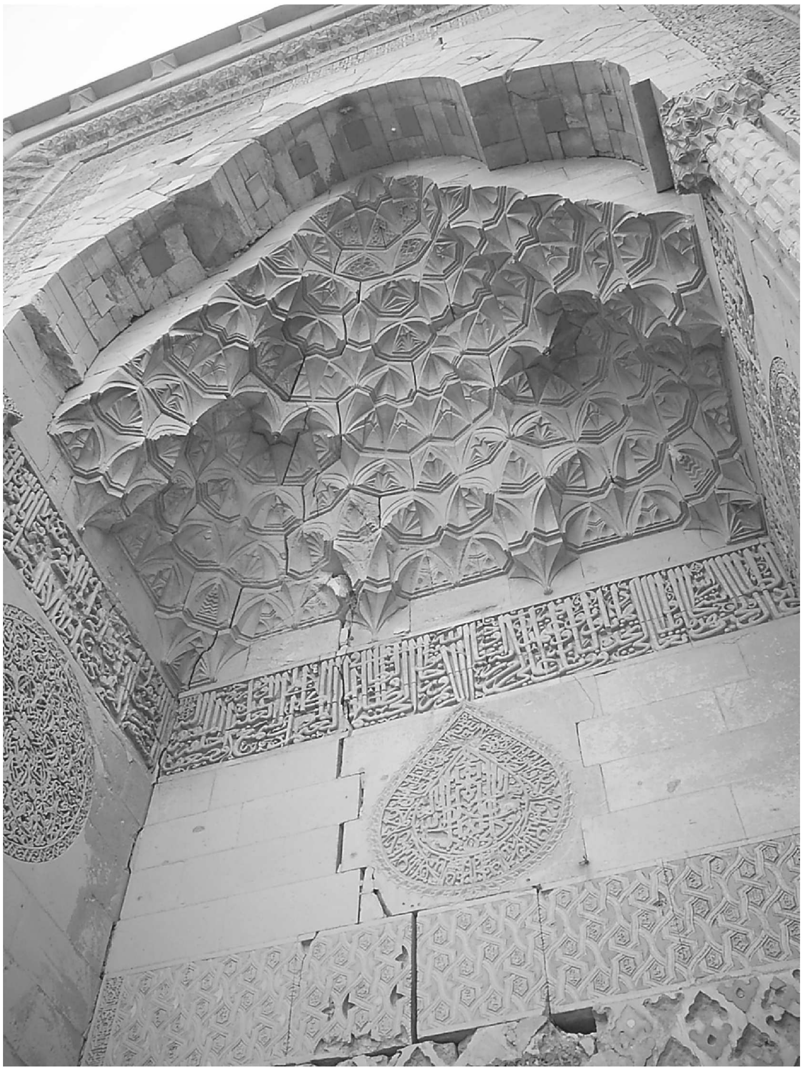
Sets, Classes, and Families

Notes on set theory

David Pierce

February 13, 2009



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 arithmetic of cardinal numbers. Axiom of choice, generalized continuum hypothesis.

The notes cover these topics and more. There have been three previous editions:

(i) In 2001, I wrote notes on ordinals and cardinals for Fundamentals of Mathematics (Math 111); a good part of what I wrote went beyond what that course had time for.
(ii) I revised those notes for use in Math 320 as taught by Ayşe Berkman in the spring semester of the 2004/5 academic year.
(iii) When I was to teach that course myself, in the spring of 2006/7, I completely rewrote the notes, using Ayşe's copy of the second edition, with her comments.

All of these editions must be considered as rough drafts. The same is true of the present edition, since it has many differences with the third edition. Starting with that third edition, I place more emphasis on the following picture of set theory:

(i) There are things called *sets*, with certain properties. Sets compose a so-called *universal class*, $\mathbf{V}$ (¶ 2.3.1).
(ii) There is a logical language for talking about sets (¶ 2.2.1); the one non-logical symbol of this language is $\in$, to express *membership* of one set in another (¶ 2.2.4). *Equality* of sets can be *defined* in terms of membership (¶ 3.1.1).
(iii) In the language of sets, a formula φ with one free variable x defines the class $\{x\colon \varphi(x)\}$ (¶ 2.3.1), which is a *subclass* of the universal class (¶ 2.3.6).
(iv) Sets are also classes (¶ 3.1.1). 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 (¶ 3.1.5).

This picture of set theory can be seen in Levy [**12**]; 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 [**22**], Kunen [**10**], and

Moschovakis [**13**]; also Enderton [**4**] for Chapter 2, and Shoenfield [**19**] and Cohen [**2**] for Ch. 8. I have used Fraenkel *et al.* [**6**] for a review of the development of the individual axioms that generate what is called Zermelo–Fraenkel set theory. Those axioms are among those listed on p. 7 below. In that list, weaker forms of the Union Axiom (called Augmentation and Binary Union respectively) are 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, weaker versions suffice.[1]

Unlike many set-theory textbooks, I aim, in Chapter 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 formalism of the logic was worked out. However, the consistency results of Chapter 8 are all about the logic.

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 Chapter 4, to see how the *class* of natural numbers can be obtained, *without* the assumption that it already exists as a set.

In studying the natural numbers, Dedekind [**3**] was clear about the distinction between *proof by induction* and *definition by recursion.* Peano [**16**] was not, though he was aware of Dedekind's work; and many writers today seem not to be clear about the distinction. In Chapters 4 and 5, I try to be clear.

Since I avoid treating classes as sets, it is useful to have a name for something that contains classes as such. I introduce the word *family* for this.

These notes are not intended for use in isolation from the classroom. Points presented here in outline may be elaborated more fully in lectures, or else simply omitted there. Lectures may take matters in a different order from the notes. I take Euclid's *Elements* [**5**] as a model. Euclid simply presents propositions and proofs, with no explanation of *why* one would want to prove these propositions. The explanation, if needed, is left to the living teacher.

The notes are divided into eight chapters; these are divided into sections (§), and these into paragraphs (¶). Some paragraphs are labelled as Axioms, Lemmas, Theorems, or (in one case) Porism. Certain displayed formulas are assigned arabic numerals in a single sequence through the notes. Terms being **defined** in the 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. All but the first chapter have exercises at the end; often the exercises ask the reader to supply the proofs of propositions (lemmas, theorems) given in the chapter.

On the front and back pages are photographs I took of Zinciriye Madrasah (built in 1385) and Mor Gabriel Monastery (founded 397), in Mardin.

[1] In the 2007 edition I remarked: '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.'

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, $\mathbf{V}$, and $\mathbf{WF}$ are the classes defined on pp. 67, 19, and 88 respectively.

(i) [p. 29] Extension:

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

(ii) [p. 31] Separation Scheme:

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

(iii) [p. 31] Pairing:

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

(iv) [p. 37] Replacement Scheme:

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

(v) [p. 42] Power Set:

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

(vi) [p. 46] Augmentation:

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

(vii) [p. 52] Binary union:

$$\exists x\; x = a \cup b.$$

(viii) [p. 68] Union:

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

(ix) [p. 70] Infinity:

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

(x) [p. 83] Choice:

Every set has a choice function.

(xi) [p. 89] Foundation:

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

The axioms of Extension, Pairing, Power Set, Union, Infinity, and Foundation, along with the axiom schemes of Separation and Replacement, are known collectively as ZF (pp. 27 and 90); but Separation is redundant in ZF (p. 39). The Augmentation and Weak Union axioms are my addition, but they are consequences of ZF. When the Axiom of Choice is added, ZF becomes ZFC. Some results on consistency and independence are worked out in Chapter 8.

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. But the set cannot be separated from its elements the way a box can be emptied of its contents. The set **comprises** its members, and 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. In ordinary speech, one may speak of a flock of pigeons, a pair of socks, a deck of cards, or a number of days. 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.*

In mathematics, the word *set* is introduced as the most abstract[1] collective noun. But now we have something new, *sets,* that can belong to something else: that something else is called a *class.* In turn, a class may belong to a *family* of classes (¶ 3.6.2). Sets are classes (¶ 3.1.1), and classes are families, but not every class is a set (¶ 3.1.5; in ¶ 8.2, we shall see how to understand a family as a class).

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.*[2] 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. A function f can be understood as the set of *ordered pairs* (a, b) such that $f(a) = b$; and the ordered pair (a, b) can be understood as the set $\{\{a\}, \{a, b\}\}$. Perhaps a and b here are *real numbers.* A real number can be understood as a certain set of *rational numbers.* A rational number can be understood as a certain set of *integers.* An integer can be understood as a certain set of *natural numbers.* A natural number can be understood as the set of its *predecessors.* There is one natural number with no predecessors; so this is the empty set.

[1] I originally called *set* the most *generally applicable* abstract noun (after *class*), but this was shown by a student, Sajjad Haider, to be a mistake: our universe of sets will not contain flocks of pigeons.

[2] We ultimately rule out this possibility with the Foundation Axiom in Ch. 8.

1.2.2. Sets form a connection between mathematics and logic, the science of reasoning. See Figure 1.1. There is, for example, 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. Indeed, if a is the set of *possible situations*[3] in which P is true, and b is the set of possible situations in which Q is true, then $a \cup b$ is the set of possible situations in which $P \vee Q$ is true.[4]

1.2.3. Sets may serve both in the definition of, and in an example of, the logical notion 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 [**5**] for organizing and developing a body of mathematical work.

1.3. Cardinals and ordinals

1.3.1. As young children, we learn to chant a sequence of numerals: *one, two, three* in English; bir, iki, üç in Turkish. We learn to use these numerals as *cardinal numbers,* to indicate the *sizes* of sets. Later we learn that *zero*/sıfır is also a size. But the sequence of numerals also has an *order,* like the order of letters in the alphabet. So we can use numerals 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.

1.3.2. So numerals have two uses, as cardinals and as ordinals; these uses are completely different; and yet the *same* underlying numerals are used in each case. For example, if a book starts on page 1 and ends on page 106, then we know two things: the book has 106 pages, and the last page of the book is the 106th page of the book.

1.3.3. Once I saw a cargo of small cardboard boxes being loaded on an airplane. Armed guards were standing by. As the boxes ascended the ramp into the cargo hold, a man counted them by writing numerals on them: 1, 2, 3, and so on. I did not see how the boxes were unloaded at the end of the flight. But 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.

1.3.4. When we count a *finite* set, then the order in which we count the elements in the set does not matter. This is a fact so basic that we do not need to learn it in school. But it fails for *infinite* sets.

1.3.5. What is a number? As suggested in ¶ 1.2.1, we can identify certain *sets* to be called 0, 1, 2, 3, and so forth. These sets will be the *natural numbers.* Each natural number n has a *successor,* which can be called $n + 1$. We *assume* that all of the natural

[3] I borrow this term from Wilfrid Hodges [**9**, § 1, p. 1], who writes: 'A set of beliefs is called *consistent* if these beliefs could all be true together in some possible situation.'

[4] However, if one is in a situation where one knows that P is true, but Q is false, then one is not likely to assert that $P \vee Q$ is true; one will probably just say that P is true and Q is false.

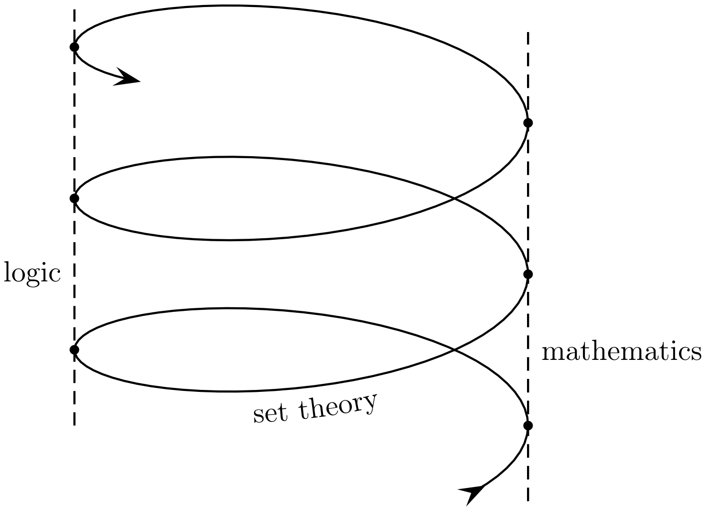


FIGURE 1.1. Logic and mathematics, connected by set theory

numbers compose a set, called[5] ω. 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 and so forth, then ω^ω, ω^{ω^ω}, and so forth. All of these new numbers are examples of *transfinite ordinal numbers.* We may be able to use these transfinite 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? The same that ω has. We can number the set alternatively so that *each* element is assigned a natural number.

1.3.6. 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 $\aleph_0$;[6] this is the first *transfinite cardinal number.* 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 transfinite cardinal numbers are *indexed* by the ordinal numbers. But while every cardinal is an ordinal, not every ordinal is a cardinal. The size of an infinite set is not obtained or defined directly by a process of counting.

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,

[5] The letter ω is not w, the so-called double u; it is the Greek minuscule *omega,* the last letter of the alphabet. See Appendix A. 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.

[6] The letter $\aleph$ is *aleph,* the first letter of the Hebrew alphabet.

> *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 Appendix A 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 defines it [**5**, 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) ’Αριθμὸς 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.[7]

1.4.2. If we take *number* seriously as a collective noun roughly equivalent to *set,* then a certain passage by Plato, in the work commonly called the *Republic,* 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 *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 [**17**] and Waterfield [**18**]. I have inserted some of the original Greek words, especially[8] 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.

[7] 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.'

[8] 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.

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

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

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

'How could it not be?'

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

'Absolutely.'

'And these things appear to lead to truth.'

'Yes, and extremely well.'

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

'Just so' he said.

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

'You speak superbly' he said.

[9] 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. For present purposes, a **logic** consists of

(i) an **alphabet** of symbols;
(ii) a way to combine the symbols of the alphabet into **formulas;**
(iii) a correspondence between the formulas and something else that we want to understand.

To understand sets, we shall develop a *predicate logic.* This will be based on a **propositional logic.**

2.1.2. The alphabet of our propositional logic will comprise

(i) **(propositional) variables** P, P', P'', $\dots$;
(ii) the **propositional connectives**[1] $\neg$ and $\Rightarrow$;
(iii) the **brackets** (and).

From the variables, connectives, and brackets, **propositional formulas** are built up **recursively** as follows.

(i) Every propositional variable is a propositional formula;
(ii) if F is a propositional formula, then so is its **negation,** $\neg F$;
(iii) If F and G are propositional formulas, then so is the **implication**[2] of G by F, namely $(F \Rightarrow G)$.

If a certain string of symbols is a propositional formula, this can be shown by means of a **tree** as in Figure 2.1. The **nodes** of the tree are the **subformulas** of the formula; these are just the formulas that are created during the construction of the formula.

