributes over* $+$*;*

(iii) $a + 0 = a$*;*

(iv) $a \cdot 1 = a$*;*

(v) $<$ *is a total ordering;*

(vi) $0 < a\ \&\ 0 < b \Rightarrow 0 < a + b\ \&\ 0 < a \cdot b$.

There is an embedding of $(\mathbb{N}, +, \cdot, 0, 1, <)$ in $(\mathbb{Z}, +, \cdot, 0, 1, <)$ by the rule taking a to $a - 0$.

Proof. Exercise 17. To validate the definitions of $+$ and $\cdot$ in $\mathbb{Z}$, one must show that, if $(a_0, b_0)\ \boldsymbol{E}\ (a_1, b_1)$ and $(c_0, d_0)\ \boldsymbol{E}\ (c_1, d_1)$, then

$$\begin{aligned}&(a_0 + c_0, b_0 + d_0)\ \boldsymbol{E}\ (a_1 + c_1, b_1 + d_1),\\ &(a_0 \cdot c_0 + b_0 \cdot d_0, b_0 \cdot c_0 + a_0 \cdot d_0)\ \boldsymbol{E}\ (a_1 \cdot c_1 + b_1 \cdot d_1, b_1 \cdot c_1 + a_1 \cdot d_1),\\ &b_0 + c_0 < a_0 + d_0 \Leftrightarrow b_1 + c_1 < a_1 + d_1.\end{aligned} \tag{5.5}$$

For (5.5), consider $(a_1 \cdot c_0 + b_1 \cdot d_0, b_1 \cdot c_0 + a_1 \cdot d_0)$. □

5·5·3. In $\mathbb{Z}$, if $a \neq 0$, if there is a solution to the equation

$$a \cdot x = b,$$

then the solution is unique by Theorem 5.3.4 (v) and can be denoted by

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

or by b/a; also a is a **divisor** of b (see ¶6.1.1). If b/a and d/c exist in $\mathbb{Z}$, then $b/a = d/c \Leftrightarrow a \cdot d = b \cdot c$ (Exercise 18).

5·5·4 Theorem and Definition. *On* $(\mathbb{Z} \smallsetminus \{0\}) \times \mathbb{Z}$*, let* $\boldsymbol{E}$ *be the binary relation given by*

$$(a, b)\ \boldsymbol{E}\ (c, d) \Leftrightarrow a \cdot d = b \cdot c.$$

Then $\boldsymbol{E}$ *is an equivalence-relation. The virtual class* $((\mathbb{Z} \smallsetminus \{0\}) \times \mathbb{Z})/\boldsymbol{E}$ *is denoted by*

$$\mathbb{Q};$$

its elements are the ***rational numbers.*** *Let the* $\boldsymbol{E}$*-class* $(a, b)\boldsymbol{E}$ *be denoted*

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

Then addition, additive inversion, multiplication, and multiplicative inversion of $\boldsymbol{E}$*-classes, and a total ordering of them, can be defined by the following rules:*

$$\begin{aligned}
\frac{b}{a} + \frac{d}{c} &= \frac{a \cdot d + b \cdot c}{ac}, \\
-\frac{b}{a} &= \frac{-b}{a}, \\
\left(\frac{b}{a}\right)^{-1} &= \frac{a}{b} \quad \textit{(if } b \neq 0\textit{)}, \\
\frac{b}{a} \cdot \frac{d}{c} &= \frac{b \cdot d}{a \cdot c}, \\
\frac{b}{a} < \frac{d}{c} &\Leftrightarrow a \cdot b \cdot c \cdot c < a \cdot a \cdot c \cdot d.
\end{aligned}$$

In $\mathbb{Q}$*, let* $0 - 0$ *be denoted by* 0*, and let* $1 - 0$ *be denoted by* 1*. Then* $(\mathbb{Q}, +, -, \cdot, 0, 1, <)$ *is an* ***ordered field:*** *it is an ordered ring, and*

$$\frac{a}{b} \cdot \frac{b}{a} = 1$$

if $a \neq 0$ *and* $b \neq 0$*. Then* $(\mathbb{Z}, +, \cdot, 0, 1, <)$ *embeds in* $(\mathbb{Q}, +, \cdot, 0, 1, <)$ *by the rule taking* a *to* $a/1$*.*

Proof. Exercise 19. □

Exercises

(1) Write down bijections from $(\boldsymbol{C} \times \boldsymbol{C}) \times \boldsymbol{C}$ to $\boldsymbol{C} \times (\boldsymbol{C} \times \boldsymbol{C})$ and $\boldsymbol{C}^3$.

(2) Show that an ordered pair $(\boldsymbol{C}, \boldsymbol{D})$ of *classes* can be defined as the class $(\boldsymbol{C} \times \{0\}) \cup (\boldsymbol{D} \times \{1\})$: that is, show that an equivalence like (3.23) holds in this case. Define an ordered triple $(\boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E})$ of classes.

(3) Prove that an iterative structure admits induction if and only if every embedding into it is a surjection onto the universe (¶5.1.6).

(4) Prove the claim in the proof of Theorem 5.2.2 that $\boldsymbol{R}_0$ is an operation on $\boldsymbol{A}$, and $\boldsymbol{R}_0 = \mathrm{id}_{\boldsymbol{A}}$, and if $\boldsymbol{R}_d$ is an operation on $\boldsymbol{A}$, then so is $\boldsymbol{R}_{d^+}$, and $\boldsymbol{R}_{d^+} = \boldsymbol{S} \circ \boldsymbol{R}_d$. Note carefully how it is required that $x \mapsto x^+ : \mathbb{N} \rightarrowtail \mathbb{N} \setminus \{0\}$.

(5) Prove Theorem 5.2.7.

(6) Prove Theorem 5.3.2. You may use, as a pattern, the proof of Theorem 5.2.2.

(7) Prove Lemma 5.3.3.

(8) Prove Theorem 5.3.4.

(9) As an alternative to the proofs of Theorems 5.2.2 and 5.3.2, use the Recursion Theorem to show that addition and multiplication exist as operations on $\mathbb{N}$.

(10) Prove Theorem 5.3.7.

(11) Find an example showing that exponentiation does not always exist as a binary operation on $\boldsymbol{A}$.

(12) Show that there is a unique binary operation $(x, y) \mapsto \binom{x}{y}$ on $\mathbb{N}$ such that $\binom{a}{0} = 1$ and $\binom{0}{b^+} = 0$ and $\binom{a^+}{b^+} = \binom{a}{b} + \binom{a}{b^+}$. Can such an operation be defined on any structures that admit induction, but are *not* arithmetic?

(13) Prove Theorem 5.4.4.

(14) Prove $1 + a \cdot b \leqslant (1 + a)^b$ in $\mathbb{N}$.

(15) Prove $3 < a \Rightarrow a^2 < 2^a$ in $\mathbb{N}$.

(16) Prove the claims in ¶5.5.1.

(17) Prove Theorem 5.5.2. $a \neq 0$.

(18) Prove the claims in ¶5.5.3.

(19) Prove Theorem 5.5.4.

Chapter 6

Ordinality

6.1 Well-ordered classes

6.1.1. A natural number is **prime** if its only divisors are 1 and itself, and these are distinct. So 0 and 1 are not prime, though 2 is prime. Then every natural number different from 1 has a prime divisor; but this does not follow from a simple proof by induction. Indeed, if a has a prime divisor, this does not tell us anything about $a + 1$. However, 2 (like all primes) is a prime divisor of 0; also, if $a > 1$ and is not prime, then a has a divisor b that is not 1 or a. Then $b < a$. Any divisor of b is a divisor of a; so if b is known to have a prime divisor, then a has one also. By *strong induction,* as defined below (¶6.1.2), we can conclude that every natural number different from 1 has a prime divisor.

6.1.2. Suppose $(\boldsymbol{C}, <)$ is a (strict) total order. Following ¶5.4.3, if $a \in \boldsymbol{C}$, we can define

$$\mathrm{pred}(a) = \{x\colon x \in \boldsymbol{C} \mathbin{\&} x < a\}.$$

(One might write $\mathrm{pred}(a, \boldsymbol{C}, <)$ instead of just $\mathrm{pred}(a)$, for complete precision.) Such a class can be called a **section** of $(\boldsymbol{C}, <)$; it is a kind of proper initial segment. If $\boldsymbol{C}$ is *well-ordered* by $<$, then every proper initial segment $\boldsymbol{C}_0$ is equal to the section $\mathrm{pred}(a)$, where a is the least element of $\boldsymbol{C} \smallsetminus \boldsymbol{C}_0$. In any case, let us say that:

(i) $(\boldsymbol{C}, <)$ **admits (proof by) strong** or **trans-finite induction** if

$$\forall x\ (x \in \boldsymbol{C} \mathbin{\&} \mathrm{pred}(x) \subseteq \boldsymbol{C}_0 \Rightarrow x \in \boldsymbol{C}_0) \Rightarrow \boldsymbol{C}_0 = \boldsymbol{C}$$

whenever $\boldsymbol{C}_0 \subseteq \boldsymbol{C}$;

(ii) $(\boldsymbol{C}, <)$ **admits (definition by) strong** or **trans-finite recursion** if:

(a) all sections are sets;

(b) for every class $\boldsymbol{D}$, if

$$\boldsymbol{F}\colon \{x\colon \exists y\ (y \in \boldsymbol{C}\ \&\ x \in {}^{\mathrm{pred}(y)}\boldsymbol{D})\} \to \boldsymbol{D} \tag{6.1}$$

(that is, if $\boldsymbol{F}$ is such that, if g is a function from a section of $\boldsymbol{D}$ into $\boldsymbol{D}$, then $\boldsymbol{F}(g) \in \boldsymbol{D}$), then there is a unique function $\boldsymbol{G}$ from $\boldsymbol{C}$ into $\boldsymbol{D}$ such that

$$\boldsymbol{G}(a) = \boldsymbol{F}(\boldsymbol{G} \restriction \mathrm{pred}(a)).$$

6.1.3 Theorem. *Suppose $(\boldsymbol{C}, <)$ is a total order whose every sub-class* $\mathrm{pred}(a)$ *is a set. The following conditions are equivalent:*

(i) $(\boldsymbol{C}, <)$ *is well-ordered,*

(ii) $(\boldsymbol{C}, <)$ *admits strong induction;*

(iii) $(\boldsymbol{C}, <)$ *admits strong recursion.*

Proof. There are three implications to prove.

(i) Suppose $(\boldsymbol{C}, <)$ is well-ordered. If $\boldsymbol{C}_0 \subset \boldsymbol{C}$, then $\boldsymbol{C} \setminus \boldsymbol{C}_0$ is non-empty, so it has a least element a; then $\mathrm{pred}(a) \subseteq \boldsymbol{C}_0$, but $a \notin \boldsymbol{C}_0$. The contrapositive of this implication is that $(\boldsymbol{C}, <)$ admits strong recursion.

(ii) Suppose $(\boldsymbol{C}, <)$ admits strong induction, $\boldsymbol{D}$ is a class, and $\boldsymbol{F}$ is as in (6.1).

(a) We first show that there is at most one function $\boldsymbol{G}$ from $\boldsymbol{C}$ into $\boldsymbol{D}$ such that $\boldsymbol{G}(a) = \boldsymbol{F}(\boldsymbol{G} \restriction \mathrm{pred}(a))$. Suppose $\boldsymbol{G}_0$ and $\boldsymbol{G}_1$ are two such functions, and that they agree on $\mathrm{pred}(a)$. Then

$$\boldsymbol{G}_0(a) = \boldsymbol{F}(\boldsymbol{G}_0 \restriction \mathrm{pred}(a)) = \boldsymbol{F}(\boldsymbol{G}_1 \restriction \mathrm{pred}(a)) = \boldsymbol{G}_1(a).$$

By strong induction, $\boldsymbol{G}_0$ and $\boldsymbol{G}_1$ agree on $\boldsymbol{C}$. So there is at most one function $\boldsymbol{G}$ as desired.

(b) We next show that there is at least one function $\boldsymbol{G}$ as desired. Suppose that, for each b in $\mathrm{pred}(a)$, there is at least one function $\boldsymbol{G}_b$ from $\mathrm{pred}(b) \cup \{b\}$ into $\boldsymbol{D}$ such that

$$\boldsymbol{G}_b(c) = \boldsymbol{F}(\boldsymbol{G}_b \restriction \mathrm{pred}(c)).$$

Then by what we have just shown, there is at *most* one function $\boldsymbol{G}_b$ (see Exercise 1). In particular then, if $c < b < a$, then $\boldsymbol{G}_c$ and $\boldsymbol{G}_b$ must agree on the domain of the former, so $\boldsymbol{G}_c \subseteq \boldsymbol{G}_b$. Therefore the union $\bigcup\{\boldsymbol{G}_b\colon b < a\}$ is a function $\boldsymbol{H}$, whose domain is

$$\bigcup\{\mathrm{pred}(b) \cup \{b\}\colon b < a\},$$

which is pred(a). Now let

$$\boldsymbol{G}_a = \boldsymbol{H} \cup \{(a, \boldsymbol{F}(\boldsymbol{H}))\}.$$

We have just shown, by strong induction, that for every a in $\boldsymbol{C}$, there is a function $\boldsymbol{G}_a$ from pred(a) $\cup \{a\}$ into $\boldsymbol{D}$ such that $\boldsymbol{G}_a(b) = \boldsymbol{F}(\boldsymbol{G}_a \restriction \text{pred}(b))$. We have also shown that the union $\bigcup\{\boldsymbol{G}_a \colon a \in \boldsymbol{C}\}$ is a function; its domain is $\boldsymbol{C}$; so it is the desired function $\boldsymbol{G}$.