2.1.3. For our convenience, we may write a formula in abbreviated form:
(i) a formula $(F \Rightarrow G)$ can be written as $F \Rightarrow G$, with outer brackets removed;
(ii) a subformula $(\cdots \Rightarrow (H \Rightarrow (G \Rightarrow F))\cdots)$ can be written as

$$(\cdots \Rightarrow H \Rightarrow G \Rightarrow F),$$

so that repeated signs of implication are applied from the right. Note then that $P \Rightarrow P'$ is *not* a subformula of $P \Rightarrow P' \Rightarrow P''$.

[1] In place of the double-shafted arrow $\Rightarrow$, the single-shafted arrow $\rightarrow$ is often used; but this might be confused with the arrow used in denoting functions (¶ 3.6.4).

[2] 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.

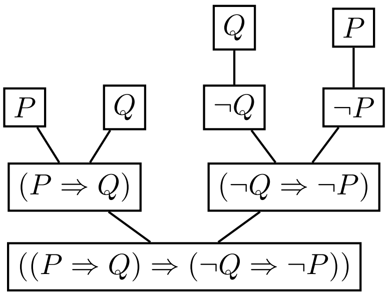

Q
P
P
Q
$\neg Q$
$\neg P$
$(P \Rightarrow Q)$
$(\neg Q \Rightarrow \neg P)$
$((P \Rightarrow Q) \Rightarrow (\neg Q \Rightarrow \neg P))$

FIGURE 2.1. A propositional formula with its tree

We may also use Q and R as variables, instead of P', P'', ...

2.1.4. It is an important fact that a given propositional formula can be constructed in only one way. Obviously a propositional variable by itself is neither a negation nor an implication; and a negation is not an implication. Obviously a particular negation takes the form $\neg F$ for some *unique* subformula F. Not so obviously, a particular implication takes the form $(F \Rightarrow G)$ for some *unique* subformulas F and G.

2.1.5. A **truth-assignment** is an assignment of a **truth-value**—true or false—to each propositional variable. Under a truth-assignment, by ¶ 2.1.4, a formula takes on a unique truth-value 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.

The truth-values of a formula under all possible truth-assignments can be given in a **truth-table.** Usually when one computes a truth-table for a formula, one includes the possible truth-values of the subformulas. A convenient way to do this is in a table where

(i) false appears as 0, and true as 1;[3]
(ii) the value of a variable is written beneath it;
(iii) the value of $\neg F$ is written beneath the $\neg$;
(iv) the value of $F \Rightarrow G$ is written beneath the $\Rightarrow$.

These rules are expressed by three tables:

P
0
1

$\neg$	F
1	0
0	1

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

A full example is worked out in stages in Figure 2.2. The formula in that table happens to be a **(propositional) tautology:** that is, it is true under every truth-assignment.

2.1.6. It is convenient to use the following abbreviations for certain formulas:

(i) $F \vee G$ stands for $\neg F \Rightarrow G$;
(ii) $F \mathbin{\&} G$ stands for $\neg(F \Rightarrow \neg G)$;

[3] Some writers, as Stoll [**21**, 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

Figure 2.2. The filling-out of a truth-table

(iii) $F \Leftrightarrow G$ stands for $\neg((F \Rightarrow G) \Rightarrow \neg(G \Rightarrow F))$;
(iv) 0 stands for $\neg(P \Rightarrow P)$;
(v) 1 stands for $P \Rightarrow P$.

In writing with these abbreviations, we may follow the convention whereby $\&$ and $\vee$ are applied before $\Rightarrow$ and $\Leftrightarrow$, so that, for example, $F \Rightarrow G \mathbin{\&} G \Rightarrow F$ stands for $F \Rightarrow (G \mathbin{\&} G) \Rightarrow F$ (which stands for $F \Rightarrow ((G \mathbin{\&} G) \Rightarrow F)$ by ¶ 2.1.3).

2.1.7. 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

$$F \sim G.$$

Hence for example

$$\neg\neg F \sim F, \qquad F \vee G \sim \neg(\neg F \mathbin{\&} \neg G), \qquad F \mathbin{\&} G \sim \neg(\neg F \vee \neg G),$$

$$F \Leftrightarrow G \sim (F \Rightarrow G) \mathbin{\&} (G \Rightarrow F), \qquad F \Leftrightarrow G \sim (F \mathbin{\&} G) \vee (\neg F \mathbin{\&} \neg G).$$

Two formulas F and G are equivalent if and only if the formula $F \Leftrightarrow G$ is a tautology.

2.1.8. Every truth-table is the truth-table of some formula, in the following sense. Suppose we are asked to find a formula F whose variables are P, Q, and R only and which takes on values as in the following table.

P	Q	R	F
0	0	0	0
1	0	0	1
0	1	0	1
1	1	0	0
0	0	1	1
1	0	1	0
0	1	1	0
1	1	1	0

Then we can let F be $(\neg R \Rightarrow G) \mathbin{\&} (R \Rightarrow H)$, where G and H are as in the following table derived from the one above.

P	Q	G	H
0	0	0	1
1	0	1	0
0	1	1	0
1	1	0	0

That is, G takes the value of F when R is 0, and H takes the value of F when R is 1. Analyzing G and H as we analyzed F, we get

$$G \sim (\neg Q \Rightarrow P) \mathbin{\&} (Q \Rightarrow \neg P), \qquad\qquad H \sim (\neg Q \Rightarrow \neg P) \mathbin{\&} (Q \Rightarrow 0),$$

so that

$$F \sim (\neg R \Rightarrow (\neg Q \Rightarrow P) \mathbin{\&} (Q \Rightarrow \neg P)) \mathbin{\&} (R \Rightarrow (\neg Q \Rightarrow \neg P) \mathbin{\&} (Q \Rightarrow 0)).$$

Alternatively, looking at where F is 1 in the original table, we obtain the **disjunctive normal form:**

$$F \sim (P \mathbin{\&} \neg Q \mathbin{\&} \neg R) \lor (\neg P \mathbin{\&} Q \mathbin{\&} \neg R) \lor (\neg P \mathbin{\&} \neg Q \mathbin{\&} R).$$

2.1.9. When we say that F is a propositional formula, 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 the language that 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]. I shall not worry further about identifying syntactical variables as such.

2.2. Predicate logic

2.2.1. Our logic for talking about sets will be a so-called **predicate logic.** Its alphabet will comprise

(i) the **binary predicate** $\in$, the sign of **membership** in a set;[4]
(ii) **(individual) variables** x, x', x'', ...;
(iii) **(individual) constants** a, a', a'', ...;
(iv) the propositional connectives $\neg$ and $\Rightarrow$;
(v) the **existential quantifier,** $\exists$;

[4] The $\in$ can be understood as a form of the Greek letter ϵ (epsilon), standing for ἐστί *is*.

(vi) the brackets (and).

We may use x, y, z, ... to stand for variables; and a, b, c, ..., for constants. 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 ¶ 2.1.2, but with an additional possibility:

(i) Atomic formulas are formulas;
(ii) if φ is a formula, then so is $\neg\varphi$;
(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 propositional connectives, as in ¶ 2.1.6.

2.2.2. The **subformulas** of a formula φ are the formulas obtained during the construction of φ. There is a recursive definition:

(i) Every formula is a subformula of itself;
(ii) φ is a subformula of $\neg\varphi$ and of $\exists x\ \varphi$;
(iii) φ and ψ are subformulas of $(\varphi \Rightarrow \psi)$;
(iv) if φ is a subformula of ψ, and ψ is a subformula of χ, then φ is a subformula of χ.

2.2.3. If φ is a formula and x is a variable, then x may or may not occur in φ. Indeed, x occurs once in $x \in y$ (if y is not x), twice in $x \in x$, and three times in $\exists x\ x \in x$. If φ is a subformula of ψ, then every occurrence of x in φ is also an occurrence in ψ. Some occurrences of x are **free,** while all others are **bound.bound variable** The definition of free occurrences is recursive:

(i) All occurrences of x in an atomic formula are free.
(ii) The free occurrences of x in $\neg\varphi$ are just those in φ.
(iii) The free occurrences of x in $\varphi \Rightarrow \psi$ are just those in φ and ψ.
(iv) If y is a variable different from x, then the free occurrences of x in $\exists y\ \varphi$ are just those in φ.
(v) There are no free occurrences of x in $\exists x\ \varphi$.

If some occurrence of a variable in a formula is free, then that variable is one of the **free variables** of the formula. 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 (if y is not x). If t is a term, then let the expression

$$(\varphi)^x_t$$

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 $(\varphi)^x_a$ is $a \in y \Rightarrow \exists x\ x \in y$. A formula with at most one free variable is a **singulary** or **unary** formula.[5] If that variable is x, and the formula is φ,

[5] 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 [**15**]—are SINGUL-, BIN-, TERN-, QUATERN-, QUIN-. The latter sequence gives us *binary, ternary, quaternary,* and *quinary.* So *singulary* appears to be a better word than *unary.* In fact, *singulary* does

then we may write φ also as

$$\varphi(x).$$

A formula with no free variables is a **sentence.** Sentences are also singulary formulas.

2.2.4. 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 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 ¶ 2.1.5.

(iii) Suppose $\exists x\ \varphi$ is a sentence, and the constant a does not appear in φ. Then $\exists x\ \varphi$ is true just in case the sentence $(\varphi)^x_a$ 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 $(\varphi)^x_a$ is $a \in a$; this will ultimately be false, by the Foundation Axiom (¶ 8.2.1); 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 $(\varphi)^x_b$ is true.

2.2.5. Suppose φ is $\exists y\ x \in y$, where y is not x. Then φ is $\varphi(x)$, that is, φ has no free variable other than x. In fact, x is free in φ. But y is not free in $(\varphi)^x_y$, since this formula is the sentence $\exists y\ y \in y$. We may say then that y is not **substitutable** for x in φ. As usual, there is a recursive definition:

(i) In every atomic formula, y is substitutable for x.

(ii) If y is substitutable for x in φ and ψ, then it is so in $\neg\varphi$ and $\varphi \Rightarrow \psi$.

(iii) If y is substitutable for x in φ, and z is not y, then y is substitutable for x in $\exists z\ \varphi$.

(iv) If x is not free in φ, then y is substitutable for x in $\exists y\ \varphi$.

If x is free in φ, then y is *not* substitutable for x in $\exists y\ \varphi$. Trivially, a variable is always substitutable for itself. A constant is always substitutable for a variable. If φ is $\varphi(x)$, and t is substitutable for x in φ, then we can write $(\varphi)^x_t$ as

$$\varphi(t).$$

If $\exists y\ \psi$ is known to be a sentence, then ψ is $\psi(y)$, so $(\psi)^y_a$ can be written as $\psi(a)$.

2.2.6. 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. These are

not appear in the original *Oxford English Dictionary* [**14**]. 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 *μονάς*, *μοναδ*- (for which see ¶ 1.4.1). The on-line *OED* (2nd ed., 1989; `http://dictionary.oed.com`, accessed January 26, 2009) does have *singulary,* for which Quine is quoted from 1940.

the **pure** or the **hereditary sets**. (So a set is hereditary if and only if all its members are hereditary.) We shall see that this is not a real limitation, mathematically speaking.

2.2.7. We may write

$$s \notin t$$

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

$$\forall x\ \varphi$$

instead of $\neg\exists x\ \neg\varphi$; here $\forall$ is the **universal quantifier.** If $\forall x\ \varphi$ 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.*

2.3. Classes

2.3.1. If any constants in a singulary formula $\varphi(x)$ have been interpreted, then the formula may be said to **define** something, namely the **class** of those sets a (assuming a does not appear in φ) such that $\varphi(a)$ is true. Then the class **comprises** such a, and the a are **members** or **elements** of the class, and these **compose** the class. The class defined by φ can be denoted by

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

or by $\{y\colon \varphi(y)\}$ if y is substitutable for x in φ. Arbitrary classes can be denoted by boldface capital letters, as $\boldsymbol{C}$, $\boldsymbol{D}$, and so on.[6] 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, $\{x\colon x \in x \Rightarrow x \in x\}$ is the same as the class $\{x\colon x \in x \Rightarrow x \in x \Rightarrow x \in x\}$, namely the class of all sets, which can be denoted by

$$\mathbf{V};$$

this is the **universal class.**

2.3.2. A class appears to be something like a set. However, we are defining classes *in terms of* sets. We should not expect all classes to be sets. At the beginning of the study of sets in the nineteenth century, no possibility of a distinction between classes and sets was recognized. This led to problems, as in Theorem 3.1.5 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;[7] 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

[6] This convention is followed by Kunen [**10**, Ch. 1, §9], for example, though not by Moschovakis [**13**, **3.19**]; Levy [**12**, I.4.1] uses plainface capital letters for classes. The handwritten version of a boldface letter is the letter with a wavy line underneath; so $\boldsymbol{C}$ can be written by hand as $\underset{\sim}{C}$.

[7] William Wordsworth:

> To me the meanest flower that blows can give
> Thoughts that do often lie too deep for tears.

—from the *Ode: Intimations of Immortality from Recollections of Early Childhood.*

full-blown; we do not (to my knowledge) have any earlier texts to tell us how the idea of erecting a mathematical theory on axioms was discovered.

2.3.3. If we let $\boldsymbol{C}$ denote a class $\{x\colon \varphi(x)\}$, and t is substitutable for x in φ, then, in any formula, we may use the expression

$$t \in \boldsymbol{C} \tag{1}$$

instead of $\varphi(t)$. If t is a variable y that is perhaps *not* substitutable for x in φ, we can still write $y \in \boldsymbol{C}$; now this means $\varphi^*(y)$, where $\varphi*$ is the result of replacing *each* occurrence of y in φ with some variable z that does not appear in φ.

2.3.4. Theorem. *If φ^* is as in ¶ 2.3.3, then y is substitible for x in φ^*, and*

$$\forall x\ (\varphi \Leftrightarrow \varphi^*). \tag{2}$$

Proof. We use induction on *singulary* formulas, in the following sense. Suppose φ is atomic. Since we assume that φ has no free variables other than x, φ must be $x \in a$ or $x \in x$ or $a \in x$ for some constant a. Then φ^* is the same, so the claim (2) is trivially true. Suppose the claim is true when φ is ψ or χ. Since $(\neg\psi)^*$ is $\neg\psi^*$, and $(\psi \Rightarrow \chi)^*$ is $\psi^* \Rightarrow \chi^*$, the claim holds when φ is $\neg\psi$ or $\psi \Rightarrow \chi$. In the next step, we should like to suppose that the claim holds when φ is a formula ρ, and then prove it when φ is $\exists u\ \rho$ for some variable u. But then ρ may have both u and x free. So our inductive hypothesis should be that (2) holds when φ is $(\rho)^u_b$ for some b. Now say φ is $\exists u\ \rho$. We may assume that u is not z. If u is not x or y, then φ^* is $\exists u\ \varphi^*$, and

$$\begin{array}{lll}
\varphi^*(a) & \text{if and only if} & (\exists u\ \rho^*)^x_a \\
 & \text{if and only if} & \exists u\ (\rho^*)^x_a \\
 & \text{if and only if} & ((\rho^*)^x_a)^u_b \text{ for some } b \\
 & \text{if and only if} & ((\rho^*)^u_b)^x_a \text{ for some } b \\
 & \text{if and only if} & (((\rho)^u_b)^*)^x_a \text{ for some } b \\
 & \text{if and only if} & ((\rho)^u_b)^x_a \text{ for some } b \quad \text{[by inductive hypothesis]} \\
 & \text{if and only if} & ((\rho)^x_a)^u_b \text{ for some } b \\
 & \text{if and only if} & \exists u\ (\rho)^x_a \\
 & \text{if and only if} & (\exists u\ \rho)^x_a \\
 & \text{if and only if} & \varphi(a),
\end{array}$$

and the claim holds. If u is y, then φ^* is $\exists z\ \rho^*$, and the argument is as before, with z for u. If u is x, then φ is already a sentence, so the inductive hypothesis has settled the matter. This completes the induction. □

2.3.5. We can now use the sentence

$$\boldsymbol{C} = \boldsymbol{D} \tag{3}$$

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

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

2.3.6. A class $\boldsymbol{C}$ is a **subclass** 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 $\forall x\ (x \in \boldsymbol{C} \Rightarrow x \in \boldsymbol{D})$. The definition of equality of classes (¶ 2.3.1) is now expressed by the sentence

$$\boldsymbol{C} = \boldsymbol{D} \Leftrightarrow \boldsymbol{C} \subseteq \boldsymbol{D} \mathbin{\&} \boldsymbol{D} \subseteq \boldsymbol{C}. \tag{4}$$

Instead of $\neg\, \boldsymbol{C} \subseteq \boldsymbol{D}$, we may write

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

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

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

and say that **$\boldsymbol{D}$ properly includes $\boldsymbol{C}$**, or that $\boldsymbol{C}$ is **properly included** in $\boldsymbol{D}$. Instead of $\neg\, \boldsymbol{C} \subset \boldsymbol{D}$, we may write

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

2.3.7. 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

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

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

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

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

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

(iv) The **empty class** is $\{x \colon x \in x \mathbin{\&} x \notin x\}$, denoted by

$$\varnothing.$$

(v) The **difference** of $\boldsymbol{C}$ from $\boldsymbol{D}$ is $\{x \colon x \in \boldsymbol{C} \mathbin{\&} x \notin \boldsymbol{D}\}$ 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

$$\boldsymbol{C} \mathbin{\triangle} \boldsymbol{D}.$$

Hence for example (Exercise 2.4)

$$\varnothing^{\mathrm{c}} = \mathbf{V}, \qquad \boldsymbol{C} = \boldsymbol{D} \Leftrightarrow \boldsymbol{C} \mathbin{\triangle} \boldsymbol{D} = \varnothing. \tag{5}$$

2.3.8. None of the operations on classes defined in ¶ 2.3.7 made any real use of *membership* of sets. Here are two that do.

(i) The **union** of a single class $\boldsymbol{C}$ is the class $\{x \colon \exists y\ (y \in \boldsymbol{C} \mathbin{\&} x \in y)\}$ 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

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

2.4. Relations

2.4.1. *Equality* and *inclusion* of classes are examples of **relations.** In particular, equality is a relation that is

(i) **reflexive,** because $\boldsymbol{C} = \boldsymbol{C}$ for all classes $\boldsymbol{C}$;
(ii) **symmetric,** because $\boldsymbol{C} = \boldsymbol{D} \Leftrightarrow \boldsymbol{D} = \boldsymbol{C}$ for all $\boldsymbol{C}$ and $\boldsymbol{D}$;
(iii) **transitive,** because $\boldsymbol{C} = \boldsymbol{D}\ \&\ \boldsymbol{D} = \boldsymbol{E} \Rightarrow \boldsymbol{C} = \boldsymbol{E}$ for all $\boldsymbol{C}$, $\boldsymbol{D}$, and $\boldsymbol{E}$.

Therefore equality is the prototypical example of an **equivalence-relation.** Inclusion, like equality, is reflexive and transitive; but it is also **anti-symmetric,** because

$$\boldsymbol{C} \subseteq \boldsymbol{D}\ \&\ \boldsymbol{D} \subseteq \boldsymbol{C} \Rightarrow \boldsymbol{C} = \boldsymbol{D}.$$

Because it is reflexive, anti-symmetric, and transitive, inclusion is called an **ordering.** *Proper* inclusion is symmetric and transitive, but also **irreflexive,** because always

$$\boldsymbol{C} \not\subset \boldsymbol{C}.$$

Still, proper inclusion is refered to as an ordering: it is a **strict** ordering.

2.4.2. 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, by writing

$$(\varphi)^{x\,y}_{s\,t},$$

we mean 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 $x \in x$. If s is substitutable for x, and t for y, in φ, then we can write $(\varphi)^{x\,y}_{s\,t}$ as

$$\varphi(s, t).$$

2.4.3. If any constants in a binary formula $\varphi(x, y)$ have been interpreted, then the formula may be said to **define** something, namely a **binary relation** between sets. If this relation is called $\boldsymbol{R}$, then in any formula, we may use the expression

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

instead of $\varphi(s, t)$, and $\neg(s\ \boldsymbol{R}\ t)$ for $\neg\varphi(s, t)$. (We may have to replace φ with φ^* as in ¶ 2.3.3.) The **domain** of $\boldsymbol{R}$ is the class $\{x\colon \exists y\ x\ \boldsymbol{R}\ y\}$; this can be denoted by

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

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

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

If $\operatorname{dom}(\boldsymbol{R}) \subseteq \boldsymbol{C}$, and $\operatorname{rng}(\boldsymbol{R}) \subseteq \boldsymbol{D}$, then $\boldsymbol{R}$ can be called a relation **from $\boldsymbol{C}$ to $\boldsymbol{D}$**. If also $\boldsymbol{C} = \boldsymbol{D}$, then $\boldsymbol{R}$ is a relation **on $\boldsymbol{C}$**.

2.4.4. A relation $\boldsymbol{R}$ is **reflexive** if $\forall x\ x\ \boldsymbol{R}\ x$; **irreflexive,** if $\forall x\ \neg(x\ \boldsymbol{R}\ x)$. The relation defined by a formula $(\varphi(x) \Leftrightarrow \varphi(y)) \lor \psi(x,y)$, for example, is reflexive; by $\neg(\varphi(x) \Leftrightarrow \varphi(y))\ \&\ \psi(x,y)$, irreflexive.

2.4.5. A binary relation $\boldsymbol{R}$ has a **converse,** which can be denoted by

$$\breve{\boldsymbol{R}};$$

it is defined by the same formula as $\boldsymbol{R}$, but with the two free variables considered in the other order. That is, if $\varphi(x,y)$ defines $\boldsymbol{R}$, then $\varphi(y,x)$ defines $\breve{\boldsymbol{R}}$. We have

$$\operatorname{dom}(\breve{\boldsymbol{R}}) = \operatorname{rng}(\boldsymbol{R}), \tag{6}$$

$$\operatorname{rng}(\breve{\boldsymbol{R}}) = \operatorname{dom}(\boldsymbol{R}), \tag{7}$$

$$\breve{\breve{\boldsymbol{R}}} = \boldsymbol{R} \tag{8}$$

(Exercise 2.7). If $\breve{\boldsymbol{R}} = \boldsymbol{R}$, then $\boldsymbol{R}$ is called **symmetric** (compare ¶ 2.4.1). So, the relation defined by $\varphi(x,y)$ is symmetric if and only if

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

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

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

is the **composite** of $\boldsymbol{R}$ and $\boldsymbol{S}$, denoted by[8]

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

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

$$(\boldsymbol{R}/\boldsymbol{S})/\boldsymbol{T} = \boldsymbol{R}/(\boldsymbol{S}/\boldsymbol{T}). \tag{9}$$

If $\boldsymbol{R}/\boldsymbol{R} \subseteq \boldsymbol{R}$, then $\boldsymbol{R}$ is called **transitive** (again, compare ¶ 2.4.1). In this case, instead of $a\ \boldsymbol{R}\ b\ \&\ b\ \boldsymbol{R}\ c$, we may write

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

The relation defined by a formula $\varphi(x) \Rightarrow \varphi(y)$ is transitive; so is the relation defined by $\forall z\ (\psi(x,z) \Rightarrow \psi(y,z))$, and so on. As in ¶ 2.4.1, a reflexive, symmetric, transitive relation is an **equivalence-relation.**

2.5. Proof

2.5.1. A **truth-assignment** in predicate logic is an assignment of truth-values to the atomic sentences. Let $\mathfrak{A}$ be such an assignment. This determines a truth-value $\sigma^{\mathfrak{A}}$ for each sentence σ by the following rules.

(i) If σ is atomic, then $\sigma^{\mathfrak{A}}$ is whatever value is assigned to σ by $\mathfrak{A}$.
(ii) $(\neg\sigma)^{\mathfrak{A}}$ is true if and only if $\sigma^{\mathfrak{A}}$ is false.
(iii) $(\sigma \Rightarrow \tau)^{\mathfrak{A}}$ is true if and only if $\sigma^{\mathfrak{A}}$ is false or $\tau^{\mathfrak{A}}$ is true.
(iv) $(\exists x\ \varphi)^{\mathfrak{A}}$ is true if and only if $\varphi(a)^{\mathfrak{A}}$ is true for some constant a.

[8] Tarski [**23**, § 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 [**22**, § 3.1, Definition 7, p. 63] also uses the notation.

If $\sigma^{\mathfrak{A}}$ is true, we may write also

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

saying σ is **true in** $\mathfrak{A}$; otherwise,

$$\mathfrak{A} \nvDash \sigma,$$

and σ is **false in** $\mathfrak{A}$. A sentence that is true in every truth-assignment is **logically true.** The proof of Theorem 2.3.4 is really a proof that (2) is logically true.

2.5.2. Suppose F is a tautology of propositional logic, and for each variable P appearing in F, for some formula φ, we replace each instance of P in F with φ. The resulting formula is a **tautology** of predicate logic. If, to an arbitrary formula, we prefix $\forall x$ for some variables x, among which are all of the free variables of the formula, then the result, a sentence, is called a **generalization** of the original formula. Generalizations of tautologies are logically true.

2.5.3. There follows a list of several kinds of formulas whose generalizations are logically true sentences; we shall refer to all of these generalizations as **logical axioms:**

(i) tautologies;
(ii) $\theta \Rightarrow \forall x\ \theta$, where x is not free in θ;
(iii) $\forall x\ (\varphi \Rightarrow \psi) \Rightarrow \forall x\ \varphi \Rightarrow \forall x\ \psi$;
(iv) $(\varphi)^x_t \Rightarrow \exists x\ \varphi$, where t is a term that is substitutable for x in φ.

2.5.4. The **logical theorems** are obtained from the logical axioms by repeated application of a **rule of inference,** called ***Modus Ponens*** in Latin and **Detachment** in English. This means:

(i) the logical axioms are logical theorems;
(ii) if σ and $\sigma \Rightarrow \tau$ are logical theorems, then so is τ.

More generally, if Γ is a list[9] of sentences, then sentences are **deducible** from Γ by the following rules:

(i) logical axioms, and sentences in Γ, are deducible from Γ;
(ii) if σ and $\sigma \Rightarrow \tau$ are deducible from Γ, then so is τ.

If σ is deducible from Γ, we may say also that Γ **entails** σ, writing

$$\Gamma \vdash \sigma.$$

As a special case, if σ is a logical theorem, we may write

$$\vdash \sigma.$$

The logical axioms, together with the rule of inference, constitute a **proof system.** In general, if $\Gamma \vdash \sigma$, this can be shown with a **formal proof,** namely a *finite*[10] list of sentences, ending with σ, such that each sentence τ on the list is a logical axiom, is in Γ, or is preceded in the list by sentences θ and $\theta \Rightarrow \tau$. In general, capital Greek letters will stand for lists of sentences (or formulas); then an expression like

$$\Gamma, \sigma$$

will stand for the list obtained from the list Γ by adding σ.

[9] I use the word *list* here mainly to avoid using *set,* although the word *list* should suggest an ordering.

[10] A definition of finite sets will come later, in ¶ 7.1.1; meanwhile, we must rely on our intuition.

2.6. Completeness

2.6.1. Theorem (Deduction). $\Gamma, \sigma \vdash \tau$ *if and only if*

$$\Gamma \vdash \sigma \Rightarrow \tau.$$

Proof. A formal proof that $\Gamma \vdash \sigma \Rightarrow \tau$ is also a formal proof that $\Gamma, \sigma \vdash \sigma \Rightarrow \tau$. If $\Delta, \sigma \Rightarrow \tau$ is such a proof, then $\Delta, \sigma \Rightarrow \tau, \sigma, \tau$ is a formal proof that $\Gamma, \sigma \vdash \tau$.

Suppose conversely that $\Gamma, \sigma \vdash \tau$. Suppose Δ is a proof of this. We convert Δ to a proof that $\Gamma \vdash \sigma \Rightarrow \tau$ by replacing the sentences in Δ one by one. Suppose θ is such a sentence.

1. If θ is a logical axiom or a sentence of Γ, then we replace θ with the list

$$\theta,\ \theta \Rightarrow \sigma \Rightarrow \theta,\ \sigma \Rightarrow \theta,$$

which is itself a formal proof that $\Gamma \vdash \sigma \Rightarrow \theta$, since $\theta \Rightarrow \sigma \Rightarrow \theta$ is a tautology.

2. If θ is σ, we replace it with the logical axiom $\sigma \Rightarrow \sigma$.

3. If θ is preceded by sentences ρ and $\rho \Rightarrow \theta$, then we replace θ with the list

$$(\sigma \Rightarrow \rho \Rightarrow \theta) \Rightarrow (\sigma \Rightarrow \rho) \Rightarrow \sigma \Rightarrow \theta,\ (\sigma \Rightarrow \rho) \Rightarrow \sigma \Rightarrow \theta,\ \sigma \Rightarrow \theta,$$

which is a formal proof that $\sigma \Rightarrow \rho, \sigma \Rightarrow \rho \Rightarrow \theta \vdash \sigma \Rightarrow \theta$.