Thus $(\boldsymbol{C}, <)$ admits strong recursion.

(iii) Suppose $(\boldsymbol{C}, <)$ is not well-ordered. Then some sub-class $\boldsymbol{C}_0$ of $\boldsymbol{C}$ has no least element. Then the sub-class $\{x \colon x \in \boldsymbol{C} \mathbin{\&} \exists y\ (y \in \boldsymbol{C} \mathbin{\&} y \leqslant x)\}$ also has no least element. Call this sub-class $\boldsymbol{C}_1$. Let

$$\boldsymbol{F} \colon \{x \colon \exists y\ (y \in \boldsymbol{C} \mathbin{\&} x \in {}^{\text{pred}(y)}2)\} \to 2,$$

where $\boldsymbol{F}(h) = 1 \Leftrightarrow 1 \in \text{rng}(h)$. If $e \in 2$, let $\boldsymbol{G}_e$ be the function from $\boldsymbol{C}$ into 2 given by

$$\boldsymbol{G}_e(x) = \begin{cases} 0, & \text{if } x \in \boldsymbol{C} \setminus \boldsymbol{C}_1; \\ e, & \text{if } x \in \boldsymbol{C}_1. \end{cases}$$

Then $\boldsymbol{G}_e(a) = \boldsymbol{F}(\boldsymbol{G}_e \restriction \text{pred}(a))$:

(a) If $a \in \boldsymbol{C} \setminus \boldsymbol{C}_1$, and $b < a$, then $b \notin \boldsymbol{C}_1$; so $\boldsymbol{G}_e[\,\text{pred}(a)\,] = \{0\}$, hence $\boldsymbol{F}(\boldsymbol{G}_e \restriction \text{pred}(a)) = 0 = \boldsymbol{G}_e(a)$.

(b) If $a \in \boldsymbol{C}_1$, then a is not the least element of $\boldsymbol{C}_1$, which means $e \in \boldsymbol{G}_e[\,\text{pred}(a)\,]$. Then $\boldsymbol{F}(\boldsymbol{G}_e \restriction \text{pred}(a)) = e = \boldsymbol{G}_e(a)$.

Thus $(\boldsymbol{C}, <)$ does not admit strong recursion. □

6.1.4. Suppose $(\boldsymbol{C}, <)$ is a non-empty well-ordered class. Then $\boldsymbol{C}$ has a least element, which we may call 0 (or more precisely $0^{(\boldsymbol{C}, <)}$). Every element a of $\boldsymbol{C}$ has a **successor,** which (following ¶5.1.7) we may denote by

$$a^+;$$

it is the least element of $\boldsymbol{C} \setminus \{x \colon x \in \boldsymbol{C} \mathbin{\&} x \leqslant a\}$, that is, $\boldsymbol{C} \setminus (\text{pred}(a) \cup \{a\})$. Then an element of $\boldsymbol{C}$ is a **successor,** if it is the successor of some element. An element of $\boldsymbol{C}$ that is neither a successor nor 0 is a **limit.** So a is a limit if and only if

(i) $a \neq 0$ and

(ii) $b < a \Rightarrow b^+ < a$.

Note that ω contains no limits. We do not yet have an example that does; after ¶6.3.7, we shall.

6.1.5 Theorem. *Suppose $(\boldsymbol{C}, <)$ is a well-ordered class, and $\boldsymbol{D}$ is a sub-class such that:*

(i) $0 \in \boldsymbol{D}$ *(that is,* $0^{(\boldsymbol{C},<)} \in \boldsymbol{D}$*),*

(ii) $a \in \boldsymbol{D} \Rightarrow a^+ \in \boldsymbol{D}$*;*

(iii) if a is a limit and $\mathrm{pred}(a) \subseteq \boldsymbol{D}$*, then* $a \in \boldsymbol{D}$*.*

Then $\boldsymbol{D} = \boldsymbol{C}$.

Proof. Exercise 3. □

6.1.6 Theorem. *Suppose $(\boldsymbol{C}, <)$ is a well-ordered class, $\boldsymbol{F}\colon \boldsymbol{D} \to \boldsymbol{D}$, and*

$$\boldsymbol{G}\colon \{x\colon \exists y\ (y \in \boldsymbol{C}\ \&\ \forall z\ (z \in \boldsymbol{C} \Rightarrow z^+ \neq y)\ \&\ x \in {}^{\mathrm{pred}(y)}\boldsymbol{C})\} \to \boldsymbol{D},$$

Then there is a unique function $\boldsymbol{H}$ from $\boldsymbol{C}$ to $\boldsymbol{D}$ such that

(i) $\boldsymbol{H}(a^+) = \boldsymbol{F}(\boldsymbol{H}(a))$*,*

(ii) $\boldsymbol{H}(d) = \boldsymbol{G}(\boldsymbol{H} \restriction \mathrm{pred}(d))$ *if d is limit or* 0*.*

Proof. Exercise 4. □

6.2 Order-types

6.2.1. By definition (¶4.5.3), every ordinal α is well-ordered by containment. We may then understand α to denote the structure $(\alpha, \in)$. Likewise, since **ON** is also well-ordered by containment (¶4.5.6), we may take **ON** to denote the structure $(\mathbf{ON}, \in)$. Finally, since for ordinals, $\alpha \in \beta \Leftrightarrow \alpha \subset \beta$ (¶4.5.4), we may use $\in$ and $\subset$ interchangeably in **ON**; we may also use $<$ for either of them, and we may use $\leqslant$ in place of $\subseteq$. But let us continue to use α' (rather than α^+) to denote the successor $\alpha \cup \{\alpha\}$ of the ordinal α.

6.2.2. Suppose $(a, <)$ is an order, and $b \subseteq a$, and $c \in a$. Then c is an **upper bound** for b (with respect to $<$) if $d \in b \Rightarrow d \leqslant c$; and c is a **strict upper bound** if $d \in b \Rightarrow d < c$. If b has a *least* upper bound, then this is unique and is the **supremum** of b; it is denoted by

$$\sup(b).$$

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

$$\exists x\ x = \bigcup a.$$

6.2.4 Lemma. *The union of a set of ordinals is an ordinal, which is the supremum of the set:*

$$a \subset \mathbf{ON} \Rightarrow \bigcup a = \sup(a).$$

Proof. Let a be a set of ordinals. Ordinals are sets of ordinals (Lemma 4.5.4), so $\bigcup a$ is a sub-class of **ON**. Hence $\bigcup a$ is well-ordered by inclusion. Since $\bigcup a$ is a set by the Union Axiom. Also, let $\beta \in \bigcup a$. Then some element α of a contains β. But then α is transitive, so $\beta \subseteq \alpha \subseteq \bigcup a$. Therefore $\bigcup a$ is an ordinal.

Finally, if $\alpha \in a$, then $\alpha \subseteq \bigcup a$; so $\bigcup a$ is an upper bound of a. If $\beta < \bigcup a$, then β belongs to an element of a; that is, β is less than that element, so β is not an upper bound of a. □

6.2.5 Lemma. *If a is a set of ordinals, then $\bigcup\{x' : x \in a\}$ is the least ordinal that is greater than the ordinals in a.*

Proof. Exercise 6. □

6.2.6 Theorem and Definition. *Every well-ordered set $\mathfrak{a}$ is uniquely isomorphic to a unique ordinal, denoted by*

$$\mathrm{ord}(\mathfrak{a})$$

*and called the **order-type** or the **ordinality** of $\mathfrak{a}$. Every well-ordered proper class is uniquely isomorphic to* **ON**.

Proof. Let $(\boldsymbol{C}, <)$ be a well-ordered class. Suppose $\boldsymbol{F}$ is an isomorphism from this to α or **ON**. This means that, if $\boldsymbol{C}$ contains a and b, and $b < a$, then $\boldsymbol{F}(b) \in \boldsymbol{F}(a)$. Thus $\boldsymbol{F}(a)$ is greater than the ordinals in $\boldsymbol{F}[\,\mathrm{pred}(a)\,]$. Since $\boldsymbol{F}$ is surjective onto an initial segment of **ON**, we conclude that $\boldsymbol{F}(a)$ is the *least* ordinal greater than those in $\boldsymbol{F}[\,\mathrm{pred}(a)\,]$. Therefore, by Lemma 6.2.5,

$$\boldsymbol{F}(a) = \bigcup\{x' : x \in \mathrm{rng}(\boldsymbol{F} \restriction \mathrm{pred}(a))\}.$$

By strong recursion, there is exactly one such function $\boldsymbol{F}$ from $\boldsymbol{C}$ into **ON**.

It remains to show that $\boldsymbol{F}$ is indeed the desired isomorphism. Again by Lemma 6.2.5, $\boldsymbol{F}$ is a homomorphism; therefore it is an isomorphism onto its range (Exercise 5). We have to show that $\boldsymbol{F}[\,\boldsymbol{C}\,]$ is an initial segment of **ON**. Suppose $\alpha \subseteq \boldsymbol{F}[\,\boldsymbol{C}\,]$. In particular then, being an ordinal, α is an initial segment of $\boldsymbol{F}[\,\boldsymbol{C}\,]$, so $\boldsymbol{F}^{-1}[\,\alpha\,]$ is an initial segment of $(\boldsymbol{C}, <)$. If it is a *proper* initial segment, then (¶6.1.2) it is a section $\mathrm{pred}(a)$, and then $\boldsymbol{F}(a) = \alpha$ by definition of $\boldsymbol{F}$. By strong induction, either $\boldsymbol{F}[\,\boldsymbol{C}\,] = \mathbf{ON}$, or $\boldsymbol{F}[\,\boldsymbol{C}\,] = \alpha$ for some ordinal α. □

6.3 Ordinal addition

6.3.1. Suppose $(\boldsymbol{C}, <)$ and $(\boldsymbol{D}, <)$ are total orders. The **(right) lexicographic ordering** of $\boldsymbol{C} \times \boldsymbol{D}$ is given by

$$(a, b) < (c, d) \Leftrightarrow b < d \vee (b = d \,\&\, a < c).$$

This is a *total* ordering of $\boldsymbol{C} \times \boldsymbol{D}$ (Exercise 8; see also Fig. 6.1). If $(\boldsymbol{C}, <)$ and $(\boldsymbol{D}, <)$ are *well-ordered,* then the lexicographic order well-orders $\boldsymbol{C} \times \boldsymbol{D}$: Indeed, if $\boldsymbol{E} \subseteq \boldsymbol{C} \times \boldsymbol{D}$, then its least element is (a, b), where b is the least element of $\{y\colon \exists x\ (x, y) \in \boldsymbol{E}\}$, and a is the least element of $\{x\colon (x, b) \in \boldsymbol{C}\}$.

6.3.2. The **(ordinal) sum** of two ordinals α and β is the ordinality of a well-ordered set that is like α *followed by* β. Suppose there are embeddings f and g of α and β respectively in a common well-ordered set such that $f(\gamma) < g(\delta)$ whenever $\gamma \in \alpha$ and $\delta \in \beta$; then we can take the sum of α and β to be the ordinality of $f[\,\alpha\,] \cup g[\,\beta\,]$. To be precise, we define

$$\alpha + \beta = \operatorname{ord}((\alpha \times \{0\}) \cup (\beta \times \{1\})),$$

where $(\alpha \times \{0\}) \cup (\beta \times \{1\})$ has the lexicographic ordering of $(\alpha \cup \beta) \times 2$. But we must confirm that this definition agrees with the definition in ¶5.2.2 when the ordinals are natural numbers.

6.3.3 Theorem. *For all ordinals α, β, and γ,*

(i) $0 + \alpha = \alpha + 0 = \alpha$,

(ii) $\alpha' = \alpha + 1$,

(iii) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$,

(iv) $\alpha < \beta \Rightarrow \gamma + \alpha < \gamma + \beta$,

(v) if $\alpha < \beta$, then $\alpha + x = \beta$ has a unique ordinal solution.

Proof. Exercise 9. In (v), the solution is $\operatorname{ord}(\alpha \smallsetminus \beta)$; it is unique by (iv). □

6.3.4 Theorem. *Ordinal addition has the recursive definition*

(i) $\alpha + 0 = \alpha$,

(ii) $\alpha + \beta' = (\alpha + \beta)'$,

(iii) $\alpha + \beta = \bigcup\{\alpha + x\colon x \in \beta\}$ *when β is a limit.*

In particular, addition on ω is as defined in ¶5.2.2.