In the end, Δ is converted to a formal proof that $\Gamma \vdash \sigma \Rightarrow \tau$. □

2.6.2. Lemma. *If y is substitutable for x in $\varphi(x)$, then $\vdash \forall y\ (\forall x\ \varphi(x) \Rightarrow \varphi(y))$.*

Proof. Here is a formal proof:

$$\begin{aligned}
&\forall y\ \big((\neg\varphi(y) \Rightarrow \exists x\ \neg\varphi(x)) \Rightarrow \forall x\ \varphi(x) \Rightarrow \varphi(y)\big) && \text{[Axiom (i)]}\\
&\forall y\ \big((\neg\varphi(y) \Rightarrow \exists x\ \neg\varphi(x)) \Rightarrow \forall x\ \varphi(x) \Rightarrow \varphi(y)\big) \Rightarrow\\
&\qquad \Rightarrow \forall y\ (\neg\varphi(y) \Rightarrow \exists x\ \neg\varphi(x)) \Rightarrow \forall y\ \big(\forall x\ \varphi(x) \Rightarrow \varphi(y)\big) && \text{[Axiom (iii)]}\\
&\forall y\ (\neg\varphi(y) \Rightarrow \exists x\ \neg\varphi(x)) \Rightarrow \forall y\ \big(\forall x\ \varphi(x) \Rightarrow \varphi(y)\big) && \text{[Detachment]}\\
&\forall y\ (\neg\varphi(y) \Rightarrow \exists x\ \neg\varphi(x)) && \text{[Axiom (iv)]}\\
&\forall y\ \big(\forall x\ \varphi(x) \Rightarrow \varphi(y)\big); && \text{[Detachment]}
\end{aligned}$$

thus $\vdash \forall y\ \big(\forall x\ \varphi(x) \Rightarrow \varphi(y)\big)$. □

2.6.3. Theorem (Generalization). *If φ is a formula $\varphi(x)$, which does not feature a, and $\Gamma \vdash \varphi(a)$, where Γ is a list of sentences that do not feature a, then*

$$\Gamma \vdash \forall x\ \varphi(x).$$

Proof. Under the given conditions, there is a formal proof Δ that $\Gamma \vdash \varphi(a)$. Let y be a variable not appearing in Δ. Each sentence in Δ can be understood as $\psi(a)$, where ψ has no variable other than y free (so ψ is $\psi(y)$), and a does not appear in ψ.

1. If $\psi(a)$ is a logical axiom, then we replace it with $\forall y\ \psi$, which is also a logical axiom.

2. If $\psi(a)$ is from Γ, then $\psi(a)$ must be ψ, a sentence, and we replace it with

$$\psi,\ \psi \Rightarrow \forall y\ \psi,\ \forall y\ \psi,$$

which is a formal proof that $\Gamma \vdash \forall y\ \psi$.

3. If $\psi(a)$ is preceded in Δ by $\chi(a)$ and $\chi(a) \Rightarrow \rho(a)$, then we replace $\psi(a)$ in Δ with

$$\forall y\ (\chi \Rightarrow \rho) \Rightarrow \forall y\ \chi \Rightarrow \forall y\ \rho,\ \forall y\ \chi \Rightarrow \forall y\ \rho,\ \forall y\ \rho,$$

which is a formal proof that $\forall y\ \chi, \forall y\ (\chi \Rightarrow \rho) \vdash \forall y\ \rho$.

In the end, Δ is converted to a formal proof Δ^* that $\Gamma \vdash \forall y\ \varphi(y)$. Note that x is substitutable for y in $\varphi(y)$. Now can we continue Δ^* with

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

where the missing steps exist by Lemma 2.6.2; thus we get a proof that $\Gamma \vdash \forall x\ \varphi(x)$. □

2.6.4. Let us denote the negation of an arbitrary logically true sentence by

$$\bot.$$

A list of sentences is **inconsistent** if it entails every sentence; otherwise it is **consistent.**

2.6.5. Lemma. *Let Γ be a list of sentences; and σ, a sentence.*

(i) Γ is inconsistent if and only if $\Gamma \vdash \bot$.
(ii) Γ is consistent if and only if each finite sublist is consistent.
(iii) $\Gamma, \neg\sigma$ is inconsistent if and only if $\Gamma \vdash \sigma$.
(iv) Γ, σ is inconsistent if and only if $\Gamma \vdash \neg\sigma$.
(v) Γ, σ is inconsistent if and only if $\Gamma, \sigma \vdash \neg\sigma$.
(vi) If Γ is consistent, then at least one of Γ, σ and $\Gamma, \neg\sigma$ is consistent.

Proof. We prove only (ii); the rest is Exercise 2.9. If Γ is inconsistent, then there is a formal proof Δ that $\Gamma \vdash \bot$. Let Γ_0 comprise the sentences from Γ that actually appear in Δ. Then Γ_0 is finite, and $\Gamma_0 \vdash \bot$. □

2.6.6. Theorem (Completeness). *Every logically true sentence is a logical theorem.*

Proof. Suppose σ is not a logical theorem. We shall find a truth-assignment $\mathfrak{A}$ such that $\mathfrak{A} \models \neg\sigma$, so that σ is not logically true.

Arrange *all* sentences in an (infinite) list Γ that starts with $\neg\sigma$ (Exercise 2.12). We recursively construct a new list Δ_θ for each sentence θ in Γ. To start with, $\Delta_{\neg\sigma}$ is just $\neg\sigma$, which is a finite list that is consistent by Lemma 2.6.5 (iii). Suppose Δ_θ is finite and consistent, and θ is followed by θ' in Γ. We consider several cases.

1. If Δ_θ, θ' is inconsistent, then $\Delta_\theta, \neg\theta'$ is consistent by Lemma 2.6.5 (vi), so we let $\Delta_{\theta'}$ be $\Delta_\theta, \neg\theta'$.

2. If Δ_θ, θ' is consistent, but θ' does *not* have the form $\exists x\ \neg\neg\varphi(x)$ for any formula φ and variable x, then we let $\Delta_{\theta'}$ be Δ_θ, θ'.

3. Suppose Δ_θ, θ' is consistent, and θ' has the form $\exists x\ \neg\neg\varphi(x)$ for some formula φ and variable x. Then we let $\Delta_{\theta'}$ be $\Delta_\theta, \theta', \varphi(a)$, where a is a constant not appearing in Δ_θ or φ. Note that this list is consistent. Indeed, suppose $\Lambda, \theta, \varphi(a)$ is inconsistent. By Lemma 2.6.5 (iv), we have

$$\Lambda, \theta \vdash \neg\varphi(a).$$

By the Generalization Theorem, 2.6.3, we have $\Lambda, \theta \vdash \neg\theta$. By Lemma 2.6.5 (v), we have that Λ, θ is inconsistent.

Now let Δ consist of the sentences belonging to the Δ_θ. Then Δ is consistent by Lemma 2.6.5 (ii) because every finite sublist of Δ is a sublist of some Δ_θ. By construction then, for each sentence τ, the list Δ features either τ or $\neg\tau$, but not both. Hence we can define $\mathfrak{A}$ so that, for every atomic sentence α, we have $\mathfrak{A} \models \alpha$ if and only if α is in Δ. We now show that $\mathfrak{A} \models \tau$ if and only if τ is in Δ, for all sentences τ.

1. The claim is true when τ is atomic, by definition of $\mathfrak{A}$.
2. If the claim is true when τ is ρ, then it is true when τ is $\neg\rho$, since ρ belongs to Δ if and only if $\neg\rho$ does not.
3. If the claim is true when τ is ρ or θ, then it is true when τ is $\rho \Rightarrow \theta$, since $\rho \Rightarrow \theta$ belongs to Δ if and only if $\neg\rho$ or θ does.
4. Suppose for some $\varphi(x)$ that the claim holds whenever ρ is $\varphi(a)$ for some constant a. If $\exists x\ \varphi(x)$ belongs to Δ, then by construction $\varphi(a)$ is in Δ for some a, so $\mathfrak{A} \models \varphi(a)$ and therefore $\mathfrak{A} \models \exists x\ \varphi(x)$. Conversely, if $\mathfrak{A} \models \exists x\ \varphi(x)$, then $\mathfrak{A} \models \varphi(a)$ for some a, so $\varphi(a)$ is in Γ and therefore $\Gamma \vdash \exists x\ \varphi(x)$, so $\exists x\ \varphi(x)$ must be in Γ.

As a special case, since $\neg\sigma$ is in Δ, we have $\mathfrak{A} \models \neg\sigma$. □

2.6.7. Porism. *For every consistent list of sentences, there is a truth-assignment in which all of the sentences are true.*

Proof. If Σ is a consistent list of sentences, then, in the preceding proof, we can start the list Γ with Σ instead of $\neg\sigma$. Maybe Σ is infinite; but the finiteness of the lists Δ_θ in the proof above serve only to ensure that there are constants not appearing in Δ_θ. All we really need is that there is an infinite list b, b', b'', ... of constants not appearing in Σ. □

2.7. Set theory

2.7.1. A consequence of Porism 2.6.7 is **Compactness.** A truth-assignment in which every sentence on a given list is true is a **model** of the list. Suppose every finite sublist of some list Γ of sentences has a model. Then those finite sublists are consistent, so Γ is consistent by Lemma 2.6.5 (ii), so it has a model. The interest for us is that if σ is true in every model of Γ, then $\Gamma \vdash \sigma$, since if Γ does not entail σ, then $\Gamma, \neg\sigma$ is consistent by Lemma 2.6.5 (iii), so by the porism, Γ has a model in which σ is false. Thus we can establish entailment without actually exhibiting a formal proof. This is what we shall do from now on.

2.7.2. We now aim to identify a list of sentences that are *true* in the sense of ¶ 2.2.4 and that entail all true sentences. We aim at this; but we must fail, by *Gödel's Incompleteness Theorem.* We settle for a list, called ZFC, of true sentences that entail everything about sets that is useful for mathematics. The sentences of ZFC are not logical axioms; they are not logically true; but they are **axioms** of **set theory.** We say that the sentences in ZFC are true. If they *are* true, then ZFC is consistent. However (also by Gödel) there is no proof that ZFC is consistent; we must accept our intuition. If our intuition were wrong, this could be shown by a formal proof of $\bot$ from ZFC. But we cannot prove our intuition *correct.*

2.7.3. We shall sometimes make claims about *all* classes. Such a claim cannot be expressed by a sentence of our predicate logic; rather, it is expressed by a **scheme** of sentences, one for each singulary formula. Some of the axioms in ZFC are really schemes of axioms.

Exercises

2.1. Draw the tree showing the construction of

$$\neg(\neg P \Rightarrow (Q \Rightarrow P)) \Rightarrow \neg(R \Rightarrow Q) \Rightarrow \neg\neg R.$$

2.2. In at least two ways, find a formula F with the following truth-table:

P	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
Q	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
R	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
S	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
F	0	1	1	0	1	0	0	0	0	1	0	1	1	1	1	0

2.3. Draw a tree showing the construction of the sentence

$$\forall x\ (\varphi \Rightarrow \exists y\ (\psi \Rightarrow \forall z\ (\chi \Rightarrow \rho)))$$

(treating φ, ψ, χ, and ρ as atomic formulas).

2.4. Prove sentences (5) (p. 21).

2.5. Determine whether

(i) $\boldsymbol{C} \subseteq \boldsymbol{D} \Rightarrow \bigcap \boldsymbol{D} \subseteq \bigcap \boldsymbol{C}$;
(ii) $\exists x\ \exists y\ \forall z\ (x \in x \Leftrightarrow z \notin y)$.

2.6. What are $\bigcap \varnothing$ and $\bigcup \varnothing$?

2.7. Verify (6), (7), and (8) on p. 23.

2.8. Using a singulary formula $\varphi(x)$, write down a formula that defines an equivalence-relation.

2.9. Prove Lemma 2.6.5.

2.10. Let σ be $\forall x\ \varphi \Rightarrow \exists x\ \varphi$, where φ is $\varphi(x)$.

(i) Prove that σ is a logical theorem.
(ii) Prove that σ is a logical theorem, without using the Completeness Theorem.
(iii) Give a formal proof of σ.

2.11. Determine whether $\varphi(a) \Rightarrow \psi(a) \vdash \forall x\ \varphi(x) \Rightarrow \forall x\ \psi(x)$.

2.12. Show that all sentences can be arranged in one infinite list, as the proof of the Completeness Theorem requires.

CHAPTER 3

Functions

3.1. Equality of sets

3.1.1. We have defined equality of classes (¶ 2.3.1), but not sets. This we now do. We introduce the formula $s = t$ as an *abbreviation* for

$$\forall x\ (s \in x \Leftrightarrow t \in x)$$

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. We now *assert* that two sets are equal if (and only if) they have the same extension, in the sense of ¶ 2.3.1. This means a set a is equal to the class $\{x \colon x \in a\}$. This assertion is the first axiom of set theory.

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

$$a = b \Leftrightarrow \forall x\ (x \in a \Leftrightarrow x \in b). \tag{10}$$

3.1.3. Since sets are classes, the notions and notations used in §§ 2.3 and 2.4 can be used for sets. A subclass that is also a set can be called a **subset.** We can rewrite (10) as

$$a = b \Leftrightarrow a \subseteq b \ \&\ b \subseteq a. \tag{11}$$

Two sets are equal just in case each one is a subset of the other.

3.1.4. Theorem. *For all formulas $\varphi(x)$,*

$$a = b \Rightarrow (\varphi(a) \Leftrightarrow \varphi(b)).$$

Proof. Exercise 3.1 □

3.1.5. 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 ¶ 3.1.3 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,

$$r \notin r, \tag{12}$$

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

$$r \in \boldsymbol{R}. \tag{13}$$

The last two conclusions—(12) and (13)—imply $\boldsymbol{R} \nsubseteq r$, so $r \neq \boldsymbol{R}$. □

3.1.6. Often Theorem 3.1.5 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 $P \mathbin{\&} \neg P$. 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 ¶ 3.1.5, we happen to learn something more than the truth of the theorem: namely that no subset of $\boldsymbol{R}$ is a member of itself.[1]

3.2. New classes

3.2.1. There is 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 **power class** of $\boldsymbol{C}$, and denote it by

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

The power class of a *set* will later be called the *power set* of the set; the power set will *be* a set, by ¶ 4.1.7; but for now, it is simply a class.

3.2.2. Having defined equality of sets, we can put two sets a and b into a class, namely the class

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

this class is commonly denoted by

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

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

$$a \cup b = \bigcup \{a, b\}; \tag{14}$$

$$a \cap b = \bigcap \{a, b\} \tag{15}$$

(Exercise 3.4). 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, the union of a set will be a set; but we shall not need to use this until after ¶ 6.2.3. Meanwhile, if $\boldsymbol{C} \subseteq \mathscr{P}(b)$, then $\bigcup \boldsymbol{C} \subseteq b$, so $\bigcup \boldsymbol{C}$ will be a set by ¶ 3.3.2 (Exercise 3.7).

3.3. New sets

3.3.1. Classes that are not sets are called **proper classes.**[2] The proof of the Russell Paradox (3.1.5) 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:

[1] The argument for this—$r \subseteq \boldsymbol{R} \mathbin{\&} r \in r \Rightarrow r \notin r$—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 $(P \Rightarrow \neg P) \Rightarrow \neg P$.

[2] A proper class need not be a proper subclass of any class; nor need a proper subclass of some class be a proper class (Exercise 3.6).

3.3.2. Axiom Scheme (Separation). *Every subclass of a set is a set: For every singulary formula* φ,

$$\exists x\ \forall y\ (y \in x \Leftrightarrow y \in a \mathbin{\&} \varphi(y)). \tag{16}$$

3.3.3. The set whose existence is expressed by (16) 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 (16), because of the Extension Axiom (¶ 3.1.2). As a first consequence of the Separation 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 3.8). 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 Separation, 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 (¶ 3.2.2) that, from two sets a and b, we can form the class denoted by $\{a, b\}$. But if any class is a set, surely this class is:

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

$$\exists x\ (a \in x \mathbin{\&} b \in x).$$

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 subclass of some set. By Separation, $\{a, b\}$ is itself a set. This set is an **(unordered) pair.** In case $a = b$, the set is a **singleton,** denoted by

$$\{a\}.$$

3.4. Relations

3.4.1. Equality and inclusion are relations between classes, but they are not classes themselves: they are just features of our logic. In ¶ 2.4.3 we defined relations between sets as being defined by binary formulas; but we did not establish such relations as classes. This we now do.

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

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

However, this is also the class determined in the same way by $\varphi(x, y) \lor \varphi(y, x)$ (Exercise 3.12).

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

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

$$(a, b),$$

such that

$$(a, b) = (c, d) \Leftrightarrow a = c \mathbin{\&} b = d. \tag{17}$$

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

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

(See Exercises 3.13 and 3.14.) Then the class $\{z \colon \exists x\ \exists y\ (z = (x, y) \mathbin{\&} \varphi(x, y))\}$ 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

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

this class is denoted by

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

and it is a subclass of $\mathscr{P}(\mathscr{P}(\boldsymbol{C} \cup \boldsymbol{D}))$ (Exercise 3.15). A relation from $\boldsymbol{C}$ to $\boldsymbol{D}$ (¶ 2.4.3) can now be understood as a subclass of $\boldsymbol{C} \times \boldsymbol{D}$.

3.5. Kinds of relations

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

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

Then

$$\Delta_{\mathrm{dom}(\boldsymbol{R})} \subseteq \boldsymbol{R}/\breve{\boldsymbol{R}}, \tag{18}$$

$$\Delta_{\mathrm{rng}(\boldsymbol{R})} \subseteq \breve{\boldsymbol{R}}/\boldsymbol{R}, \tag{19}$$

for every binary relation $\boldsymbol{R}$ (Exercise 3.16). We defined reflexive and irreflexive relations in ¶ 2.4.4; symmetric, in ¶ 2.4.5; transitive, in ¶ 2.4.6. Some alternative formulations of definitions are now possible. A binary relation $\boldsymbol{R}$ is called

(i) **reflexive,** if $\Delta_{\mathbf{V}} \subseteq \boldsymbol{R}$;
(ii) **irreflexive,** if $\boldsymbol{R} \cap \Delta_{\mathbf{V}} = \varnothing$;
(iii) **anti-symmetric,** if $\boldsymbol{R} \cap \breve{\boldsymbol{R}} \subseteq \Delta_{\mathbf{V}}$.

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 $\boldsymbol{R} \cap \breve{\boldsymbol{R}} \cap (\boldsymbol{C} \times \boldsymbol{C}) \subseteq \Delta_{\boldsymbol{C}}$
(iv) **transitive on $\boldsymbol{C}$**, if $\boldsymbol{R} \cap (\boldsymbol{C} \times \boldsymbol{C})$ is transitive.

3.5.2. 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 3.18). The diagonal $\Delta_{\boldsymbol{C}}$ is an equivalence-relation on every subclass of $\boldsymbol{C}$.

3.5.3. 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

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

If $\boldsymbol{R}$ is an ordering of $\boldsymbol{C}$, then $\boldsymbol{C}$ is **ordered** by $\boldsymbol{R}$, and we may refer to the ordered pair $(\boldsymbol{C}, \boldsymbol{R})$ as an **ordered class** or simply an **order.** Note however that if $\boldsymbol{C}$ is a proper class, then $(\boldsymbol{C}, \boldsymbol{R})$ is not literally an ordered pair in the sense of ¶ 3.4.3. Note also that we do not assume that $\boldsymbol{R}$ is a relation *on* $\boldsymbol{C}$. 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 (Exercise 3.20). 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 $\boldsymbol{R} \smallsetminus (\boldsymbol{C} \times \boldsymbol{C})$ is). A strict ordering $\boldsymbol{S}$ of $\boldsymbol{C}$ has the corresponding reflexive ordering $\boldsymbol{S} \cup \Delta_{\boldsymbol{C}}$.

3.5.4. There are standard examples. (Compare ¶ 2.4.1.)

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

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

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 $>$. **Membership** of sets (the relation $\{(x, y) \colon x \in y\}$ or $\in$) is not yet an example of anything in particular; but it will be.

3.5.5. Suppose $<$ is a strict ordering of $\boldsymbol{C}$. If a and b are in $\boldsymbol{C}$, and $a < b$, then a is **less** than b, and b is **greater** than a.[4] 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** or **minimum** 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 (Exercise 3.21); but they need not exist. With respect to a total ordering, a minimal element is a least element.

3.5.6. Suppose again that $<$ is a strict ordering of $\boldsymbol{C}$. An **initial segment** of $\boldsymbol{C}$ with respect to $<$ is a subclass $\boldsymbol{D}$ of $\boldsymbol{C}$ such that

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

[4] Likewise, if $a \leqslant b$, then a is less than *or equal to* b. This standard language brings out a point about English. If the letter b is replaced with the first-person pronoun, then we should say a is less than *I* am or is equal to *me.* The pronoun retains a distinct objective form. But a noun—which is what we are using the letter b as—has no such form.

A **proper initial segment** is an initial segment that is a proper subclass. If $a \in \boldsymbol{C}$, then the class $\{x\colon x \in \boldsymbol{C} \;\&\; x < a\}$ is the class of **predecessors** of a in $\boldsymbol{C}$ with respect to $<$; it can be denoted by

$$\operatorname{pred}(a, \boldsymbol{C}, <),$$

or more simply by $\operatorname{pred}(a)$ if there will be no ambiguity. Such a class is a particular kind of proper initial segment, called a **section.**

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

(i) $<$ is a strict total ordering of $\boldsymbol{C}$;
(ii) every section 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 subclass, with an element a, then either a is its least element, or else its least element is the least element of $\boldsymbol{D} \cap \operatorname{pred}(a, \boldsymbol{C}, <)$. If $<$ well-orders $\boldsymbol{C}$, then the order $(\boldsymbol{C}, <)$ is in particular a **well-ordered set.**

3.6. Functions

3.6.1. The various operations on classes defined in ¶¶ 2.3.7, 2.3.8, 3.2.1, 3.2.2, and 3.4.4 are examples of **functions.** The union operation in ¶ 2.3.7 is the function by which the class $\boldsymbol{C} \cup \boldsymbol{D}$ is obtained from the classes $\boldsymbol{C}$ and $\boldsymbol{D}$; the power class operation in ¶ 3.2.1 is a function converting $\boldsymbol{C}$ into $\mathscr{P}(\boldsymbol{C})$. In this sense, a function is not a set or a class; it is a feature of our logic.

3.6.2. Orderings determine functions as in ¶ 3.5.6. More generally, suppose $\boldsymbol{R}$ is a binary relation. Then we can make the definitions:

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

If $\boldsymbol{R}$ is a relation *on* $\boldsymbol{C}$ that strictly orders $\boldsymbol{C}$, then $Ra = \operatorname{pred}(a)$. In general, 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 3.24).

3.6.3. Suppose $\boldsymbol{E} \subseteq \boldsymbol{C} \times \boldsymbol{C}$ and is an equivalence-relation on $\boldsymbol{C}$. If $a \in \boldsymbol{C}$, then $a\boldsymbol{E}$ is the **equivalence-class** of a with respect to $\boldsymbol{E}$, or the $\boldsymbol{E}$**-class** of a, and a is a **representative** of this class; every other member of the class is also a representative. We may understand the $\boldsymbol{E}$-classes as composing a **family,** denoted by

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

Then $\boldsymbol{D}$ belongs to $\boldsymbol{C}/\boldsymbol{E}$ if and only if $\boldsymbol{D} = a\boldsymbol{E}$ for some a in $\boldsymbol{C}$. Since classes contain only sets, and equivalence-classes may be proper classes, the family $\boldsymbol{C}/\boldsymbol{E}$ may not be a class. However, see §8.2.

3.6.4. 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,

$$a \;\boldsymbol{F}\; b \;\&\; a \;\boldsymbol{F}\; c \Rightarrow b = c. \tag{20}$$

This is equivalent to $\breve{\boldsymbol{F}}/\boldsymbol{F} = \Delta_{\mathrm{rng}(\boldsymbol{F})}$ (Exercise 3.26). 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

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

these are abbreviations of the sentence

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

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

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

This notation is consistent with our definition of equality of sets, by which equality is an equivalence-relation (¶ 3.5.2). Indeed, now the implication (20) becomes

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

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

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

One must not confuse $\boldsymbol{F}(a)$ with $\boldsymbol{R}a$ as defined in ¶ 3.6.2.

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

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

(iii) 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 $\boldsymbol{F}$.

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

(v) 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 by

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

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

(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 by

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

or π_0; it is **projection** onto the first coordinate. Likewise, there is $(x, y) \mapsto y$ or π_1.

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

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

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

If also $\boldsymbol{H}$ is functional, and $\operatorname{rng}(\boldsymbol{G}) \subseteq \operatorname{dom}(\boldsymbol{H})$, then, as we have (9), so

$$\boldsymbol{H} \circ (\boldsymbol{G} \circ \boldsymbol{F}) = (\boldsymbol{H} \circ \boldsymbol{G}) \circ \boldsymbol{F}. \tag{21}$$

3.6.7. Suppose $\boldsymbol{F} \colon \boldsymbol{C} \to \boldsymbol{D}$, and $\boldsymbol{F}/\breve{\boldsymbol{F}} \subseteq \Delta_{\mathbf{V}}$, equivalently, $\boldsymbol{F}/\breve{\boldsymbol{F}} = \Delta_{\operatorname{dom}(\boldsymbol{F})}$ (Exercise 3.27). 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 3.28), called the **inverse** of the function; when it exists, the inverse of $\boldsymbol{F}$ is denoted

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

3.6.8. Suppose again $\boldsymbol{F} \colon \boldsymbol{C} \to \boldsymbol{D}$. If $\boldsymbol{D} = \operatorname{rng}(\boldsymbol{F})$, 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

$$\boldsymbol{F} \colon \boldsymbol{C} \rightarrowtail\mathrel{\mkern-14mu}\twoheadrightarrow \boldsymbol{D}.$$

3.6.9. 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 3.29. □

3.7. Functions from functions

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

(i) If $\boldsymbol{C}$ is a subclass of $\operatorname{dom}(\boldsymbol{F})$, then the class

$$\{y \colon \exists x\ (x \in \boldsymbol{C}\ \&\ \boldsymbol{F}(x) = y)\} \tag{22}$$

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

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

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

(ii) The class
$$\{x\colon x \in \operatorname{dom}(\boldsymbol{F}) \mathbin{\&} \boldsymbol{F}(x) \in \boldsymbol{C}\} \tag{23}$$
is the **pre-image** of $\boldsymbol{C}$ under $\boldsymbol{F}$ and can be denoted by
$$\boldsymbol{F}^{-1}[\,\boldsymbol{C}\,].$$
Pre-images of all classes exist, regardless of whether the function $\boldsymbol{F}$ itself has an inverse (¶ 3.6.7). In particular, $\operatorname{dom}(\boldsymbol{F}) = \boldsymbol{F}^{-1}[\,\boldsymbol{C}\,]$ whenever $\operatorname{rng}(\boldsymbol{F}) \subseteq \boldsymbol{C}$.

Note the great difference in form between (22) and (23). The difference is reflected in the following.

3.7.2. Theorem. *Suppose $\boldsymbol{F}$ is a function. Then*
$$\boldsymbol{F}[\,\boldsymbol{C}\cup\boldsymbol{D}\,] = \boldsymbol{F}[\,\boldsymbol{C}\,]\cup\boldsymbol{F}[\,\boldsymbol{D}\,], \tag{24}$$
$$\boldsymbol{F}[\,\boldsymbol{C}\cap\boldsymbol{D}\,] \subseteq \boldsymbol{F}[\,\boldsymbol{C}\,]\cap\boldsymbol{F}[\,\boldsymbol{D}\,], \tag{25}$$
$$\boldsymbol{F}[\,\operatorname{dom}(\boldsymbol{F}) \smallsetminus \boldsymbol{C}\,] \supseteq \operatorname{rng}(\boldsymbol{F}) \smallsetminus \boldsymbol{F}[\,\boldsymbol{C}\,] \tag{26}$$
for all subclasses $\boldsymbol{C}$ and $\boldsymbol{D}$ of $\operatorname{dom}(\boldsymbol{F})$*, and*
$$\boldsymbol{F}^{-1}[\boldsymbol{C}\cup\boldsymbol{D}\,] = \boldsymbol{F}^{-1}[\boldsymbol{C}\,]\cup\boldsymbol{F}^{-1}[\boldsymbol{D}\,], \tag{27}$$
$$\boldsymbol{F}^{-1}[\boldsymbol{C}\cap\boldsymbol{D}\,] = \boldsymbol{F}^{-1}[\boldsymbol{C}\,]\cap\boldsymbol{F}^{-1}[\boldsymbol{D}\,], \tag{28}$$
$$\boldsymbol{F}^{-1}[\,\operatorname{rng}(\boldsymbol{F}) \smallsetminus \boldsymbol{C}\,] = \operatorname{dom}(\boldsymbol{F}) \smallsetminus \boldsymbol{F}^{-1}[\boldsymbol{C}\,] \tag{29}$$
for all classes $\boldsymbol{C}$ and $\boldsymbol{D}$. Moreover, the following statements are equivalent:

(i) $\boldsymbol{F}$ is injective;
(ii) $\boldsymbol{C}\cup\boldsymbol{D} \subseteq \operatorname{dom}(\boldsymbol{F}) \Rightarrow \boldsymbol{F}[\,\boldsymbol{C}\cap\boldsymbol{D}\,] = \boldsymbol{F}[\,\boldsymbol{C}\,]\cap\boldsymbol{F}[\,\boldsymbol{D}\,]$ for all classes $\boldsymbol{C}$ and $\boldsymbol{D}$;
(iii) $\boldsymbol{C} \subseteq \operatorname{dom}(\boldsymbol{F}) \Rightarrow \boldsymbol{F}[\,\operatorname{dom}(\boldsymbol{F}) \smallsetminus \boldsymbol{C}\,] = \operatorname{rng}(\boldsymbol{F}) \smallsetminus \boldsymbol{F}[\,\boldsymbol{C}\,]$ for all classes $\boldsymbol{C}$;
(iv) $a\cup b \subseteq \operatorname{dom}(\boldsymbol{F}) \Rightarrow \boldsymbol{F}[\,a\cap b\,] = \boldsymbol{F}[\,a\,]\cap\boldsymbol{F}[\,b\,]$ for all sets a and b;
(v) $a \subseteq \operatorname{dom}(\boldsymbol{F}) \Rightarrow \boldsymbol{F}[\,\operatorname{dom}(\boldsymbol{F}) \smallsetminus a\,] = \operatorname{rng}(\boldsymbol{F}) \smallsetminus \boldsymbol{F}[\,a\,]$ for all sets a.