Proof. Exercise 10. □

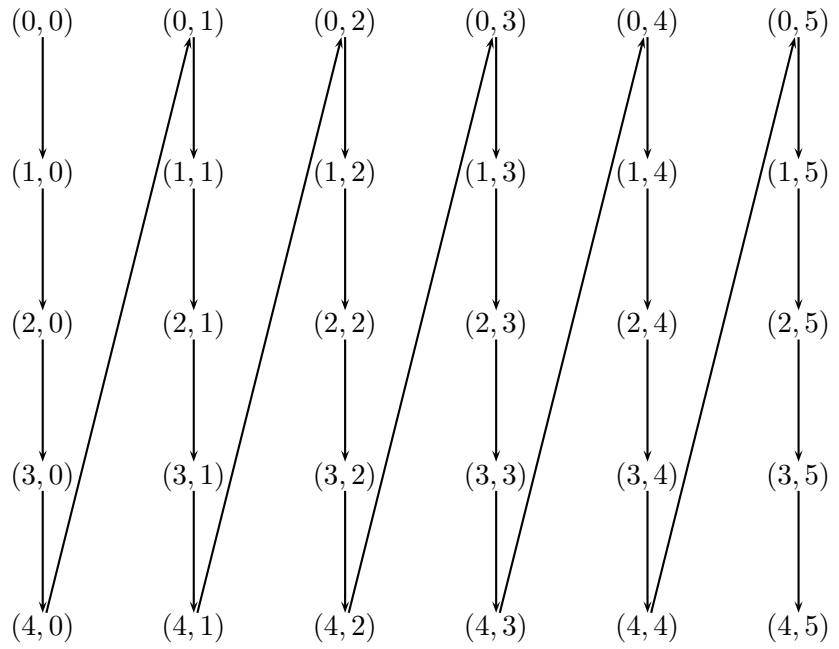

Figure 6.1: The lexicographic ordering of 5×6

6.3.5 Lemma. *Suppose* $\mathfrak{a}$ *and* $\mathfrak{b}$ *are well-ordered sets, and* $f\colon \mathfrak{a} \to \mathfrak{b}$. *Then* $\operatorname{ord}(\mathfrak{a}) \leqslant \operatorname{ord}(\mathfrak{b})$.

Proof. The homomorphism f is an embedding, which induces an embedding g of $\operatorname{ord}(\mathfrak{a})$ in $\operatorname{ord}(\mathfrak{b})$. We shall show by induction that $\alpha < \operatorname{ord}(\mathfrak{a}) \Rightarrow \alpha \leqslant g(\alpha)$. Suppose this is so when $\alpha < \beta$. Say also $\beta < \operatorname{ord}(\mathfrak{a})$. Then $\alpha \leqslant g(\alpha) < g(\beta)$, so $\alpha < g(\beta)$. Briefly, $\alpha < \beta \Rightarrow \alpha < g(\beta)$; so $\beta \leqslant g(\beta)$. This completes the induction. Since $g(\alpha) < \operatorname{ord}(\mathfrak{b})$ whenever $\alpha < \operatorname{ord}(\mathfrak{a})$, we conclude $\alpha < \operatorname{ord}(\mathfrak{a}) \Rightarrow \alpha < \operatorname{ord}(\mathfrak{b})$, and hence $\operatorname{ord}(\mathfrak{a}) \leqslant \operatorname{ord}(\mathfrak{b})$. □

6.3.6 Theorem. $\alpha \leqslant \beta \Rightarrow \alpha + \gamma \leqslant \beta + \gamma$.

Proof. If $\alpha \leqslant \beta$, then $(\alpha \times \{0\}) \cup (\gamma \times \{1\}) \subseteq (\beta \times \{0\}) \cup (\gamma \times \{1\})$; this inclusion is an embedding of the well-ordered sets, so $\alpha + \gamma \leqslant \beta + \gamma$ by Lemma 6.3.5. □

6.3.7 Axiom (Infinity)**.** *The class of natural numbers is a set:*

$$\exists x\ x = \omega.$$

6.3.8 Theorem. $n < \omega \Rightarrow n + \omega = \omega$.

Proof. Define f from ω into $(n \times \{0\}) \cup (\omega \times \{1\})$ by

$$f(x) = \begin{cases} (x, 0), & \text{if } x < n; \\ (y, 1), & \text{if } x = n + y. \end{cases}$$

Then f is an isomorphism. □

6.3.9. By Theorem 6.3.3 (iv), along with the Axiom of Infinity, we have the following initial segment of **ON**:

$$\{0, 1, 2, \dots; \omega, \omega + 1, \omega + 2, \dots; \omega + \omega, \omega + \omega + 1, \dots; \omega + \omega + \omega, \dots\}.$$

Here the ordinals following the semicolons (;) are limits. By Theorem 6.3.8, addition involving infinite ordinals is not commutative; also the ordering in Theorem 6.3.6 cannot be made strict: we have $0 < 1$, but $0 + \omega = \omega = 1 + \omega$.

6.4 Ordinal multiplication

6.4.1. The **(ordinal) product** of two ordinals is the ordinality of their product with the lexicographic ordering (¶6.3.1):

$$\alpha \cdot \beta = \operatorname{ord}(\alpha \times \beta).$$

We must confirm that this agrees with ¶5.3.2 on ω.

6.4.2 Theorem. *For all ordinals α, β, and γ,*

(i) $1 \cdot \alpha = \alpha \cdot 1 = \alpha$;

(ii) $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$;

(iii) $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$;

(iv) $\alpha < \beta \;\&\; 0 < \gamma \Rightarrow \gamma \cdot \alpha < \gamma \cdot \beta$;

(v) $\alpha \leqslant \beta \Rightarrow \alpha \cdot \gamma \leqslant \beta \cdot \gamma$;

(vi) $\alpha \cdot x + y = \beta \;\&\; y < \alpha$ *has a unique ordinal solution if* $0 < \alpha$.

Proof. Exercise 11. For (vi), show $\beta \leqslant \alpha \cdot \beta$. If $\beta < \alpha \cdot \beta$, then β is isomorphic to a section $\operatorname{pred}((\gamma, \delta))$ of $\alpha \times \beta$. But $\operatorname{pred}((\gamma, \delta)) = (\alpha \times \delta) \cup (\gamma \times \{\delta\})$; so $\alpha \cdot \delta + \gamma = \beta$ and $\gamma < \alpha$. Now show uniqueness. □

6.4.3 Theorem. *Ordinal multiplication has the recursive definition*

(i) $\alpha \cdot 0 = 0$,

(ii) $\alpha \cdot \beta' = \alpha \cdot \beta + \alpha$,

(iii) $\alpha \cdot \beta = \bigcup\{\alpha \cdot x \colon x \in \beta\}$ *when* β *is a limit.*

In particular, multiplication on ω *is as in* ¶5.3.2.

Proof. Exercise 12. □

6.4.4 Theorem. $0 < n < \omega \Rightarrow n \cdot \omega = \omega$.

Proof. Exercise 13. (Write out an isomorphism from $n \times \omega$ to ω.) □

6.4.5. We can now extend the initial segment of **ON** in ¶6.3.9:

$$\{0, 1, \dots; \omega, \omega + 1, \dots; \omega \cdot 2, \dots; \omega \cdot 3, \dots; \omega \cdot \omega, \dots; \omega \cdot \omega \cdot \omega, \dots\}.$$

By Theorem 6.4.2 (iii) and Theorem 6.4.4, ordinal multiplication is not commutative, nor does it distribute from the right over addition:

$$(1 + 1) \cdot \omega = 2 \cdot \omega = \omega < \omega + \omega = 1 \cdot \omega + 1 \cdot \omega = \omega \cdot 1 + \omega \cdot 1 = \omega \cdot 2;$$

also the ordering in Theorem 6.4.2 (v) cannot be made strict.

6.5 Ordinal exponentiation

6.5.1. An endomorphism $\boldsymbol{F}$ of **ON** is called **normal** if

$$\boldsymbol{F}(\alpha) = \sup(\boldsymbol{F}[\,\alpha\,])$$

(that is, $\boldsymbol{F}(\alpha) = \bigcup \boldsymbol{F}[\,\alpha\,]$) whenever α is a limit. For example, the function $x \mapsto \alpha + x$ is an endomorphism of **ON** by Theorem 6.3.3 (iv); it is normal by Theorem 6.3.4. Likewise, if $0 < \alpha$, then $x \mapsto \alpha \cdot x$ is normal by Theorems 6.4.2 (iv) and 6.4.3.

6.5.2 Theorem and Definition. *There is a unique binary operation of* ***(ordinal) exponentiation*** *on* **ON**, *denoted*

$$(x, y) \mapsto x^y,$$

such that:

(i) $0^\beta = 0$ *when* $\beta \neq 0$;

(ii) $\alpha^0 = 1$;

(iii) $\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$;

(iv) $\alpha^\beta = \bigcup\{\alpha^x \colon x \in \beta\}$ *when* β *is a limit and* $\alpha \neq 0$.

On ω, *this operation is as given in* ¶5.3.5. *If* $1 < \alpha$, *then* $x \mapsto \alpha^x$ *is normal.*

Proof. Exercise 14. (Normality follows immediately, once one has

$$0 < \alpha \,\&\, \beta < \gamma \Rightarrow \alpha^\beta < \alpha^\gamma,$$

which can be proved by induction on γ.) □

6.5.3 Lemma. *If $\boldsymbol{F}$ is normal and $c \subset \mathbf{ON}$, then*

$$\boldsymbol{F}(\sup(c)) = \sup(\boldsymbol{F}[\,c\,]).$$

Proof. Let $\alpha = \sup(c)$. If $\alpha \in c$, then α is the greatest element of c, so $\sup(\boldsymbol{F}[\,c\,]) = \boldsymbol{F}(\alpha)$ since $\boldsymbol{F}$ preserves order. Suppose $\alpha \notin c$. Then $c \subseteq \alpha$, so $\sup(\boldsymbol{F}[\,c\,]) \leqslant \sup(\boldsymbol{F}[\,\alpha\,])$. Also, if $\beta < \alpha$, then $\beta \leqslant \gamma < \alpha$ for some γ in c, so $\sup(\boldsymbol{F}[\,\alpha\,]) \leqslant \bigcup \boldsymbol{F}[\,c\,]$. Therefore $\sup(\boldsymbol{F}[\,\alpha\,]) = \bigcup \boldsymbol{F}[\,c\,]$. But α must be a limit, so $\sup(\boldsymbol{F}[\,\alpha\,]) = \boldsymbol{F}(\alpha)$ by normality of $\boldsymbol{F}$. □

6.5.4 Theorem. *For all ordinals α, β, and γ,*

(i) $0 < \alpha \Rightarrow 0^\alpha = 0$,

(ii) $1^\alpha = 1$,

(iii) $\alpha^{\beta+\gamma} = \alpha^\beta \cdot \alpha^\gamma$,

(iv) $\alpha^{\beta\cdot\gamma} = (\alpha^\beta)^\gamma$.

Proof. Exercise 16. For (iii), by Lemma 6.5.3, if $1 < \alpha$ and γ is a limit, one has by induction

$$\begin{aligned}
\alpha^\beta \cdot \alpha^\gamma &= \alpha^\beta \cdot \sup(\{\alpha^x : x \in \gamma\}) \\
&= \sup(\{\alpha^\beta \cdot \alpha^x : x \in \gamma\}) \\
&= \sup(\{\alpha^{\beta+x} : x \in \gamma\}) \\
&= \alpha^{\sup(\{\beta+x\,:\,x\in\gamma\})} \\
&= \alpha^{\beta+\sup(\{x\,:\,x\in\gamma\})} \\
&= \alpha^{\beta+\gamma}.
\end{aligned}$$

Likewise for (iv). □

6.5.5. We now have the following initial segment of **ON**:

$$\{0, 1, \ldots; \omega, \omega+1, \ldots, \omega\cdot 2, \ldots; \omega^2, \omega^2+1, \ldots; \omega^2+\omega, \ldots;$$
$$\omega^2\cdot 2, \ldots; \omega^3, \ldots; \omega^\omega, \ldots; \omega^{\omega\cdot 2}, \ldots; \omega^{\omega^2}, \ldots; \omega^{\omega^\omega}, \ldots; \omega^{\omega^{\omega^\omega}}, \ldots\}.$$

This set is closed under the operations that we have defined so far. Its supremum is denoted by

$$\epsilon_0;$$

this is $\bigcup \mathrm{rng}(\boldsymbol{F})$, where $\boldsymbol{F}$ is the function on ω defined recursively by

(i) $\boldsymbol{F}(0) = 1$,

(ii) $\boldsymbol{F}(n+1) = \omega^{\boldsymbol{F}(n)}$.

Exercises

(1) Show that every subset of an ordered set that admits strong induction admits strong induction.

(2) Let $(\boldsymbol{C}, <)$ be a total order whose every section is a set. Show that $(\boldsymbol{C}, <)$ admits recursion if and only if, for every class $\boldsymbol{D}$ and every function $\boldsymbol{F}$ from $\mathcal{P}(\boldsymbol{D})$ into $\boldsymbol{D}$, there is a unique function $\boldsymbol{G}$ from $\boldsymbol{C}$ into $\boldsymbol{D}$ such that $\boldsymbol{G}(a) = \boldsymbol{F}(\boldsymbol{G}[\,\mathrm{pred}(a)\,])$.

(3) Prove Theorem 6.1.5.

(4) Prove Theorem 6.1.6.

(5) Suppose $(\boldsymbol{A}, \boldsymbol{R})$ and $(\boldsymbol{B}, \boldsymbol{S})$ are strict total orders, and h is a homomorphism from the former to the latter. Show that $h : (\boldsymbol{A}, \boldsymbol{R}) \xrightarrow{\cong} (\mathrm{rng}(h), \boldsymbol{S})$.

(6) Prove Lemma 6.2.5.

(7) Prove that an ordinal α is 0 or a limit if and only if $\alpha = \bigcup \alpha$.

(8) Prove that the lexicographic ordering is total (¶6.3.1).

(9) Prove Theorem 6.3.3.

(10) Prove Theorem 6.3.4.

(11) Prove Theorem 6.4.2.

(12) Prove Theorem 6.4.3.

(13) Prove Theorem 6.4.4.

(14) Prove Theorem 6.5.2.

(15) Prove $1 < \alpha \Rightarrow \beta \leqslant \alpha^{\beta}$.

(16) Prove Theorem 6.5.4.