Proof. Exercise 3.31. □

3.7.3. If $\operatorname{dom}(\boldsymbol{F})$ is a set, then $\boldsymbol{F}^{-1}[\,\boldsymbol{E}\,]$ is a set, by Separation (¶ 3.3.2). Likewise, if $\operatorname{rng}(\boldsymbol{F})$ is a set, and $\boldsymbol{E} \subseteq \operatorname{dom}(\boldsymbol{F})$, then $\boldsymbol{F}[\,\boldsymbol{E}\,]$ is a set. Now, possibly $\operatorname{rng}(\boldsymbol{F})$ is a set, while $\operatorname{dom}(\boldsymbol{F})$ 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. Theorem.

(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 $\mathscr{P}(a)$ to $\mathscr{P}(b)$, and $y \mapsto f^{-1}[y]$ is a well-defined function from $\mathscr{P}(b)$ to $\mathscr{P}(a)$. Moreover,

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

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

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

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

Proof. Exercise 3.34. □

3.7.6. Now functions whose domains are sets may be denoted by constants like f and g. The class of functions from a to $\boldsymbol{C}$ is denoted by

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

So $^{a}\boldsymbol{C} \subseteq a \times \boldsymbol{C}$. In particular, if we let $0 = \varnothing$, $1 = \{0\}$, and $2 = \{0, 1\}$, then

$$f\colon {}^{a}2 \rightarrowtail\!\!\!\!\twoheadrightarrow \mathscr{P}(a),$$

where $f(g) = g^{-1}[\{1\}]$.

Exercises

3.1. Prove Theorem 3.1.4.

3.2. Prove carefully that $\{a\} = \{b\} \Leftrightarrow a = b$.

3.3. In ¶ 2.3.6, the sentence (4) is said to be a validity. But then (11) is just a special case. Why then do we need the Extension Axiom?

3.4. Show (14) and (15).

3.5. Find three proper classes.

3.6. Find:

(i) a proper class that is not a proper subclass of any class;
(ii) a class with a proper subclass that is not a proper class.

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

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

3.9. Determine whether $\boldsymbol{C} = \bigcup \mathscr{P}(\boldsymbol{C})$.

3.10. Give three examples of sets.

3.11. Using only $\varnothing$, $\mathbf{V}$, and the braces { and }, compute

(i) $\bigcap\{\varnothing\}$ and $\bigcup\{\varnothing\}$,

(ii) $\bigcap \mathbf{V}$ and $\bigcup \mathbf{V}$.

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

3.13. Show the definition of ordered pair in ¶ 3.4.3 causes (17) to be true.

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

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

3.16. Verify (18) and (19).

3.17. Suppose $\boldsymbol{E}$ is a reflexive binary relation such that $a \mathrel{\boldsymbol{E}} c \mathbin{\&} b \mathrel{\boldsymbol{E}} c \Rightarrow a \mathrel{\boldsymbol{E}} b$. Show that $\boldsymbol{E}$ is an equivalence-relation.

3.18. Prove that $\Delta_{\mathbf{V}}$ is an equivalence-relation using only the definition of equality in ¶ 3.1.1 (and not for example the Extension Axiom).

3.19. Show that an irreflexive transitive relation is a strict ordering.

3.20. Show that the converse of an ordering is an ordering.

3.21. Show that, if an ordered class has a least element, it is unique.

3.22. Find a set with minimal elements, but no least element, with respect to some ordering.

3.23. Is there an ordered class with a proper initial segment that is not a section?

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

3.25. If $\boldsymbol{R}$ and $\boldsymbol{S}$ are binary relations, prove that they are equal if and only if $a\boldsymbol{R} = a\boldsymbol{S}$ for all a.

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

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

3.28. Show that the converse of an injective function is a function.

3.29. Prove Theorem 3.6.9.

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

3.31. Prove Theorem 3.7.2.

3.32. 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).

3.33. Prove the Separation Scheme from the Replacement Scheme.

3.34. Prove Theorem 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

$$\boldsymbol{C} \approx \boldsymbol{D}, \tag{30}$$

and we may also say that $\boldsymbol{C}$ and $\boldsymbol{D}$ are **equipollent.** So we have a relation between classes, called **equipollence,** which is an equivalence-relation in the sense of ¶ 2.4.1. Note however that the expression in (30) does *not* generally stand for a sentence of the logic of sets (Exercise 4.1). If a *set* is equipollent with a class, then that class is also a set, by Replacement (¶ 3.7.4). We shall generally be concerned only with equipollence of sets. If $a \approx b$, then we may write also

$$\operatorname{card}(a) = \operatorname{card}(b); \tag{31}$$

this *does* stand for a sentence of the logic of sets (Exercise 4.2). Here, $\operatorname{card}(a)$ is the **cardinality** of a: for now, we can understand this as the equivalence-class $\{x \colon x \approx a\}$. (See Exercise 4.3.) However, we shall ultimately (see ¶ 7.2.1) be able to define $\operatorname{card}(a)$ as a certain *set* that is equipollent with a. In any case, 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 $\boldsymbol{D}$, and 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

$$\boldsymbol{C} \preccurlyeq \boldsymbol{D} \quad \text{if and only if} \quad \boldsymbol{C} \prec \boldsymbol{D} \quad \text{or} \quad \boldsymbol{C} \approx \boldsymbol{D}.$$

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

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

The relations $\preccurlyeq$ and $\prec$ are transitive; the former is reflexive, but the latter is irreflexive (Exercise 4.4). Thus $\prec$ induces a strict ordering on cardinalities. By the following theorem, $\preccurlyeq$ is anti-symmetric on cardinalities; so it induces a reflexive ordering on cardinalities.

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 $g\colon b \rightarrowtail a$. We want to find a bijection h from a to b. Since we have only f and g to work with, we shall look for a subset c of b such that we can define h by

$$h(x) = \begin{cases} g^{-1}(x), & \text{if } x \in g[\,c\,]; \\ f(x), & \text{if } x \in a \smallsetminus g[\,c\,]. \end{cases} \tag{32}$$

To ensure that h is injective, since f and g^{-1} are already injective, we need only have $g^{-1}[\,g[\,c\,]\,] \cap f[\,a \smallsetminus g[\,c\,]\,] = \varnothing$, that is,

$$c \cap (f[\,a\,] \smallsetminus (f \circ g)[\,c\,]) = \varnothing. \tag{33}$$

To ensure that h is surjective onto b, we need only have $g^{-1}[\,g[\,c\,]\,] \cup f[\,a \smallsetminus g[\,c\,]\,] = b$, that is,

$$c \cup (f[\,a\,] \smallsetminus (f \circ g)[\,c\,]) = b. \tag{34}$$

These two conditions on c, given in (33) and (34), are equivalent to the conditions

$$c \subseteq b, \qquad\qquad b \smallsetminus c = f[\,a\,] \smallsetminus (f \circ g)[\,c\,]$$

and hence to

$$c \subseteq b, \qquad\qquad (b \smallsetminus f[\,a\,]) \cup (f \circ g)[\,c\,] = c. \tag{35}$$

Now let $\boldsymbol{D}$ be the class

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

Then $\boldsymbol{D}$ contains b. In particular, $\boldsymbol{D}$ is non-empty, so (by ¶ 3.3.3) its intersection is a set. If $c = \bigcap \boldsymbol{D}$, then (35) holds (Exercise 4.5). □

4.1.4. 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:

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

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.

4.1.5. THEOREM (Cantor). *The power class of a set is larger than the set itself:*

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

[1] The theorem is also called the Cantor–Bernstein Theorem, as for example by Levy [**12**, III.2.8, p. 85], who nonetheless observes that Dedekind gave the first proof in 1887. The proof given here is due to Zermelo [**26**, p. 208].

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

$$c \in b \Rightarrow c \in b \smallsetminus f(c), \tag{36}$$

$$c \in a \smallsetminus b \Rightarrow c \in f(c) \smallsetminus b \tag{37}$$

(Exercise 4.7). Thus, if $c \in a$, then $f(c) \neq b$. So $b \notin \operatorname{rng}(f)$. Therefore, there is no bijection from a to $\mathscr{P}(a)$; so $a \prec \mathscr{P}(a)$. □

4.1.6. Note well how the preceding proof requires b to be a set (Exercise 4.8). If a were a proper class, then the proof would fail. In particular, we cannot yet address the cardinality of $\mathscr{P}(\mathscr{P}(a))$, since we have not established that $\mathscr{P}(a)$ is a set. This is what the following axiom is for.

4.1.7. Axiom (Power Set). *The power class of a set is a set: that is,*

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

4.1.8. As foretold in ¶ 2.3.8, we may now refer to the power class of a set as its **power set.** We can form *chains:*

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

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

4.2.1. A class is **infinite** if it is equipollent with a proper subclass of itself. Let us first observe that there *is* an infinite class:

4.2.2. Theorem. *The universal class* $\mathbf{V}$ *is infinite.*

Proof. By the Power Set Axiom (¶ 4.1.7), we have a function $x \mapsto \mathscr{P}(x)$ 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 $a \smallsetminus b \in \mathscr{P}(a) \smallsetminus \mathscr{P}(b)$ and hence $\mathscr{P}(a) \neq \mathscr{P}(b)$ (Exercise 4.9). Also, every power set contains $\varnothing$, but $\varnothing$ contains nothing, so $\varnothing$ is not a power set; thus, $x \mapsto \mathscr{P}(x)$ is not surjective onto $\mathbf{V}$. So $\mathbf{V}$ is infinite, by definition. □

4.2.3. By the Russell Paradox (¶ 3.1.5), $\mathbf{V}$ has a subclass that is not a set; so $\mathbf{V}$ itself is not a set, by Separation (¶ 3.3.2). Is there an infinite *set?* We cannot now *prove* that there is an infinite set, or that there is not. In ¶ 6.3.7, we shall *assume* that there is an infinite set; but even without this assumption, we shall be able to find an infinite *class* that we can define as the class of *natural numbers.*

4.2.4. Indeed, suppose $\boldsymbol{C}$ is an infinite class. This means there is an injection $\boldsymbol{F}$ from $\boldsymbol{C}$ into itself, and there is an element i of $\boldsymbol{C}$ that is not in $\boldsymbol{F}[\boldsymbol{C}]$. It appears we can form the list

$$i,\ \boldsymbol{F}(i),\ \boldsymbol{F}(\boldsymbol{F}(i)),\ \boldsymbol{F}(\boldsymbol{F}(\boldsymbol{F}(i))),\ \ldots \tag{38}$$

whose entries are distinct elements of $\boldsymbol{C}$. This list would appear to have the essential properties of the natural numbers. However, it is not yet clear that the members of the list compose a *subclass* of $\boldsymbol{C}$.

4.2.5. Suppose all we know is that $\boldsymbol{F}$ is a function, and $\boldsymbol{C}$ is a subclass of $\operatorname{dom}(\boldsymbol{F})$ such that $\boldsymbol{F}[\boldsymbol{C}] \subseteq \boldsymbol{C}$. Then $\boldsymbol{C}$ is **closed under**closed under an — $\boldsymbol{F}$. Suppose also $i \in \boldsymbol{C}$. Then we can still form a list as in (38), although possibly not all of the entries are distinct. Indeed, in the simplest example, $\boldsymbol{C} = \{\varnothing\}$, and $\boldsymbol{F}$ is the identity on this set, and i is the unique element $\varnothing$ of the set. Then the list (38) is just $\varnothing, \varnothing, \varnothing, \varnothing, \dots$ In general, there would appear to be the three kinds of possibilities depicted in Figure 4.1, namely (i) no entry in (38) is repeated, (ii) some entries are repeated, but not i, or (iii) every entry is repeated. In any case, let us refer to a class $\boldsymbol{C}$, with element i, and closed under $\boldsymbol{F}$, as an **iterative structure.** I choose this term[2] simply because we can start with i and apply $\boldsymbol{F}$ over and over, while staying within $\boldsymbol{C}$. (Structures in general will be defined in §5.1.) We may denote this iterative structure by

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

It is not clear whether this symbolism denotes a class; but nor does it really matter. An iterative structure is simply a class 'equipped with' a distinquished element and a function under which the class is closed; we write these three things together, to indicate that we are thinking of them together. For example, from $\{\varnothing\}$ we obtained the iterative structure $(\{\varnothing\}, \varnothing, \mathrm{id}_{\{\varnothing\}})$; in the proof of Theorem 4.2.2, we used the iterative structure $(\mathbf{V}, \varnothing, x \mapsto \mathscr{P}(x))$. In general, if $(\boldsymbol{C}, i, \boldsymbol{F})$ is an iterative structure, then we may call i the **initial element** of the structure, and $\boldsymbol{F}$ the operation of **succession;** then $\boldsymbol{F}(a)$ is the **successor** of a, whenever $a \in \boldsymbol{C}$.

4.2.6. Given a set i belonging to the domain of a function $\boldsymbol{F}$, we may consider the **family** of classes that contain i and are closed under $\boldsymbol{F}$. (Compare ¶ 3.6.3.) This family (like all families of classes) is ordered by inclusion (¶ 2.4.1). The notions of **least** and **minimal** member, defined in ¶ 3.5.5, can be adapted to this ordered family. A *minimal* member $\boldsymbol{C}$ of the family can be said to admit **proof by (ordinary) induction;** more precisely, the corresponding iterative structure $(\boldsymbol{C}, i, \boldsymbol{F})$ admits induction. This means the following. Suppose that $\boldsymbol{D} \subseteq \boldsymbol{C}$ and that we can establish two claims, namely

(i) $i \in \boldsymbol{D}$ (the **base** of the induction); and
(ii) if $a \in \boldsymbol{D}$ (the **inductive hypothesis**), then $\boldsymbol{F}(a) \in \boldsymbol{D}$.

Then $\boldsymbol{D}$ belongs to the family of which $\boldsymbol{C}$ is a minimal member, and so we have proved, by induction, that $\boldsymbol{D} = \boldsymbol{C}$. A basic lemma proved by induction is the following.

4.2.7. LEMMA. *Suppose the iterative structure $(\boldsymbol{C}, i, \boldsymbol{F})$ admits induction. Then*

$$\boldsymbol{C} = \{i\} \cup \boldsymbol{F}[\boldsymbol{C}].$$

PROOF. The subclass $\{i\} \cup \boldsymbol{F}[\boldsymbol{C}]$ of $\boldsymbol{C}$ contains i, and if it contains an element a of $\boldsymbol{C}$, then it contains $\boldsymbol{F}(a)$; therefore it *is* $\boldsymbol{C}$. □

4.2.8. For $\boldsymbol{C}$ to admit proof by induction, it is sufficient that $\boldsymbol{C}$ be the *least* member of the family of classes that contain the set i and are closed under the function $\boldsymbol{F}$. In this

[2] Stoll [**21**] uses the term *unary system,* though he assumes $\boldsymbol{C}$ is a set.

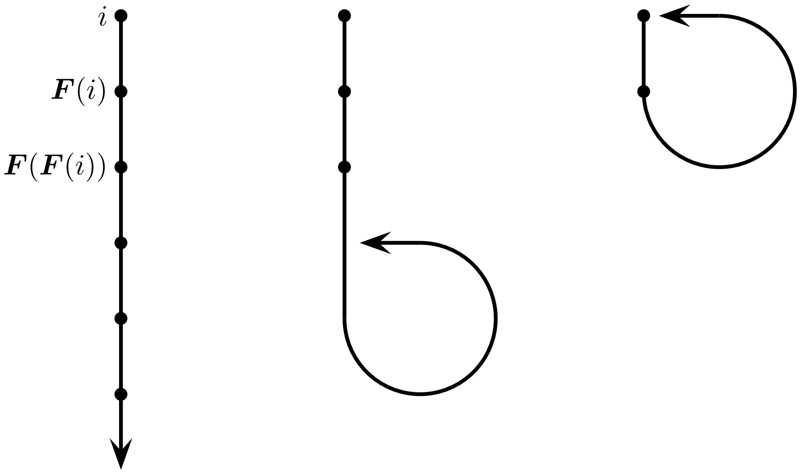


Figure 4.1. Some iterative structures

case, the iterative structure $(\boldsymbol{C}, i, \boldsymbol{F})$ can be said to be **recursive;** and $\boldsymbol{C}$ is **recursively defined,** or defined by **(ordinary) recursion.** This means that

(i) $\boldsymbol{C}$ is a class $\boldsymbol{D}$ such that
  (a) $i \in \boldsymbol{D}$, and
  (b) $a \in \boldsymbol{D} \Rightarrow \boldsymbol{F}(a) \in \boldsymbol{D}$; and
(ii) whenever $\boldsymbol{D}$ is such a class, then $\boldsymbol{C} \subseteq \boldsymbol{D}$.

Then $\boldsymbol{C}$ can be denoted suggestively by

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

Note well that possibly $\boldsymbol{F} \restriction \boldsymbol{C}$ is not injective, and possibly $i \in \boldsymbol{F}[\boldsymbol{C}]$. (However, these possibilities seem to be mutually exclusive.)

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

Proof. The 'only-if' part was implicit in ¶ 4.2.8: *Least* elements of a class (with respect to some ordering) are minimal, and therefore recursive structures admit induction. Now suppose $(\boldsymbol{C}, i, \boldsymbol{F})$ is *not* recursive. Then there is a class $\boldsymbol{D}$ containing i and closed under $\boldsymbol{F}$ of which $\boldsymbol{C}$ is not a subclass. Then $\boldsymbol{C} \cap \boldsymbol{D}$ is a proper subclass of $\boldsymbol{C}$ that contains i and is closed under $\boldsymbol{F}$ (by Theorem 3.7.2). Hence $(\boldsymbol{C}, i, \boldsymbol{F})$ does not admit induction. □

4.3. Ordinary recursion

4.3.1. Suppose $(\boldsymbol{C}, i, \boldsymbol{F})$ is a recursive structure, and $(\boldsymbol{D}, j, \boldsymbol{G})$ is another iterative structure (not necessarily recursive). There *may be* a function $\boldsymbol{H}$ from $\boldsymbol{C}$ to $\boldsymbol{D}$ satisfying two conditions:

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

These conditions can be depicted as in Figure 4.2. The first rule says what $\boldsymbol{H}(i)$ is; the second says how to obtain $\boldsymbol{H}(\boldsymbol{F}(a))$ from $\boldsymbol{H}(a)$. If such a function $\boldsymbol{H}$ does exist, then, since $(\boldsymbol{C}, i, \boldsymbol{F})$ admits induction by Lemma 4.2.9, we can use induction to prove

$$\begin{tikzcd}
\{i\} \arrow[r, "\mathrm{id}_{\{i\}}"] \arrow[d, "\boldsymbol{H}\restriction\{i\}"'] & \boldsymbol{C} \arrow[r, "\boldsymbol{F}"] \arrow[d, "\boldsymbol{H}"] & \boldsymbol{C} \arrow[d, "\boldsymbol{H}"] \\
\{j\} \arrow[r, "\mathrm{id}_{\{j\}}"'] & \boldsymbol{D} \arrow[r, "\boldsymbol{G}"'] & \boldsymbol{D}
\end{tikzcd}$$

FIGURE 4.2. Here is a *commutative diagram* showing the definition of a function $\boldsymbol{H}$ by ordinary recursion. Paths through the diagram that follow the arrows represent compositions of functions. If two paths start from the same point and end at the same point, then the corresponding compositions of functions are equal: this is why the diagram is called *commutative.*

that $\boldsymbol{H}$ is uniquely determined by these rules. We may then say that $\boldsymbol{H}$ is **recursively defined** by the given rules; also, $\boldsymbol{H}$ is a **homomorphism** from $(\boldsymbol{C}, i, \boldsymbol{F})$ to $(\boldsymbol{D}, j, \boldsymbol{G})$. See Corollary 4.3.4 below.

4.3.2. Note first another possible kind of recursive definition. If $(\boldsymbol{C}, i, \boldsymbol{F})$ is recursive, $\boldsymbol{E} \subseteq \boldsymbol{D}$, and $\boldsymbol{G} : \boldsymbol{D} \to \boldsymbol{D}$, then perhaps there is a subclass $\boldsymbol{R}$ of $\boldsymbol{C} \times \boldsymbol{D}$ such that (in the notation of ¶ 3.6.2)

(i) $i\boldsymbol{R} = \boldsymbol{E}$,
(ii) $a \in \boldsymbol{C} \Rightarrow \boldsymbol{F}(a)\boldsymbol{R} = \boldsymbol{G}[\,a\boldsymbol{R}\,]$.

(Note that $\boldsymbol{F}(a)\boldsymbol{R}$ means $b\boldsymbol{R}$, where $b = \boldsymbol{F}(a)$.) Then $\boldsymbol{R}$ too is uniquely determined by these rules, so it too is **recursively defined,** by the following.

4.3.3. THEOREM. *Suppose $(\boldsymbol{C}, i, \boldsymbol{F})$ is recursive, $\boldsymbol{E} \subseteq \boldsymbol{D}$, and $\boldsymbol{G} : \boldsymbol{D} \to \boldsymbol{D}$. Then there is at most one relation $\boldsymbol{R}$ as in ¶ 4.3.2.*

PROOF. Suppose $\boldsymbol{R}_0$ and $\boldsymbol{R}_1$ are two such relations. Let

$$\boldsymbol{C}_1 = \{x \colon x \in \boldsymbol{C} \mathbin{\&} x\boldsymbol{R}_0 = x\boldsymbol{R}_1\}.$$

Since $i\boldsymbol{R}_0 = \boldsymbol{E} = i\boldsymbol{R}_1$, we have $i \in \boldsymbol{C}_1$. Suppose $a \in \boldsymbol{C}_1$, so that $a \in \boldsymbol{C}$ and $a\boldsymbol{R}_0 = a\boldsymbol{R}_1$. Then

$$\boldsymbol{F}(a)\boldsymbol{R}_0 = \boldsymbol{G}[\,a\boldsymbol{R}_0\,] = \boldsymbol{G}[\,a\boldsymbol{R}_1\,] = \boldsymbol{F}(a)\boldsymbol{R}_1,$$

so $\boldsymbol{F}(a) \in \boldsymbol{C}_1$. By (Lemma 4.2.9 and) induction, $\boldsymbol{C}_1 = \boldsymbol{C}$. Since $\operatorname{dom}(\boldsymbol{R}_0) \subseteq \boldsymbol{C}$ and $\operatorname{dom}(\boldsymbol{R}_1) \subseteq \boldsymbol{C}$, we conclude (by Exercise 3.25) that $\boldsymbol{R}_0 = \boldsymbol{R}_1$. □

4.3.4. COROLLARY. *Suppose $(\boldsymbol{C}, i, \boldsymbol{F})$ is recursive, and $(\boldsymbol{D}, j, \boldsymbol{G})$ is iterative. Then there is at most one homomorphism from $(\boldsymbol{C}, i, \boldsymbol{F})$ to $(\boldsymbol{D}, j, \boldsymbol{G})$.*

PROOF. Supposing such a homomorphism $\boldsymbol{H}$ exists, it is a relation, namely a subclass $\boldsymbol{R}$ of $\boldsymbol{C} \times \boldsymbol{D}$. Let $\boldsymbol{E} = \{\boldsymbol{H}(i)\}$. Then

(i) $i\boldsymbol{R} = \{\boldsymbol{H}(i)\} = \boldsymbol{E}$;
(ii) $a \in \boldsymbol{C} \Rightarrow \boldsymbol{F}(a)\boldsymbol{R} = \{\boldsymbol{H}(\boldsymbol{F}(a))\} = \{\boldsymbol{G}(\boldsymbol{H}(a))\} = \boldsymbol{G}[\,\{\boldsymbol{H}(a)\}\,] = \boldsymbol{G}[\,a\boldsymbol{R}\,]$.

By the theorem, $\boldsymbol{R}$ is unique, so $\boldsymbol{H}$ is unique. □

4.4. Countably infinite classes

4.4.1. Suppose $\boldsymbol{C}$ is infinite. By ¶ 4.2.1, this means:

(i) there is an element i of $\boldsymbol{C}$, and
(ii) there is a function $\boldsymbol{F}$ under which $\boldsymbol{C}$ is closed, such that
(iii) $i \notin \boldsymbol{F}[\boldsymbol{C}]$, and
(iv) $\boldsymbol{F} \upharpoonright \boldsymbol{C}$ is injective.

Let us refer then to the iterative structure $(\boldsymbol{C}, i, \boldsymbol{F})$ as an **infinitary structure.** Suppose also

(v) the infinitary structure $(\boldsymbol{C}, i, \boldsymbol{F})$ admits induction (or is recursive).

Then $\boldsymbol{C}$ is called **countably infinite.** I propose to refer to $(\boldsymbol{C}, i, \boldsymbol{F})$ as an **arithmetic structure.**[3] The five numbered conditions here are sometimes referred to as the **Peano axioms.**[4] Does any class $\boldsymbol{C}$ meet these conditions? The following gives a sufficient condition.

4.4.2. Theorem. *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 $a \smallsetminus \{i\}$. 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 4.12) and is a set a_0. Then (a_0, i, f) is recursive, so a_0 is countably infinite. □

4.4.3. 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 (Exercise 4.13); the Binary Union Axiom (¶ 4.7.2) will be a generalization.

4.4.4. Axiom (Augmentation). *The union of a set and a singleton is a set:*

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

4.4.5. Theorem (Recursion). *Suppose* $(\boldsymbol{C}, i, \boldsymbol{F})$ *is arithmetic,* $\boldsymbol{E} \subseteq \boldsymbol{D}$*, and* $\boldsymbol{G} \colon \boldsymbol{D} \to \boldsymbol{D}$*. Then there is (uniquely, by Theorem 4.3.3) one recursively defined relation* $\boldsymbol{R}$ *as in ¶ 4.3.2.*

[3] Stoll [**21**] uses the term *integral system,* assuming $\boldsymbol{C}$ is a set.

[4] In § 1 of Chapter 1 of the *Foundations of Analysis* [**11**], Landau formulates conditions like these as Axioms 1–5 (respectively) for the *natural numbers;* he refers to these axioms as Peano's axioms in his 'Preface for the Teacher'. However, for Peano [**16**], the five conditions are numbered 1, 6, 8, 7, and 9 in a list of *nine* axioms for the positive integers. (The other four of these axioms are purely logical.) A year earlier, Dedekind [**3**, II.VI, (71), p. 67] defines the natural numbers as the elements of a set N on which a function φ is defined, and which has an element 1, such that

(i) $\varphi[N] \subseteq N$;
(ii) N is the intersection of the class of all sets that contain 1 and are closed under φ [however, Dedekind does not have the distinction between sets and classes];
(iii) $1 \notin \varphi[N]$;
(iv) φ is injective.