Chapter 7

Cardinality

7.1 Finite sets

7.1.1. A set is called *Dedekind-infinite* if it is equipollent with a proper subset of itself (¶4.2.1). The definition is an attempt to capture what we mean by calling something infinite. One attempt to capture what we mean by *finite* runs as follows. First, let a subset b of a power-set $\mathcal{P}(a)$ be called a **inductive** subset of $\mathcal{P}(a)$ if

(i) $\varnothing \in b$,

(ii) $c \in b \,\&\, d \in a \Rightarrow c \cup \{d\} \in b$.

If $c \subseteq a$, then $b \cap \mathcal{P}(c)$ is an inductive subset of $\mathcal{P}(c)$ (Exercise 1). A set is called **finite** if it is an element of each inductive subset of its power-set. Trivially then, $\varnothing$ is finite.

7.1.2 Lemma. *If a is finite, then so is $a \cup \{b\}$.*

Proof. Assuming a finite, let c be an inductive subset of $\mathcal{P}(a \cup \{b\})$; we must show $a \cup \{b\} \in c$. But $c \cap \mathcal{P}(a)$ is an inductive subset of $\mathcal{P}(a)$ (¶7.1.1). Hence $a \in c$. Therefore $a \cup \{b\} \in c$, by definition of an inductive set. □

7.1.3 Theorem (Induction)**.** *Suppose $\boldsymbol{C}$ is a class of finite sets such that*

(i) $\varnothing \in \boldsymbol{C}$,

(ii) $a \in \boldsymbol{C} \Rightarrow a \cup \{b\} \in \boldsymbol{C}$.

Then $\boldsymbol{C}$ contains all finite sets.

Proof. Let a be finite. By the hypothesis on $\boldsymbol{C}$, the set $\{x \in \mathcal{P}(a) \colon x \in \boldsymbol{C}\}$ is an inductive subset of $\mathcal{P}(a)$; therefore it contains a. So $a \in \boldsymbol{C}$. □

7.1.4 Lemma. *The image of a finite set is finite.*

Proof. Let $\boldsymbol{C}$ comprise those finite sets whose images are all finite. Trivially, $\varnothing \in \boldsymbol{C}$. Suppose $a \in \boldsymbol{C}$ and $a \cup \{b\} \subseteq \mathrm{dom}(\boldsymbol{F})$. Then $\boldsymbol{F}[\,a\,]$ is finite, and $\boldsymbol{F}[\,a \cup \{b\}\,] = \boldsymbol{F}[\,a\,] \cup \{\boldsymbol{F}(b)\}$, so this is also finite. Thus $a \cup \{b\} \in \boldsymbol{C}$. By induction (¶7.1.3), $\boldsymbol{C}$ contains all finite sets. □

7.1.5 Theorem. *A set is finite if and only if it is equipollent with a natural number.*

Proof. All natural numbers are finite, by induction:

(i) $\varnothing$ is finite;

(ii) if n is finite, then so is n', that is, $n \cup \{n\}$, by Lemma 7.1.2.

Hence all sets equipollent with natural numbers are finite, by Lemma 7.1.4.

Conversely, every finite set is equipollent with a natural number, by induction (¶7.1.3):

(i) $\varnothing \approx \varnothing$;

(ii) if $a \approx n$, then $a \cup \{b\} \approx n \cup \{n\}$ (assuming $b \notin a$). □

7.1.6 Lemma. *Suppose a and b are in ω, and $f\colon a' \rightarrowtail b'$. Then $g\colon a \rightarrowtail b$, where*

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

Also, $f\colon a' \twoheadrightarrow b' \Leftrightarrow g\colon a \twoheadrightarrow b$.

Proof. If f is a bijection, then

$$\mathrm{rng}(g) = f[\,a' \smallsetminus \{f^{-1}(b)\}\,] = f[\,a'\,] \smallsetminus \{b\} = b' \smallsetminus \{b\} = b,$$

so g is surjective. If g is a bijection, then

$$\mathrm{rng}(f) = \mathrm{rng}(g) \cup f[\,f^{-1}[\,\{b\}\,]\,] = b \cup \{b\} = b',$$

so f is surjective. For the injectivity of g, assuming the injectivity of f, there are two cases. If $f(a) = b$, then g is simply $f \restriction a$, so it preserves the injectivity of f. Suppose $f(c) = b$, where $c < b$. Then $g(c) = f(a)$, but $a \notin \mathrm{dom}(g)$, so $g(d) = f(d) \neq f(a)$ when $d \in a \smallsetminus \{c\}$. Again g is injective. □

7.1.7 Theorem. *No finite set is Dedekind-infinite.*

Proof. By Theorem 7.1.5, it is enough to show that no natural number is Dedekind-infinite. We show that every injection from a natural number into itself is surjective. This is trivially true for 0, and if it is true for a, then it is true for a', by Lemma 7.1.6: if $f\colon a' \rightarrowtail a'$, then there is an injection from a into itself, but this is then a surjection too by inductive hypothesis, so f is surjective. □

7.1.8 Theorem. *Equipollent natural numbers are equal.*

Proof. We show

$$a \in \omega \,\&\, b \in \omega \,\&\, a + b \approx a \Rightarrow b = 0.$$

This is trivially true when $a = 0$. Suppose it is true when $a = c$. Say $a' + b \approx a'$. Since $a' + b = (a + b)'$, by Lemma 7.1.6 there is a bijection from $a + b$ to a. By the inductive hypothesis then, $b = 0$. □

7.2 Cardinals

7.2.1. If a set a can be well-ordered, then a is equipollent with some ordinal number (Theorem 6.2.6). In this case, we can define the **cardinality** of a as the *least* ordinal that is equipollent with a. If a cannot be well-ordered, then we still have to understand its **cardinality** as the class $\{x\colon x \approx a\}$, as in ¶4.1.1. In either case, as suggested in ¶4.1.1, the cardinality of a is denoted by

$$\mathrm{card}(a).$$

We shall ultimately (with ¶7.5.6) decide that there are *no* sets that cannot be well-ordered; but not everything that can be said about cardinalities requires this assumption. So, for now, we have two kinds of cardinalities:

(i) those that are ordinal numbers: these can be called **cardinal numbers** or just **cardinals;**

(ii) the other cardinalities: the classes $\{x\colon x \approx a\}$: where a cannot be well-ordered.

The cardinal *numbers* compose the class denoted by

$$\mathbf{CN};$$

this is a sub-class of **ON**. Then **CN** inherits the ordering of **ON**, usually denoted by $<$ as suggested in ¶6.2.1; and this ordering coincides with $\prec$ (¶4.1.2; Exercise 2).

7.2.2 Lemma (Hartogs)**.** *For every set, there is an ordinal that does not embed in it.*

Proof. If a is a set, let b be the set of well-ordered sets $(c, <)$ such that $c \subseteq a$. If $\mathrm{ord}(c, <) = \beta$, and $\gamma < \beta$, then $\mathrm{ord}(d, <) = \gamma$ for some section d of c. This shows that $\{\mathrm{ord}(\mathfrak{c})\colon \mathfrak{c} \in b\}$ is a transitive subset of **ON**; so it is an ordinal α. If $f : \beta \rightarrowtail a$, then f determines an element of b whose ordinality is β; so $\beta \in \alpha$. Since $\alpha \notin \alpha$, there is no injection of α in a. □

7.2.3. As a consequence of the Hartogs Lemma, for every cardinal κ, there is a cardinal α such that $\kappa < \alpha$, but $\kappa \not\approx \alpha$. Therefore $\kappa < \mathrm{card}(\alpha)$. Thus κ has a **successor,**

$$\kappa^+;$$

it is the *least* of the cardinals that are greater than κ.

7.2.4 Lemma. *The union of a set of cardinals is a cardinal.*

Proof. Let a be a set of cardinals. The union $\bigcup a$ is $\sup(a)$, an ordinal (Lemma 6.2.4). Hence, if κ is a cardinal less than $\sup(a)$, then $\kappa < \lambda$ for some λ in a, and therefore $\kappa \neq \mathrm{card}(\sup(a))$. Therefore $\sup(a)$ must be a cardinal (namely its own cardinality). □

7.2.5 Theorem and Definition. *The class* $\{x\colon x \in \mathbf{CN} \ \&\ \omega \leqslant x\}$ *of infinite cardinals is a proper class; there is an isomorphism*

$$x \mapsto \aleph_x$$

from **ON** *onto this class, given by*

(i) $\aleph_0 = \omega$,

(ii) $\aleph_{\alpha'} = \aleph_\alpha{}^+$,

(iii) $\aleph_\alpha = \bigcup\{\aleph_x\colon x \in \alpha\}$ *if* α *is a limit.*

If $0 < \alpha$, *then* $\aleph_\alpha$ *is called* ***uncountable.***

Proof. Exercise 3. □

7.3 Cardinal addition and multiplication

7.3.1. The **(cardinal) sum** and **(cardinal) product** of two cardinals κ and λ are defined by

$$\kappa + \lambda = \mathrm{card}((\kappa \times \{0\}) \cup (\lambda \times \{1\})),$$
$$\kappa \cdot \lambda = \mathrm{card}(\kappa \times \lambda).$$

If κ and λ are merely cardinalities, not ordinals, then

$$\kappa + \lambda = \mathrm{card}((a \times \{0\}) \cup (b \times \{1\})), \tag{7.1}$$
$$\kappa \cdot \lambda = \mathrm{card}(a \times b), \tag{7.2}$$

where $a \in \kappa$ and $b \in \lambda$ (Exercise 4). One must distinguish the cardinal operations from the ordinal operations of ¶¶ 6.3.2 and 6.4.1. However, the cardinal operations, when involving infinite *cardinals* (and not merely cardinalities), will turn out to be very simple.

7.3.2 Theorem. *If κ, λ, and μ are cardinalities, then*

(i) $\kappa + \lambda = \lambda + \kappa$,

(ii) $\kappa + 0 = \kappa$,

(iii) $(\kappa + \lambda) + \mu = \kappa + (\lambda + \mu)$,

(iv) $\kappa \cdot \lambda = \lambda \cdot \kappa$,

(v) $\kappa \cdot 1 = \kappa$,

(vi) $(\kappa \cdot \lambda) \cdot \mu = \kappa \cdot (\lambda \cdot \mu)$,

(vii) $\kappa \cdot (\lambda + \mu) = \kappa \cdot \lambda + \kappa \cdot \mu$,

(viii) $\kappa \leqslant \lambda \Rightarrow \kappa + \mu \leqslant \lambda + \mu$,

(ix) $\kappa \leqslant \lambda \Rightarrow \kappa \cdot \mu \leqslant \lambda \cdot \mu$.

The cardinal operations agree with the ordinal operations on ω.

Proof. Exercise 7. □

7.3.3 Lemma. *The class* **ON** *of ordinals becomes isomorphic to* $\mathbf{ON} \times \mathbf{ON}$ *when the latter is ordered by $<$, where*

$$(\alpha, \beta) < (\gamma, \delta) \Leftrightarrow \max(\alpha, \beta) < \max(\gamma, \delta) \vee \\ \vee \Big(\max(\alpha, \beta) = \max(\gamma, \delta) \;\&\; \big(\alpha < \gamma \vee (\alpha = \gamma \;\&\; \beta < \delta)\big)\Big).$$

(Here $\max(\alpha, \beta)$ is the maximal element of $\{\alpha, \beta\}$. See Fig. 7.1.)

Proof. Exercise 8. After showing that $<$ is a total ordering, one can show that the least element of a non-empty subset a of $\mathbf{ON} \times \mathbf{ON}$ is (β, γ), where

$$\gamma = \min\{y\colon (\beta, y) \in a\},$$
$$\beta = \min\{x\colon \exists y\ \max(x, y) = \alpha\},$$
$$\alpha = \min\{\max(x, y)\colon (x, y) \in a\}.$$

Now show that every proper initial segment is a set. □

7.3.4 Theorem. *If κ and λ are cardinals, $0 < \kappa \leqslant \lambda$, and λ is infinite, then*

$$\kappa + \lambda = \kappa \cdot \lambda = \lambda.$$

That is,

$$\aleph_\alpha + \aleph_\beta = \aleph_\alpha \cdot \aleph_\beta = \aleph_{\max(\alpha,\beta)}.$$

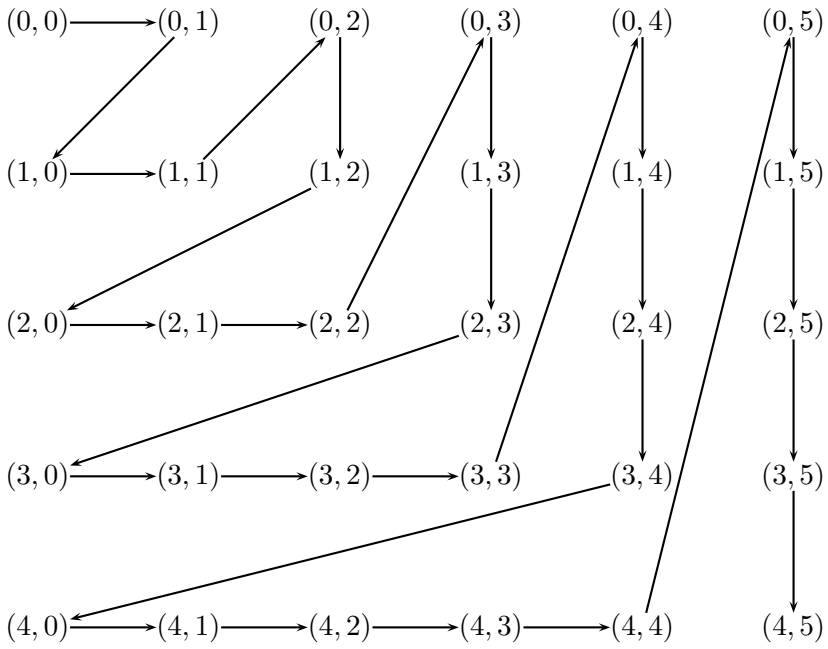

Figure 7.1: **ON** × **ON**, well-ordered

Proof. We shall show

$$\lambda \cdot \lambda = \lambda. \tag{7.3}$$

Then we can complete the argument by observing

$$\lambda \leqslant \kappa + \lambda \leqslant \lambda + \lambda = \lambda \cdot 2 \leqslant \lambda \cdot \lambda,$$
$$\lambda \leqslant \kappa \cdot \lambda \leqslant \lambda \cdot \lambda.$$

We establish (7.3) by induction on the infinite cardinals. Suppose the equation holds whenever $\omega \leqslant \lambda < \mu$, for some cardinal μ. Let $\boldsymbol{F}$ be the isomorphism from **ON** × **ON** onto **ON** guaranteed by Lemma 7.3.3. As $\alpha \times \alpha$ is always a section of **ON** × **ON**, so $\boldsymbol{F}[\,\alpha \times \alpha\,]$ must be a section of **ON**: that is, $\boldsymbol{F}[\,\alpha \times \alpha\,]$ is an ordinal. Suppose $\boldsymbol{F}[\,\mu \times \mu\,] = \alpha$. Then

$$\mu \leqslant \mu \cdot \mu = \mathrm{card}(\mu \times \mu) = \mathrm{card}(\alpha) \leqslant \alpha.$$

So $\mu \leqslant \alpha$. Suppose ν is an infinite cardinal and $\nu < \alpha$. Then $\nu = \boldsymbol{F}(\beta, \gamma)$, where $(\beta, \gamma) \in \mu \times \mu$. Since μ is a limit ordinal (Exercise 5), the successor δ of $\max(\beta, \gamma)$ is also less than μ. Hence

$$\nu \in \boldsymbol{F}[\,\delta \times \delta\,],$$
$$\nu \subseteq \boldsymbol{F}[\,\delta \times \delta\,],$$
$$\nu \leqslant \mathrm{card}(\delta \times \delta) = \mathrm{card}(\delta) \cdot \mathrm{card}(\delta)$$

(Exercise 6). By inductive hypothesis, $\mathrm{card}(\delta) \cdot \mathrm{card}(\delta) = \mathrm{card}(\delta)$, so $\nu \leqslant \mathrm{card}(\delta) < \delta < \mu$. In short, $\nu < \alpha \Rightarrow \nu < \mu$. Therefore $\alpha \leqslant \mu$; but since $\mu \leqslant \alpha$, we have $\mu = \alpha = \mu \cdot \mu$. □

7.4 Exponentiation

7.4.1. Cardinal exponentiation is as easy to define as cardinal addition and multiplication: If κ and λ are cardinals, then

$$\kappa^\lambda = \mathrm{card}({}^\lambda\kappa);$$

if κ and λ are merely cardinalities, containing representatives a and b respectively, then

$$\kappa^\lambda = \mathrm{card}({}^a b).$$

7.4.2 Theorem. *For all cardinalities κ, λ, μ and ν,*

(i) $\kappa^0 = 1$,

(ii) $0 < \lambda \Rightarrow 0^\lambda = 0$,

(iii) $1^\lambda = 1$,

(iv) $\kappa^1 = \kappa$,

(v) $\kappa^2 = \kappa \cdot \kappa$,

(vi) $\kappa^{\lambda+\mu} = \kappa^\lambda \cdot \kappa^\mu$,

(vii) $\kappa^{\lambda\cdot\mu} = (\kappa^\lambda)^\mu$,

(viii) $\kappa \leqslant \mu \;\&\; \lambda \leqslant \nu \Rightarrow \kappa^\lambda \leqslant \mu^\nu$.

Proof. Exercise 9. □

7.4.3 Theorem. *For all sets a,*

$$\mathrm{card}(\mathcal{P}(a)) = 2^{\mathrm{card}(a)}; \tag{7.4}$$

hence for all cardinalities κ,

$$\kappa < 2^\kappa. \tag{7.5}$$

Proof. There is a bijection between ${}^a 2$ and $\mathcal{P}(a)$ that takes the function f to the set $\{x \in a\colon f(x) = 1\}$. This establishes (7.4). Then (7.5) follows by Cantor's Theorem (¶4.1.6). □

7.4.4 Corollary. *If κ and λ are cardinals such that $2 \leqslant \kappa \leqslant \lambda$ and λ is infinite, then*

$$\kappa^\lambda = 2^\lambda.$$

Proof. $2^\lambda \leqslant \kappa^\lambda \leqslant \lambda^\lambda \leqslant (2^\lambda)^\lambda = 2^{\lambda\cdot\lambda} = 2^\lambda$ by Theorem 7.3.4. □

7.4.5 Theorem. *There is a normal embedding*

$$x \mapsto \beth_x$$

of **ON** *in the class of infinite cardinals given by*

(i) $\beth_0 = \aleph_0$,

(ii) $\beth_{\alpha'} = 2^{\beth_\alpha}$,

(iii) $\beth_\alpha = \bigcup\{\beth_x \colon x \in \alpha\}$ *if* α *is a limit.*

In particular,

$$\aleph_\alpha \leqslant \beth_\alpha. \tag{7.6}$$

Proof. The definition is a valid recursive definition. The function preserves the ordering, by Theorem 7.4.3 and induction; it then meets the definition of a normal operation (¶6.5.1). Finally, (7.6) holds, again by Theorem 7.4.3 and induction. □

7.4.6. The letter $\beth$ is *beth,* the second letter of the Hebrew alphabet. The **Continuum Hypothesis** is that $\aleph_1 = \beth_1$; the **Generalized** Continuum Hypothesis is that $\aleph_\alpha = \beth_\alpha$ for all ordinals α. The **continuum** is the set of *real numbers,* defined in § 7.7.

7.5 The Axiom of Choice

7.5.1. Suppose a is not finite. If $n \in \omega$ and $f \colon n \rightarrowtail a$, then f is not surjective; so $a \smallsetminus f[\,n\,]$ has an element b, and then $f \cup \{(n, b)\} \colon n + 1 \rightarrowtail a$. Since, trivially, $0 \colon 0 \rightarrowtail a$, we have by induction that every natural number embeds in a. However, this by itself does not allow us to conclude that ω embeds in a.

7.5.2. A **choice-function** for a set is a function that assigns, to each non-empty subset of the set, an element of that subset. Suppose f is a choice-function for a, so that

$$b \in \mathcal{P}(a) \smallsetminus \{\varnothing\} \Rightarrow f(b) \in b.$$

If also a is non-finite, then there is an embedding g of ω in a defined recursively by

$$g(n) = f(a \smallsetminus g[\,n\,]).$$

Thus every non-finite set with a choice-function is Dedekind-infinite.

7.5.3 Theorem. *A set has a choice-function if and only if the set can be well-ordered.*

Proof. Suppose a has the choice-function f. Assume also $f(\varnothing)$ is defined, but is not in a. Then there is a function $\boldsymbol{G}$ on **ON** defined recursively by

$$\boldsymbol{G}(\alpha) = f(a \smallsetminus \boldsymbol{G}[\,\alpha\,]).$$

Then $\boldsymbol{G}^{-1}[\,a\,]$ is an initial segment of **ON**, and $\boldsymbol{G} \upharpoonright \boldsymbol{G}^{-1}[\,a\,]$ is a bijection onto a. So, by means of $\boldsymbol{G}$, the ordering of $\boldsymbol{G}^{-1}[\,a\,]$ induces a well-ordering of a.

Now suppose conversely that a is well-ordered. Then there is a choice--function for a that assigns to each non-empty subset of a its least element. □

7.5.4 Theorem. *Suppose $(a, <)$ is an order such that every totally ordered subset of a has an upper bound in a. If a has a choice-function, then a has a maximal element.*

Proof. If $b \subseteq a$, let $f(b)$ be the set (possibly empty) of strict upper bounds of b. So f is order-reversing, in the sense that

$$c \subseteq b \Rightarrow f(c) \supseteq f(b).$$

Let g be a choice-function for a, and assume $g(\varnothing) \notin a$. Then define $\boldsymbol{H}$ on **ON** by

$$\boldsymbol{H}(\alpha) = g(f(a \cap \boldsymbol{H}[\,\alpha\,])).$$

Then $\boldsymbol{H}$ is an embedding of $\boldsymbol{H}^{-1}[\,a\,]$ in a, so $a \cap \mathrm{rng}(\boldsymbol{H})$ is totally ordered, but it has no strict upper bound in a. By hypothesis though, it has an upper bound; this is a maximal element of a. □

7.5.5. Let a be the set of choice-functions for subsets of b. Then a is ordered by proper inclusion. Also, the union of any totally ordered subset of a belongs to a and is an upper bound for the subset. If $f \in a$, but $\mathrm{dom}(f) = \mathcal{P}(c) \smallsetminus \{\varnothing\}$, where $c \subset b$, then $b \smallsetminus c$ contains some d. Then $f \subset g$, where g is a choice-function for $c \cup \{d\}$ such that $g(e) = d$ when $d \in e$. Thus f is not a maximal element of a. Therefore any maximal element of a is a choice-function for b.

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

7.5.7. The Axiom of Choice is a completely new kind of axiom, since it asserts the existence of sets that we do not already have as classes. However, the axiom is very convenient. In the presence of the Axiom of Choice, the conclusion of Theorem 7.5.4 is often known as Zorn's Lemma (though it appears to be due to Hausdorff). Then ¶7.5.5 shows that assuming the Axiom of Choice is *equivalent* to assuming Zorn's Lemma (this conclusion apparently *is* due to Zorn). By Theorem 7.5.3 now, every set has a cardinality in the sense of ¶7.2.1; and this too is equivalent to the Axiom of Choice.

7.6 Computations

7.6.1. From the definitions of ordinal and cardinal addition and multiplication, we have

$$\begin{aligned}\operatorname{card}(\alpha+\beta) &= \operatorname{card}(\alpha)+\operatorname{card}(\beta),\\ \operatorname{card}(\alpha\cdot\beta) &= \operatorname{card}(\alpha)\cdot\operatorname{card}(\beta).\end{aligned}$$

Here the ordinal operations are on the left; the cardinal, on the right. So these equations illustrate a convention: If an operation is applied to ordinals that are not necessarily cardinals, then the operation is the ordinal operation; if the ordinals *are* known to be cardinals, then the operation is the cardinal operation (even though the ordinal operation would also make sense). Following strictly the notation of ¶5.1.4, we would write for example

$$\operatorname{card}(\alpha +^{\mathbf{ON}} \beta) = \operatorname{card}(\alpha) +^{\mathbf{CN}} \operatorname{card}(\beta).$$

7.6.2. Exponentiation is different. We defined ordinal exponentiation so that $x \mapsto \alpha^x$ would be normal, assuming $0 < \alpha$. An alternative definition is as follows. We shall define the function $x \mapsto \alpha^x$ on $\beta+1$. First, we define a function $\exp_\alpha$ from $\beta+1$ into $\mathcal{P}({}^{\beta+1}\alpha)$ recursively:

(i) $\exp_\alpha(0) = {}^0\alpha$,

(ii) $\exp_\alpha(\gamma+1) = \{x \in {}^{\gamma+1}\alpha \colon x \restriction \gamma \in \exp_\alpha(\gamma)\}$,

(iii) $\exp_\alpha(\gamma) = \bigcup \exp_\alpha[\,\gamma\,]$ if γ is a limit.

Then $\exp_\alpha$ is a homomorphism: $\gamma < \delta \Rightarrow \exp_\alpha(\gamma) \subset \exp_\alpha(\delta)$. We can also recursively assign well-orderings to the sets $\exp_\alpha(\gamma)$:

(i) $\exp_\alpha(0)$ has the empty ordering;

(ii) if g and h are in $\exp_\alpha(\gamma+1)$, then

$$g < h \Leftrightarrow g(\gamma) < h(\gamma) \vee (g(\gamma) = h(\gamma) \,\&\, g \restriction \gamma < h \restriction \gamma);$$

(iii) if γ is a limit, then $\exp_\alpha(\gamma)$ is ordered by the union of the orderings of the $\exp_\alpha(\delta)$ where $\delta < \gamma$.

By induction, $\delta < \gamma \Rightarrow (\exp_\alpha(\delta), <) \subseteq (\exp_\alpha(\gamma), <)$, and each $\exp_\alpha(\gamma)$ is well-ordered; in particular, $\exp_\alpha(\beta)$ is well-ordered. Then we can define

$$\alpha^\beta = \operatorname{ord}(\exp_\alpha(\beta)).$$

By induction, this definition agrees with the definition in ¶6.5.2.

7.6.3. The **support** of an element of ${}^{\beta}\alpha$ is the set of ordinals where the function is not 0. We may write

$$\operatorname{supp}(f) = \{x\colon x \in \operatorname{dom}(f) \mathbin{\&} f(x) \neq 0\}.$$

Then $\exp_\alpha(\beta)$ is the set of elements of ${}^{\beta}\alpha$ with finite support, and the ordering is the (right) lexicographic ordering: that is, $f < g$ if and only if $f(\gamma) < g(\gamma)$, where γ is the greatest element of $\{x \in \beta\colon f(x) \neq g(x)\}$; because this set is finite, γ does exist. See Fig. 7.2.

7.6.4 Lemma. *If b is a non-empty set, and $x \mapsto a_x$ is a function on b such that each set a_c is non-empty, but $c \neq d \Rightarrow a_c \cap a_d = \varnothing$, and one of b and $\sup\{\operatorname{card}(a_x)\colon x \in b\}$ is infinite, then*

$$\operatorname{card}\Big(\bigcup\{a_x\colon x \in b\}\Big) = \operatorname{card}(b) \cdot \sup\{\operatorname{card}(a_x)\colon x \in b\}.$$

Proof. Let $\kappa = \operatorname{card}(\bigcup\{a_x\colon x \in b\})$ and $\lambda = \sup\{\operatorname{card}(a_x)\colon x \in b\}$. For each c in b, there is a bijection f_c from a_c onto $\operatorname{card}(a_c)$. Now let f be the function from $\bigcup\{a_x\colon x \in b\}$ into $b \times \lambda$ such that $f(d) = (c, f_c(d))$ when $d \in a_c$. Since f is injective, we have $\kappa = \operatorname{card}(\operatorname{rng}(f)) \leqslant \operatorname{card}(b) \cdot \lambda$. Also $\operatorname{rng}(f)$ includes $b \times 1$, so $\operatorname{card}(b) \leqslant \kappa$. If $\alpha < \lambda$, then $\alpha < \operatorname{card}(a_c)$ for some c in b, but then $\{c\} \times \alpha \subseteq \operatorname{rng}(f)$, so $\alpha \leqslant \kappa$. Thus $\lambda \leqslant \kappa$ (since λ is a limit ordinal). By Theorem 7.3.4, we are done. □

7.6.5. The preceding proof used the Axiom of Choice, since a bijection f_c from a_c onto $\operatorname{card}(a_c)$ was chosen for each c in b. But there may be many such bijections. To define f, we need to know more than that such functions f_c exist; we must be able to specify one of them for each c in b. So let g be a choice-function for the set e of bijections between elements of $\{a_x\colon x \in b\}$ and their cardinals. (See Exercise 12.) Then $f_c = g(\{x \in e\colon \operatorname{dom}(x) = a_c\})$, so $f(d) = (c, g(\{x \in e\colon \operatorname{dom}(x) = a_c\})(d)) \Leftrightarrow d \in a_c$.

7.6.6 Lemma. *If a is an infinite set, then*

$$\operatorname{card}(\{x \in \mathcal{P}(a)\colon \operatorname{card}(x) < \aleph_0\}) = \operatorname{card}(a).$$

Proof. Exercise 13. Use Lemma 7.6.4. □

7.6.7 Theorem. *If $1 < \alpha$, and $0 < \beta$, and one of α and β is infinite, then*

$$\operatorname{card}(\alpha^\beta) = \operatorname{card}(\alpha) \cdot \operatorname{card}(\beta).$$

Proof. Let b be the set of finite subsets of β. Then

$$\operatorname{card}(\alpha^\beta) = \operatorname{card}(\bigcup\{{}^{x}(\alpha \smallsetminus \{0\})\colon x \in b\}).$$

If α is infinite and b is finite, then $\operatorname{card}({}^{b}(\alpha \smallsetminus \{0\})) = \operatorname{card}(\alpha)$ (Exercise 10). If β is infinite, then $\operatorname{card}(b) = \operatorname{card}(\beta)$ by Lemma 7.6.6. In either case, we are done by Lemma 7.6.4. □

$$
\begin{array}{c}
(0,0,0,0,\dots)\\
(1,0,0,0,\dots)\\
(2,0,0,0,\dots)\\
\dots\dots\dots\dots\\
(0,1,0,0,\dots)\\
(1,1,0,0,\dots)\\
(2,1,0,0,\dots)\\
\dots\dots\dots\dots\\
(0,0,1,0,\dots)\\
(1,0,1,0,\dots)\\
\dots\dots\dots\dots\\
(0,1,1,0,\dots)\\
\dots\dots\dots\dots
\end{array}
$$

Figure 7.2: The ordering of $\{x \in {}^{\beta}\alpha\colon \operatorname{card}(\operatorname{supp}(x)) < \aleph_0\}$.

7.7 The real numbers

7.7.1. A **cut** of a total order $(\boldsymbol{C}, <)$ is a non-empty proper initial segment that does not contain its supremum (if it has a supremum). The cuts of **ON** are precisely the limit ordinals, and every cut of **ON** does have a supremum, namely itself. Some cuts of $\mathbb{Q}$ have suprema, others don't: $\operatorname{pred}(a, \mathbb{Q}, <)$ has the supremum a, but $\{x \in \mathbb{Q}\colon x^2 < 2 \vee x < 0\}$ has no supremum in $\mathbb{Q}$. If we picture $\mathbb{Q}$ as comprising points on a horizontal line, with a cut on the left and its complement on the right, then the two possibilities are as in Fig. 7.3. (If a is a cut of $\mathbb{Q}$, then the ordered pair $(a, \mathbb{Q} \setminus a)$ is a cut in the original sense of Dedekind [3, p. 13].)

7.7.2. Let $\mathbb{R}$ be the set of cuts of $\mathbb{Q}$; these are the so-called **real numbers.** Then $\mathbb{R}$ is totally ordered by inclusion, and

$$x \mapsto \operatorname{pred}(x)\colon (\mathbb{Q}, <) \to (\mathbb{R}, \subset).$$

Now define the ring-operations on $\mathbb{R}$. First,

$$
\begin{aligned}
a + b &= \{x + y\colon x \in a \mathbin{\&} y \in b\},\\
-a &= \{-x\colon x \in \mathbb{Q} \setminus a \mathbin{\&} \exists y\, (y \in \mathbb{Q} \setminus a \mathbin{\&} y < x)\}.
\end{aligned}
$$

Multiplication is defined by cases. First, if $0 \in a \cap b$ (so that a and b are *positive* real numbers), then

$$a \cdot b = \{x \cdot y\colon x \in a \mathbin{\&} y \in b\}.$$

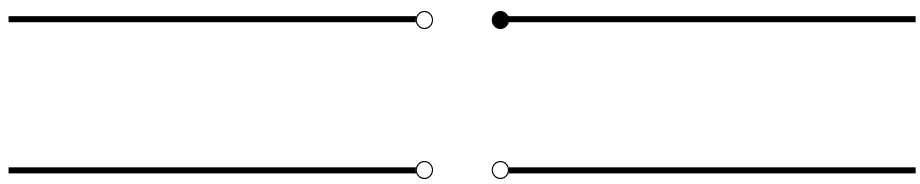

Figure 7.3: Cuts

If $a = 0^{\mathbb{R}}$ or $b = 0^{\mathbb{R}}$ (that is, if $a = \mathrm{pred}(0)$ or $b = \mathrm{pred}(0)$), then $a \cdot b = 0$. Finally,

$$a \cdot b = \begin{cases} -(-a \cdot b), & \text{if } 0 \in b \smallsetminus a; \\ -(a \cdot -b), & \text{if } 0 \in a \smallsetminus b; \\ -a \cdot -b, & \text{if } 0 \notin a \cup b. \end{cases}$$

Then $x \mapsto \mathrm{pred}(x)$ is an embedding of ordered fields (Exercise 14).

7.7.3. If a is a cut of $\mathbb{R}$, then $\bigcup a$ is a cut of $\mathbb{Q}$ and is the supremum of a (Exercise 15). Then every subset of $\mathbb{R}$ with an upper bound has a supremum: this makes $\mathbb{R}$ a **complete** ordered field.

7.7.4. We have a surjection of $\mathbb{N} \times \mathbb{N}$ onto $\mathbb{Z}$, so (Exercise 11) $\mathrm{card}(\mathbb{Z}) = \aleph_0$. We have a surjection of $\mathbb{Z} \times (\mathbb{Z} \smallsetminus \{0\})$ onto $\mathbb{Q}$, so $\mathrm{card}(\mathbb{Q}) = \aleph_0$. (See also Exercise 16.) Since $\mathbb{R} \subseteq \mathcal{P}(\mathbb{Q})$, we have

$$\mathrm{card}(\mathbb{R}) \leqslant 2^{\aleph_0}. \tag{7.7}$$

7.7.5. There is an embedding of ${}^{\omega}2$ into $\mathbb{R}$, so that $2^{\aleph_0} \leqslant \mathrm{card}(\mathbb{R})$; because of (7.7), by the Schroeder–Bernstein Theorem (¶4.1.3), we then have

$$\mathrm{card}(\mathbb{R}) = 2^{\aleph_0}.$$

To define the embedding, replace $\mathbb{Q}$ with its image in $\mathbb{R}$; that is, treat $\mathbb{Q}$ as a sub-field of $\mathbb{R}$. Let $f\colon {}^{\omega}2 \to \mathbb{R}$, where

$$f(\sigma) = \sup\Big\{\sum_{k=0}^{x} \frac{2 \cdot \sigma(k)}{3^{k+1}} \colon x \in \omega\Big\}.$$

Then f is well-defined, since, by induction,

$$\sum_{k=0}^{n} \frac{2 \cdot \sigma(k)}{3^{k+1}} \leqslant 1 - \frac{1}{3^{n+1}}.$$

Also, f is injective, since, if $\sigma \restriction n = \tau \restriction n$, but $\sigma(n) = 0 < 1 = \tau(n)$, then

$$f(\sigma) \leqslant \sum_{k=0}^{n-1} \frac{2 \cdot \sigma(k)}{3^{k+1}} + \frac{1}{3^n} < \sum_{k=0}^{n-1} \frac{2 \cdot \sigma(k)}{3^{k+1}} + \frac{2}{3^n} \leqslant f(\tau).$$

Here $f[\,{}^{\omega}2\,]$ is called the **Cantor set;** it is the intersection of the sets depicted in Fig. 7.4.

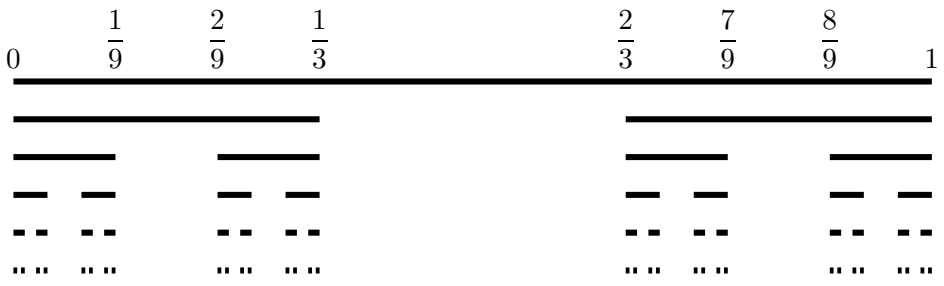

Figure 7.4: Towards the Cantor set

Exercises

(1) If b is an inductive subset of $\mathcal{P}(a)$, and $c \subseteq a$, prove that $b \cap \mathcal{P}(c)$ is an inductive subset of $\mathcal{P}(c)$.

(2) Prove that $\in$ coincides with $\prec$ on **CN**.

(3) Prove Theorem 7.2.5.

(4) Show that (7.1) and (7.2) are independent of the choice of a in κ and b in λ.

(5) Prove that infinite cardinals are limit ordinals.

(6) Show that $\mathrm{card}(a \times b) = \mathrm{card}(a) \cdot \mathrm{card}(b)$.

(7) Prove Theorem 7.3.2 without using Theorem 7.3.4.

(8) Prove Lemma 7.3.3.

(9) Prove Theorem 7.4.2.

(10) If κ is an infinite cardinal, and $1 < n < \omega$, prove $\kappa^n = \kappa$.

(11) Use the Axiom of Choice to show that, if $f\colon a \twoheadrightarrow b$, then $b \preccurlyeq a$.

(12) If a is a set, why is there a *set* of bijections between elements of a and their cardinals?

(13) Prove Lemma 7.6.6.

(14) Show that $x \mapsto \mathrm{pred}(x)$ is an embedding of $\mathbb{Q}$ in $\mathbb{R}$ as ordered fields.

(15) Show that the union of a cut of $\mathbb{R}$ is the supremum of the cut.

(16) Show that $\mathbb{Z}$ and $\mathbb{Q}$ are countable without using the Axiom of Choice.

Chapter 8

Models

8.1 Well-founded sets

8.1.1. We have noted the possibility that a set might contain itself as a member. We shall see now how to rule out this possibility, along with other pathologies like $a \in b \,\&\, b \in a$.

8.1.2. Let $\mathbf{R}$ be the function on $\mathbf{ON}$ given by

(i) $\mathbf{R}(0) = \varnothing$,

(ii) $\mathbf{R}(\alpha + 1) = \mathcal{P}(\mathbf{R}(\alpha))$,

(iii) $\mathbf{R}(\alpha) = \bigcup \mathbf{R}[\,\alpha\,]$ if α is a limit.

Now define

$$\mathbf{WF} = \bigcup \mathbf{R}[\,\mathbf{ON}\,];$$

this is the **well-founded universe.** (We do not have a particular name for the elements of $\mathbf{WF}$ as such.)

8.1.3 Lemma. *Each set* $\mathbf{R}(\alpha)$ *is transitive, so* $\mathbf{WF}$ *is transitive. If* $\beta < \alpha$, *then* $\mathbf{R}(\beta) \subseteq \mathbf{R}(\alpha)$.

Proof. Use induction. Trivially, $\mathbf{R}(0)$ is transitive; the power-set of a transitive set is transitive; the union of a set of transitive sets is transitive. (See Exercise 1.) Therefore each set $\mathbf{R}(\alpha)$ is transitive, so $\mathbf{WF}$ is transitive.

Consequently, as $\mathbf{R}(\alpha) \in \mathbf{R}(\alpha + 1)$, so $\mathbf{R}(\alpha) \subseteq \mathbf{R}(\alpha + 1)$. If $\beta < \alpha$, and α is limit, then immediately $\mathbf{R}(\beta) \subseteq \mathbf{R}(\alpha)$. By induction, $\beta < \alpha \Rightarrow \mathbf{R}(\beta) \subseteq \mathbf{R}(\alpha)$. □

8.1.4. A set is **ill-founded** if there is a function f from ω into the set such that $f(n+1) \in f(n)$ for each n. (There is no requirement that $\in$ be transitive on the set.) A set that is not ill-founded is **well-founded.**

8.1.5 Lemma. *The set a is well-founded if and only if*

$$b \subseteq a \;\&\; b \neq \varnothing \Rightarrow \exists x\, (x \in b \;\&\; x \cap b = \varnothing). \tag{8.10}$$

Proof. Suppose a is ill-founded, so it has a subset $\{b_x \colon x \in \omega\}$, where $b_{n+1} \in b_n$. Then $b_{n+1} \in b_n \cap \{b_x \colon x \in \omega\}$. Thus (8.10) fails.

Suppose conversely that (8.10) fails, so a has a non-empty subset b such that

$$c \in b \Rightarrow c \cap b \neq \varnothing.$$

Then there is a recursively defined function f from ω into b such that $f(n+1) \in f(n) \cap b$. (This uses the Axiom of Choice: assuming b is well-ordered, we let $f(n+1)$ be the *least* element of $f(n) \cap b$.) □

8.1.6. If $a \in \mathbf{WF}$, then the least ordinal α such that $a \in \mathbf{R}(\alpha)$ must be a successor, $\beta + 1$. Then β is the **rank** of a:

$$\operatorname{rank}(a) = \min\{x \in \mathbf{ON} \colon a \in \mathbf{R}(x+1)\}.$$

If $b \in a$, then $b \in \mathbf{WF}$, and $\operatorname{rank}(b) < \operatorname{rank}(a)$ (Exercise 4). Since **ON** is well-ordered, there is no infinite descending sequence $\alpha_0 > \alpha_1 > \alpha_2 > \cdots$ of ordinals. Therefore the elements of **WF** are well-founded. However, the converse may fail: If $a = \{b\}$ and $b = \{a\}$, but $a \neq b$, then $a \notin \mathbf{WF}$ (since otherwise $\operatorname{rank}(a) > \operatorname{rank}(b) > \operatorname{rank}(a)$); but a is well-founded since $a \cap b = \varnothing$. Note however $\bigcup\{a, \bigcup a\} = \{a, b\}$, which is ill-founded.

8.1.7 Lemma. *Every set is included in a transitive set. All transitive well--founded sets are in* **WF**.

Proof. Given a set a, we define the sets $\bigcup^n a$ recursively:

(i) $\bigcup^0 a = a$,

(ii) $\bigcup^{n+1} a = \bigcup \bigcup^n a$.

Now let $\bigcup^\omega a = \bigcup\{\bigcup^x a \colon x \in \omega\}$. Then $a \subseteq \bigcup^\omega a$; also, the latter is transitive, since if $b \in \bigcup^\omega a$, then $b \in \bigcup^n a$ for some n, and therefore $b \subseteq \bigcup^{n+1} a \subseteq \bigcup^\omega a$.

For the second claim, suppose a is transitive, but not in **WF**. All subsets of **WF** are elements of **WF** (Exercise 5); therefore $a \not\subseteq \mathbf{WF}$, so $a \setminus \mathbf{WF}$ has a member a_1, which is also a subset of a. Now continue: if $a_n \notin \mathbf{WF}$, then $a_n \not\subseteq \mathbf{WF}$, so we may let $a_{n+1} \in a_n \setminus \mathbf{WF}$; if also $a_n \subseteq a$, then $a_{n+1} \in a$, so $a_{n+1} \subseteq a$ (see Exercise 6). Therefore a is ill-founded. □

8.1.8 Theorem. *All sets are in* **WF** *if and only if all sets are well-founded:*

$$\forall x\; x \in \mathbf{WF} \Leftrightarrow \forall x\, (x \neq \varnothing \Rightarrow \exists y\, (y \in x \;\&\; y \cap x = \varnothing)).$$

Proof. Exercise 7. □

8.1.9 Axiom (Foundation)**.** *All sets are well-founded:*

$$a \neq 0 \Rightarrow \exists y\, (y \in a \;\&\; y \cap a = \varnothing).$$

8.2 Virtual classes

8.2.1. The Foundation Axiom is like a definition: it declares which sets we wish to study. Unlike most of our axioms, it does not assert the existence of any sets. In the following section, we shall see in a precise way that in fact we do not *need* any ill-founded sets. Meanwhile, in the present section, we show that that it is convenient not to *have* well-founded sets.

8.2.2. Along with the class-operations of § 3.2, there is an operation that assigns to the class $\boldsymbol{C}$ a *set* $\tau(\boldsymbol{C})$ such that

(i) $\tau(\boldsymbol{C}) \subseteq \boldsymbol{C}$;

(ii) $\boldsymbol{C} \neq \varnothing \Rightarrow \tau(\boldsymbol{C}) \neq \varnothing$.

Indeed, if $\boldsymbol{C} \neq \varnothing$, let $\alpha = \min\{x \in \mathbf{ON} \colon \exists y\ (y \in \boldsymbol{C}\ \&\ \mathrm{rank}(y) = x)\}$, and let

$$\tau(\boldsymbol{C}) = \mathbf{R}(\alpha) \cap \boldsymbol{C}.$$

8.2.3 Theorem. *If $\boldsymbol{E}$ is an equivalence-relation on $\boldsymbol{C}$, then there is a function $\boldsymbol{F}$ on $\boldsymbol{C}$ such that*

$$\boldsymbol{F}(a) = \boldsymbol{F}(b) \Leftrightarrow a\ \boldsymbol{E}\ b.$$

Proof. Let $\boldsymbol{F}(a) = \tau(a\boldsymbol{E})$, where τ is as in ¶8.2.2 (and $a\boldsymbol{E}$ is the equivalence-class of a, as in ¶3.6.2). □

8.2.4. Now the virtual class $\boldsymbol{C}/\boldsymbol{E}$ of ¶3.6.2 can be understood as the real class $\boldsymbol{F}[\,\boldsymbol{C}\,]$, where $\boldsymbol{F}$ is as in the theorem.

8.3 Consistency

8.3.1. We now have all of the standard axioms of set-theory. All of these axioms, besides Choice, are known collectively as ZF, in honor of Zermelo and Fraenkel. When Choice is added, the set of axioms is called ZFC. We shall now justify the Foundation Axiom by showing that all of the other axioms in ZF remain true under the assumption that $\mathbf{V} = \mathbf{WF}$.

8.3.2. If $\boldsymbol{M}$ is a class, and φ is a formula, we shall define a **relativization** of φ to $\boldsymbol{M}$, denoted by

$$\varphi^{\boldsymbol{M}}.$$

The aim is that, if φ is n-ary, and $\vec{a} \in \boldsymbol{M}^n$, then $\varphi^{\boldsymbol{M}}(\vec{a})$ means that $\varphi(\vec{a})$ is true under the assumption that all sets belong to $\boldsymbol{M}$. In short, if σ is a sentence, then $\sigma^{\boldsymbol{M}}$ means that σ is **true in** $\boldsymbol{M}$, or that $\boldsymbol{M}$ is a **model** of σ. The formal definition is recursive:

(i) $(s \in t)^{\boldsymbol{M}}$ is $s \in t$ (the same formula);

(ii) $(\neg\varphi)^{\boldsymbol{M}}$ is $\neg\varphi^{\boldsymbol{M}}$ (the negation of $\varphi^{\boldsymbol{M}}$);

(iii) $(\varphi \to \psi)^{\boldsymbol{M}}$ is $\varphi^{\boldsymbol{M}} \to \psi^{\boldsymbol{M}}$;

(iv) $(\exists x\, \varphi)^{\boldsymbol{M}}$ is $\exists x\, (x \in \boldsymbol{M} \,\&\, \varphi^{\boldsymbol{M}})$ (that is, $\exists x\, \neg(x \in \boldsymbol{M} \Rightarrow \neg\varphi^{\boldsymbol{M}})$).

Then $(\forall x\, \varphi)^{\boldsymbol{M}}$ is $\forall x\, (x \in \boldsymbol{M} \Rightarrow \varphi^{\boldsymbol{M}})$. If $\boldsymbol{R} = \{\vec{x}\colon \varphi(\vec{x})\}$, then we can write $\boldsymbol{R}^{\boldsymbol{M}}$ for $\{\vec{x} \in \boldsymbol{M}^n\colon \varphi^{\boldsymbol{M}}(\vec{x})\}$, although, strictly, this class depends on the formula φ chosen to define $\boldsymbol{R}$. (See Exercise 2.)

8.3.3. An n-ary formula φ is **absolute** for $\boldsymbol{M}$ if

$$\varphi^{\boldsymbol{M}}(\vec{a}) \Leftrightarrow \varphi(\vec{a})$$

for all $\vec{a}$ in $\boldsymbol{M}^n$. Then the atomic formula $x \in y$ is always absolute. However, $x = y$ need not be absolute, by our account. Indeed, $x = y$ is an abbreviation for $\forall x\, (x \in z \Leftrightarrow y \in z)$; so the relativization to $\boldsymbol{M}$ is

$$\forall z\, (z \in \boldsymbol{M} \Rightarrow (x \in z \Leftrightarrow y \in z)).$$

So, suppose a and b are distinct sets such that $a \notin b$ and $b \notin a$: perhaps $a = \varnothing$ and $b = \{\{\varnothing\}\}$. If $\boldsymbol{M} = \{a, b\}$, then $(a = b)^{\boldsymbol{M}}$, although $a \neq b$.

8.3.4 Lemma. ZF *is true in* $\boldsymbol{M}$*, provided*

(i) $\boldsymbol{M}$ *is transitive,*

(ii) $a \subseteq \boldsymbol{M} \Rightarrow \exists x\, (a \subseteq x \,\&\, x \in \boldsymbol{M})$*,*

(iii) the Comprehension-Scheme is absolute for $\boldsymbol{M}$*, that is,*

$$a \in \boldsymbol{M} \Rightarrow a \cap \{x \in \boldsymbol{M}\colon \varphi^{\boldsymbol{M}}(x)\} \in \boldsymbol{M}$$

for all (singulary) formulas φ *whose constants are in* $\boldsymbol{M}$*, and*

(iv) $x = y$ *is absolute for* $\boldsymbol{M}$*.*

Proof. Suppose throughout that a and b are arbitrary members of $\boldsymbol{M}$. Note first that $x \subseteq y$ is absolute for $\boldsymbol{M}$, merely because $\boldsymbol{M}$ is transitive. Indeed, this formula stands for $\forall z\, (z \in x \Rightarrow z \in y)$; so the relativization is

$$\forall z\, (z \in x \cap \boldsymbol{M} \Rightarrow z \in y).$$

But $a \cap \boldsymbol{M} = a$, by transitivity of $\boldsymbol{M}$; so $(x \subseteq y)^{\boldsymbol{M}}$ is just $x \subseteq y$.

Since the Extension Axiom can be written as

$$a = b \Leftrightarrow (a \subseteq b \,\&\, b \subseteq a)$$

(¶3.1.7), and $a = b$ is absolute for $\boldsymbol{M}$, so is the axiom.

Foundation is absolute for $\boldsymbol{M}$, by transitivity of $\boldsymbol{M}$.

The remaining axioms to consider are that certain classes $\{x\colon \varphi\}$ are sets. We have to show that the relativizations $\{x \in \boldsymbol{M}\colon \varphi^{\boldsymbol{M}}\}$ are sets that belong to $\boldsymbol{M}$ (assuming the constants of φ are in $\boldsymbol{M}$). Let $\boldsymbol{C} = \{x \in \boldsymbol{M}\colon \varphi^{\boldsymbol{M}}\}$. By Comprehension in $\boldsymbol{M}$, it will be enough to show that $\boldsymbol{C}$ is a set: Indeed, $\boldsymbol{C} \subseteq \boldsymbol{M}$ by definition; if also $\boldsymbol{C}$ is a set, then $\boldsymbol{C} \subseteq d$ for some element d of $\boldsymbol{M}$, and consequently $\boldsymbol{C} = d \cap \boldsymbol{C} \in \boldsymbol{M}$.

(i) Pairing: $\{a, b\}^{\boldsymbol{M}} = \{x \in \boldsymbol{M}\colon x = a \vee x = b\} = \{a, b\}$, a set.

(ii) Union: By transitivity of $\boldsymbol{M}$, we have

$$\begin{aligned}\Big(\bigcup a\Big)^{\boldsymbol{M}} &= \{x \in \boldsymbol{M}\colon \exists y\, (y \in \boldsymbol{M} \mathbin{\&} y \in a \mathbin{\&} x \in y)\}\\ &= \{x \in \boldsymbol{M}\colon \exists y\, (y \in a \mathbin{\&} x \in y)\}\\ &= \bigcup a.\end{aligned}$$

(iii) Infinity: $\omega^{\boldsymbol{M}} = \omega$ (Exercise 8).

(iv) Power-set: $\mathcal{P}(a)^{\boldsymbol{M}} = \boldsymbol{M} \cap \mathcal{P}(a)$.

Replacement is a bit more work. Suppose

$$(\forall x\, \forall y\, \forall z\, (\varphi(x, y) \mathbin{\&} \varphi(x, z) \Rightarrow y = z))^{\boldsymbol{M}}.$$

Then $\{(x, y) \in \boldsymbol{M} \times \boldsymbol{M}\colon \varphi^{\boldsymbol{M}}(x, y)\}$ is a function $\boldsymbol{F}$. Say $a \subseteq \mathrm{dom}(\boldsymbol{F})$. Then

$$\begin{aligned}\boldsymbol{F}[\,a\,]^{\boldsymbol{M}} &= \{y \in \boldsymbol{M}\colon (\exists x\, (x \in a \mathbin{\&} \varphi(x, y)))^{\boldsymbol{M}}\}\\ &= \{y \in \boldsymbol{M}\colon \exists x\, (x \in a \cap \boldsymbol{M} \mathbin{\&} \varphi^{\boldsymbol{M}}(x, y))\}\\ &= \{y \in \boldsymbol{M}\colon \exists x\, (x \in a \mathbin{\&} \varphi^{\boldsymbol{M}}(x, y))\}\\ &= \boldsymbol{F}[\,a\,],\end{aligned}$$

which is a set. □

8.3.5. A collection of axioms is **inconsistent** if some sentence and its negation can be proved from the axioms. Axioms not inconsistent are **consistent.** Any collection of axioms with a model (¶8.3.2) is consistent. The axioms in ZF, besides Foundation, might be called ZF^-. We believe that these axioms have the model $\mathbf{V}$; in particular, we believe that ZF^- is consistent. Since $\mathbf{WF}$ meets the conditions of the last lemma (Exercise 9), we have that $\mathbf{WF}$ is a model of ZF, so ZF is consistent. But $\mathbf{WF}$ is constructed on the assumption that $\mathbf{V}$ exists as a model of ZF^-. Briefly then, *if* ZF^- is consistent, then so is ZF itself.

8.4 Constructible sets

8.4.1. A relation on a set a is called **definable** if it is a relativization $\boldsymbol{R}^a$, where $\boldsymbol{R}$ is defined by a formula whose only constants are elements of a. Let the set of n-ary definable relations on a be denoted by

$$\mathbf{D}_n(a).$$

Then $\mathbf{D}_n(a) \subseteq \mathcal{P}(a^n)$. However, if a is infinite, then $\operatorname{card}(\mathbf{D}_n(a)) = \operatorname{card}(a)$ (Exercise 10); so not every relation on a is definable.

8.4.2. A formal definition of the definable sets can be given by recursion, paralleling the definition of formulas. First, suppose m and n are in ω, and $f\colon m \to n$. Then we obtain two related functions:

(i) Let f_* be the function

$$x \mapsto \{(u_j\colon j \in n) \in \mathbf{V}^n\colon (u_{f(i)}\colon i \in m) \in x\}$$

from $\mathcal{P}(\mathbf{V}^m)$ to $\mathcal{P}(\mathbf{V}^n)$. If a is defined by $\varphi(x_i\colon i \in m)$, then $f_*(b)$ is defined by $\varphi(x_{f(i)}\colon i \in m)$.

(ii) Let f^* be the function

$$y \mapsto \{(u_{f(i)}\colon i \in m) \in \mathbf{V}^m\colon (u_j\colon j \in n) \in y\}.$$

from $\mathcal{P}(\mathbf{V}^n)$ to $\mathcal{P}(\mathbf{V}^m)$. If b is defined by $\psi(x_j\colon j \in n)$, then $f^*(b)$ is defined by

$$\exists y_0 \cdots \exists y_{n-1}\, (\psi(y_j\colon j \in n) \mathbin{\&} x_0 = y_{f(0)} \mathbin{\&} \cdots \mathbin{\&} x_{m-1} = y_{f(m-1)}).$$

8.4.3. Now we can define the sets $\mathbf{D}_n(a)$ recursively as follows.

(i) $\{(x, y) \in a \times a\colon x \in y\} \in \mathbf{D}_2(a)$.

(ii) $b \cap a$ (that is, $\{x \in a\colon x \in b\}$) and $\{y \in a\colon b \in y\}$ are in $\mathbf{D}_1(a)$ when $b \in a$.

(iii) $a^n \smallsetminus b \in \mathbf{D}_n(a)$ when $b \in \mathbf{D}_n(a)$.

(iv) $b \cap c \in \mathbf{D}_n(a)$ when b and c are in $\mathbf{D}_n(a)$.

(v) $f_*(b) \in \mathbf{D}_n(a)$ when $b \in \mathbf{D}_m(a)$ and $f\colon m \to n$.

(vi) $f^*(c) \in \mathbf{D}_m(a)$ when $c \in \mathbf{D}_n(a)$ and $f\colon m \to n$.

8.4.4. A function $x \mapsto \mathbf{L}(x)$ on $\mathbf{ON}$ is defined recursively as follows:

(i) $\mathbf{L}(0) = \varnothing$,

(ii) $\mathbf{L}(\alpha+1) = \mathbf{D}_1(\mathbf{L}(\alpha))$,

(iii) $\mathbf{L}(\alpha) = \bigcup \mathbf{L}[\,\alpha\,]$ if α is a limit.

We now use $\mathbf{L}$ to denote $\bigcup\{\mathbf{L}(x)\colon x \in \mathbf{ON}\}$ (rather than the function $x \mapsto \mathbf{L}(x)$). Then $\mathbf{L}$ the **constructible universe,** and its elements are the **constructible sets.** As in ¶8.1.6, if $a \in \mathbf{L}$, then we can define

$$\rho(a) = \min\{x \in \mathbf{ON}\colon a \in \mathbf{L}(x+1)\}.$$

8.4.5 Lemma. $\mathbf{L}(\alpha) \in \mathbf{L}(\alpha+1)$, *and* $\mathbf{L}$ *is transitive, and* $\mathbf{L}(\alpha) \subseteq \mathbf{L}(\alpha+1)$.

Proof. Exercise 11. □

8.4.6 Theorem. ZF *is true in* $\mathbf{L}$.

Proof. Use Lemma 8.3.4. By the previous lemma, $\mathbf{L}$ is transitive. Suppose $a \subseteq \mathbf{L}$. Then $\bigcup\{\rho(x)+1\colon x \in a\}$ is an ordinal β, and $a \subseteq \mathbf{L}(\beta)$. If a and b are distinct elements of $\mathbf{L}$, then $\{a\} \in \mathbf{L}$, but $b \notin \{a\}$; this shows that $x = y$ is absolute for $\mathbf{L}$.

It remains to show that Comprehension is absolute for $\mathbf{L}$. Suppose $a \in \mathbf{L}$, so that a is a definable subset of $\mathbf{L}(\rho(a))$. If φ is an n-ary formula with constants from $\mathbf{L}$, we shall show

$$\varphi^{\mathbf{L}}(\vec{c}) \Leftrightarrow \varphi^{\mathbf{L}(\beta)}(\vec{c}) \tag{8.11}$$

for all $\vec{c}$ in $\mathbf{L}(\beta)^n$, where β is sufficiently large. In particular, we may assume $\rho(a) \leqslant \beta$, and the constants of φ are in $\mathbf{L}(\beta)$. Then a is a definable subset of $\mathbf{L}(\beta)$ by the previous lemma, so (assuming now φ is singulary), we have

$$a \cap \{x \in \mathbf{L}\colon \varphi^{\mathbf{L}}(x)\} = a \cap \{x\colon \varphi^{\mathbf{L}(\beta)}(x)\},$$

which is a definable subset of $\mathbf{L}(\beta)$; so $a \cap \{x \in \mathbf{L}\colon \varphi^{\mathbf{L}}(x)\} \in \mathbf{L}(\beta+1)$.

To prove (8.11), we use induction on the complexity of formulas. Trivially, (8.11) holds when φ is atomic. If (8.11) holds when φ is ψ or χ, then it immediately holds when φ is $\neg\psi$ or $\psi \Rightarrow \chi$. Suppose (8.11) holds when φ is $\psi(\vec{x}, y)$. Then

$$(\exists y\ \varphi(\vec{c}, y))^{\mathbf{L}(\beta)} \Rightarrow (\exists y\ \varphi(\vec{c}, y))^{\mathbf{L}} \tag{8.12}$$

when $\vec{c} \in \mathbf{L}(\beta)^n$. Conversely, for some such $\vec{c}$, if $(\exists y\ \varphi(\vec{c}, y))^{\mathbf{L}}$, then $\varphi^{\mathbf{L}}(\vec{c}, d)$ for some d in $\mathbf{L}$; but then $d \in \mathbf{L}(\beta)$ if β is large enough, so the converse of (8.12) holds. □

8.4.7 Theorem. *The Axiom of Choice and the Generalized Continuum Hypothesis are true in* $\mathbf{L}$.

Proof. I suggest the idea only. For Choice, one shows something stronger: that there is a binary formula φ such that $\varphi^{\mathbf{L}}$ defines a relation by which $\mathbf{L}$ is well-ordered. Indeed, if $\mathbf{L}(\alpha)$ is well-ordered, then the ordering can be extended to well-order $\mathbf{L}(\alpha+1)$. For the Generalized Continuum Hypothesis, one shows $\mathrm{card}(\mathbf{L}(\alpha)) = \mathrm{card}(\alpha)$ in $\mathbf{L}$, and $\mathcal{P}(\mathbf{L}(\alpha))^{\mathbf{L}} \subseteq \mathbf{L}(\mathrm{card}(\alpha)^+)$. □

Exercises

(1) Show that both the power-set of a transitive set and the union of a set of transitive sets are transitive.

(2) Find two formulas φ and ψ and a class $\boldsymbol{M}$ such that $\{\vec{x}\colon \varphi(\vec{x})\} = \{\vec{x}\colon \psi(\vec{x})\}$, but then $\{\vec{x} \in \boldsymbol{M}^n\colon \varphi^{\boldsymbol{M}}(\vec{x})\} \neq \{\vec{x} \in \boldsymbol{M}^n\colon \psi^{\boldsymbol{M}}(\vec{x})\}$.

(3) Show $\beta < \alpha \Rightarrow \mathbf{R}(\beta) \in \mathbf{R}(\alpha)$.

(4) Assuming $a \in \mathbf{WF}$ and $b \in a$, show that $b \in \mathbf{WF}$ and $\mathrm{rank}(b) < \mathrm{rank}(a)$.

(5) Show that all subsets of $\mathbf{WF}$ are elements of $\mathbf{WF}$.

(6) Using the Axiom of Choice, assuming a is transitive, but not a subset of $\mathbf{WF}$, show that there is a function f on ω such that $f(0) = a$ and $f(n+1) \in f(n) \setminus \mathbf{WF}$.

(7) Prove Theorem 8.1.8.

(8) Show that $\omega \in \boldsymbol{M}$ under the conditions of Theorem 8.3.4.

(9) Show that $\mathbf{WF}$ meets the conditions of Theorem 8.3.4.

(10) If a is infinite, show that $\mathrm{card}(\mathbf{D}_n(a)) = \mathrm{card}(a)$.

(11) Prove Lemma 8.4.5.

Bibliography

[1] Alonzo Church. *Introduction to mathematical logic. Vol. I.* Princeton University Press, Princeton, N. J., 1956.

[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] Keith Devlin. *The joy of sets*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, second edition, 1993. Fundamentals of contemporary set theory.

[5] Susanna S. Epp. *Discrete Mathematics with Applications*. PWS Publishing Company, Boston, Massachusetts, USA, 1995. 2nd edition.

[6] 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.

[7] Abraham A. Fraenkel, Yehoshua Bar-Hillel, and Azriel Levy. *Foundations of set theory*. North-Holland Publishing Co., Amsterdam, revised edition, 1973. With the collaboration of Dirk van Dalen, Studies in Logic and the Foundations of Mathematics, Vol. 67.

[8] 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.

[9] 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.

[10] Azriel Levy. *Basic set theory*. Dover Publications Inc., Mineola, NY, 2002. Reprint of the 1979 original [Springer, Berlin; MR0533962 (80k:04001)].

[11] Yiannis N. Moschovakis. *Notes on set theory*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1994.

[12] Murray et al., editors. *The Compact Edition of the Oxford English Dictionary*. Oxford University Press, 1973.

[13] Filiz Oktem. *Uygulamalı Latin Dili*. Sosyal Yayınlar, Eylül 1996.

[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] Joseph R. Shoenfield. *Mathematical logic*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1967.

[17] Robert R. Stoll. *Set theory and logic*. Dover Publications Inc., New York, 1979. Corrected reprint of the 1963 edition.

[18] Patrick Suppes. *Axiomatic set theory*. Dover Publications Inc., New York, 1972. Unabridged and corrected republication of the 1960 original with a new preface and a new section (8.4).

[19] Alfred Tarski. *Introduction to Logic and to the Methodology of Deductive Sciences*. Dover, 1995. An unabridged republication of the 9th printing, 1961, of the 1946 second, revised edition of the work originally published by Oxford University Press, New York, in 1941.

[20] Robert L. Vaught. *Set theory*. Birkhäuser Boston Inc., Boston, MA, second edition, 1995. An introduction.

Index