Proof. We want $\boldsymbol{R}$ to be a relation from $\boldsymbol{C}$ to $\boldsymbol{D}$ such that

$$i \mathrel{\boldsymbol{R}} a \Leftrightarrow a \in \boldsymbol{E}, \tag{39}$$

$$a \mathrel{\boldsymbol{R}} b \Rightarrow \boldsymbol{F}(a) \mathrel{\boldsymbol{R}} \boldsymbol{G}(b), \tag{40}$$

$$\boldsymbol{F}(a) \mathrel{\boldsymbol{R}} c \Rightarrow \exists x\ (a \mathrel{\boldsymbol{R}} x \mathbin{\&} \boldsymbol{G}(x) = c). \tag{41}$$

To this end, we define $\boldsymbol{R}$ as the union of the class comprising every subset of $\boldsymbol{C} \times \boldsymbol{D}$ whose elements can be listed as

$$(i, d),\ (\boldsymbol{F}(i), \boldsymbol{G}(d)),\ (\boldsymbol{F}(\boldsymbol{F}(i)),\ \boldsymbol{G}(\boldsymbol{G}(d))),\ \dots$$

for some d in $\boldsymbol{E}$. To be precise, let $\boldsymbol{B}$ be the class of sets a whose every element is either (i, d) for some d in $\boldsymbol{E}$ or else $(\boldsymbol{F}(b), \boldsymbol{G}(c))$ for some (b, c) in $a \cap (\boldsymbol{C} \times \boldsymbol{D})$. That is, $\boldsymbol{B}$ is the class

$$\{x \colon \forall y\ (y \in x \Rightarrow \varphi(y) \vee \psi(x, y))\},$$

where $\varphi(y)$ is the formula

$$\exists z\ (z \in \boldsymbol{E} \mathbin{\&} y = (i, z)),$$

and $\psi(x, y)$ is the formula

$$\exists u\ \exists v\ \bigl((u, v) \in x \cap (\boldsymbol{C} \times \boldsymbol{D}) \mathbin{\&} y = (\boldsymbol{F}(u), \boldsymbol{G}(v))\bigr).$$

Then $\boldsymbol{B} \subseteq \mathscr{P}(\boldsymbol{C} \times \boldsymbol{D})$, by induction. Let $\boldsymbol{R} = \bigcup \boldsymbol{B}$. If $i \mathrel{\boldsymbol{R}} a$, then $(i, a) \in b$ for some b in $\boldsymbol{B}$; but then $\neg\psi(b, (i, a))$ (since $i \notin \boldsymbol{F}[\boldsymbol{C}]$, since $(\boldsymbol{C}, i, \boldsymbol{F})$ is arithmetic); so $\varphi((i, a))$, and therefore $a \in \boldsymbol{E}$. Conversely, if $a \in \boldsymbol{E}$, then $\varphi((i, a))$, so $\{(i, a)\} \in \boldsymbol{B}$, so $i \mathrel{\boldsymbol{R}} a$. This establishes (39).

Suppose $a \mathrel{\boldsymbol{R}} b$. Then $(a, b) \in c$ for some c in $\boldsymbol{B}$, and then $\psi(c, (\boldsymbol{F}(a), \boldsymbol{G}(b)))$, so $c \cup \{(\boldsymbol{F}(a), \boldsymbol{G}(b))\} \in \boldsymbol{B}$ (here we use the Augmentation Axiom). Therefore $\boldsymbol{F}(a) \mathrel{\boldsymbol{R}} \boldsymbol{G}(b)$. Thus, (40).

Suppose finally $\boldsymbol{F}(a) \mathrel{\boldsymbol{R}} c$ for some c. Then $(\boldsymbol{F}(a), c) \in d$ for some d in $\boldsymbol{B}$, so $\psi(d, (\boldsymbol{F}(a), c))$ (again since $i \notin \boldsymbol{F}[\boldsymbol{C}]$). Hence $(\boldsymbol{F}(a), c) = (\boldsymbol{F}(e), \boldsymbol{G}(k))$ for some (e, k) in d. Since $\boldsymbol{F}$ is injective on $\boldsymbol{C}$, we have $a = e$, so $(a, k) \in d$ and $a \mathrel{\boldsymbol{R}} k$. Thus, (41), and $\boldsymbol{R}$ is as desired. □

4.4.6. Corollary (Recursion). *Suppose $(\boldsymbol{C}, i, \boldsymbol{F})$ is an arithmetic, and $(\boldsymbol{D}, j, \boldsymbol{G})$ an iterative, structure. Then there is (uniquely, by Corollary 4.3.4) a homomorphism from $\boldsymbol{C}$ to $\boldsymbol{D}$.*

Proof. Exercise 4.14. □

4.4.7. An **isomorphism** of iterative structures is a bijective homomorphism. If an isomorphism does exist, then its inverse is also an isomorphism (Exercise 4.15), and the two structures are **isomorphic.**

4.4.8. Theorem. *All countably infinite classes are equipollent; indeed, all arithmetic structures are isomorphic.*

Proof. Suppose $(\boldsymbol{C}, i, \boldsymbol{F})$ and $(\boldsymbol{D}, j, \boldsymbol{G})$ are arithmetic structures. By the corollary (4.4.6) to the Recursion Theorem, there are a homomorphism $\boldsymbol{H}$ from $(\boldsymbol{C}, i, \boldsymbol{F})$ to $(\boldsymbol{D}, j, \boldsymbol{G})$ and a homomorphism $\boldsymbol{K}$ from $(\boldsymbol{D}, j, \boldsymbol{G})$ to $(\boldsymbol{C}, i, \boldsymbol{F})$. 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 (21), 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 homomorphism from $(\boldsymbol{C},i,\boldsymbol{F})$ to itself. But so is $\mathrm{id}_{\boldsymbol{C}}$. Therefore (¶ 4.3.4) $\boldsymbol{K}\circ\boldsymbol{H}=\mathrm{id}_{\boldsymbol{C}}$. Likewise $\boldsymbol{H}\circ\boldsymbol{K}=\mathrm{id}_{\boldsymbol{D}}$. Therefore, by Theorem 3.6.9, $\boldsymbol{H}$ is a bijection from $\boldsymbol{C}$ to $\boldsymbol{D}$. □

4.5. Are there countably infinite classes?

4.5.1. There are infinite classes, by Theorem 4.2.2. We have defined countably infinite classes (¶ 4.4.1), and we have shown that all of them have the same cardinality (¶ 4.4.8). We have shown that, if there is an infinite *set,* then there is a countably infinite set (¶ 4.4.2). We have *not* shown that a countably infinite *class* exists, much less a set.

4.5.2. If we believe that there are infinite sets, then we can *postulate* their existence as an axiom: the *Axiom of Infinity.* But this is a delicate matter, since we know from the Russell Paradox that some infinite classes *cannot* be sets. Again, by Theorem 4.4.2, if there is an infinite set at all, then there must be a *countably* infinite set. So safest form of the Axiom of Infinity is that there is a countably infinite set. Then we can obtain larger sets by the Power Set Axiom and Cantor's Theorem.

4.5.3. However, all of our axioms so far, except for Extension, are assertions that certain classes are sets. One way to write the Axiom of Infinity in the same style is as follows. Pick an infinitary structure $(\boldsymbol{C},i,\boldsymbol{F})$. For example, we might pick $(\mathbf{V},\varnothing,x\mapsto\mathscr{P}(x))$. Let $\boldsymbol{D}$ be the class

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

(as in the proof of Theorem 4.4.2). All elements of $\boldsymbol{D}$ are infinite sets. If there *are* elements of $\boldsymbol{D}$, then $\bigcap\boldsymbol{D}$ is a countably infinite set (and an element of $\boldsymbol{D}$). On the other hand, if $\boldsymbol{D}$ is empty, then $\bigcap\boldsymbol{D}=\mathbf{V}$ (Exercise 2.6), which is not a set (¶ 4.2.3). So, by postulating that $\bigcap\boldsymbol{D}$ is a set, we would ensure that $\boldsymbol{D}$ is not empty, so that there would be infinite sets. However, before asserting that $\bigcap\boldsymbol{D}$ is a set, we would not have *exhibited* an element of $\boldsymbol{D}$. The question arises: Can we exhibit a countably infinite *class* without assuming that there are infinite sets?

4.5.4. We can try to define a countably infinite class as in the proof of the Recursion Theorem (4.4.5). Suppose $(\boldsymbol{C},i,\boldsymbol{F})$ is an infinitary structure, and let

$$\boldsymbol{D}=\{x\colon\forall y\ (y\in x\Rightarrow y=i\lor\exists z\ (z\in x\cap\boldsymbol{C}\ \&\ \boldsymbol{F}(z)=y))\}.\tag{42}$$

Then $\boldsymbol{D}\subseteq\mathscr{P}(\boldsymbol{C})$, 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 $(\bigcup\boldsymbol{D},i,\boldsymbol{F})$ is an infinitary structure. However, it is not clearly arithmetic. 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 $\boldsymbol{F}$. So, in this case, $(\bigcup\boldsymbol{D},i,\boldsymbol{F})$ is *not* arithmetic, and we do not know whether $\bigcup\boldsymbol{D}$ is countably infinite. We shall ultimately rule out the possibility that $a=\{a\}$ with the Foundation Axiom (8.2.1). Meanwhile, we can proceed as follows.

4.6. Well-ordered classes

4.6.1. Suppose $(\boldsymbol{C},<)$ is a strict total order.

(i) $(\boldsymbol{C},<)$ **admits (proof by) strong** or **transfinite induction** if
$$\forall x\ (x \in \boldsymbol{C}\ \&\ \mathrm{pred}(x) \subseteq \boldsymbol{C}_0 \Rightarrow x \in \boldsymbol{C}_0) \Rightarrow \boldsymbol{C}_0 = \boldsymbol{C}$$
whenever $\boldsymbol{C}_0 \subseteq \boldsymbol{C}$;

(ii) $(\boldsymbol{C},<)$ **admits (definition by) strong** or **transfinite recursion** if (a) all sections are sets, and (b) for every class $\boldsymbol{D}$ and every function $\boldsymbol{F}$ from $\mathscr{P}(\boldsymbol{D})$ to $\boldsymbol{D}$, there is a unique function $\boldsymbol{G}$ from $\boldsymbol{C}$ to $\boldsymbol{D}$ such that
$$\boldsymbol{G}(a) = \boldsymbol{F}(\boldsymbol{G}[\,\mathrm{pred}(a)\,]). \tag{43}$$

4.6.2. Theorem. *Suppose $(\boldsymbol{C},<)$ is a strict total order whose every section is a set. The following conditions are equivalent:*

(i) $(\boldsymbol{C},<)$ is well-ordered;
(ii) $(\boldsymbol{C},<)$ admits induction;
(iii) $(\boldsymbol{C},<)$ admits recursion.

Proof. There are three implications to prove.

1. Suppose (i). If $\boldsymbol{C}_0 \subset \boldsymbol{C}$, then $\boldsymbol{C} \smallsetminus \boldsymbol{C}_0$ 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 (ii).

2. Suppose next (ii), and $\boldsymbol{D}$ is a class, and $\boldsymbol{F}\colon \mathscr{P}(\boldsymbol{D}) \to \boldsymbol{D}$. We show by induction that, for all a in $\boldsymbol{C}$, there is a unique function g_a with domain $\mathrm{pred}(a) \cup \{a\}$ such that
$$g_a(c) = \boldsymbol{F}(g_a[\,\mathrm{pred}(c)\,])$$
whenever $c \leqslant a$. Suppose the claim holds whenever $a < b$. If $a < d < b$, then $g_d \restriction (\mathrm{pred}(a) \cup \{a\})$ has the defining property of g_a, so it is equal to g_a; in particular, $g_d(a) = g_a(a)$. Therefore we can define g_b by
$$g_b(x) = \begin{cases} g_x(x), & \text{if } x < b;\\ \boldsymbol{F}(\{g_y(y)\colon y < b\}), & \text{if } x = b.\end{cases} \tag{44}$$
Moreover, as before, any g_b as desired must agree with g_a on $\mathrm{pred}(a) \cup \{a\}$ when $a < b$, and then $g_b(b)$ must be as in (44). By induction, we have a function g_a as desired for all a in $\boldsymbol{C}$. Then we have (43) for all c in $\boldsymbol{C}$ if and only if $\boldsymbol{G}$ is $x \mapsto g_x(x)$.

3. Suppose finally (i) fails. Then some nonempty subclass $\boldsymbol{C}_0$ of $\boldsymbol{C}$ has no least element. Then the subclass $\{x\colon x \in \boldsymbol{C}\ \&\ \exists y\ (y \in \boldsymbol{C}_0\ \&\ y \leqslant x)\}$ of $\boldsymbol{C}$ also has no least element. Call this subclass $\boldsymbol{C}_1$. Let $\boldsymbol{F}\colon \mathscr{P}(\{0,1\}) \to \{0,1\}$, where $\boldsymbol{F}(a) = 1$ if and only if $1 \in a$. If $e \in \{0,1\}$, let $\boldsymbol{G}_e$ be the function from $\boldsymbol{C}$ into $\{0,1\}$ given by
$$\boldsymbol{G}_e(x) = \begin{cases} 0, & \text{if } x \in \boldsymbol{C} \smallsetminus \boldsymbol{C}_1;\\ e, & \text{if } x \in \boldsymbol{C}_1.\end{cases}$$
Then $\boldsymbol{G}_e(a) = \boldsymbol{F}(\boldsymbol{G}_e[\,\mathrm{pred}(a)\,])$, so (iii) fails. □

4.6.3. Theorem. *A strict total order $(\boldsymbol{C},<)$ whose every section is a set admits strong recursion if and only if, for every class $\boldsymbol{D}$, if $\boldsymbol{F}\colon \boldsymbol{E} \to \boldsymbol{D}$, where*
$$\boldsymbol{E} = \{x\colon \exists y\ (y \in \boldsymbol{C}\ \&\ x \in {}^{\mathrm{pred}(y)}\boldsymbol{D})\}, \tag{45}$$

(so $\boldsymbol{F}(g) \in \boldsymbol{D}$ whenever g is a function from a section of $\boldsymbol{D}$ into $\boldsymbol{D}$,) then there is a unique function $\boldsymbol{G}$ from $\boldsymbol{C}$ into $\boldsymbol{D}$ such that

$$\boldsymbol{G}(a) = \boldsymbol{F}(\boldsymbol{G} \upharpoonright \operatorname{pred}(a)). \tag{46}$$

Proof. Suppose first $(\boldsymbol{C}, <)$ admits strong recursion, $\boldsymbol{D}$ is a class, $\boldsymbol{E}$ is as in (45), and $\boldsymbol{F} \colon \boldsymbol{E} \to \boldsymbol{D}$. We have $\boldsymbol{E} \subseteq \mathscr{P}(\boldsymbol{C} \times \boldsymbol{D})$. Let $\boldsymbol{K}$ be a function from $\mathscr{P}(\boldsymbol{C} \times \boldsymbol{D})$ to $\boldsymbol{C} \times \boldsymbol{D}$ such that, if $f \in \boldsymbol{E}$, then $\boldsymbol{K}(f) = (a, \boldsymbol{F}(f))$, where $\operatorname{dom}(f) = \operatorname{pred}(a)$. Then there is a unique function $\boldsymbol{H}$ from $\boldsymbol{C}$ to $\boldsymbol{C} \times \boldsymbol{D}$ such that

$$\boldsymbol{H}(a) = \boldsymbol{K}(\boldsymbol{H}[\operatorname{pred}(a)]) = \boldsymbol{K}(\{\boldsymbol{H}(x) \colon x < a\}).$$

By Theorem 4.6.2, we can use induction to show $\pi_0 \circ \boldsymbol{H} = \mathrm{id}_{\boldsymbol{C}}$, where π_0 is $(x, y) \mapsto x$ (as in ¶ 3.6.5). Indeed, if $\pi_0(\boldsymbol{H}(b)) = b$ whenever $b < a$, then $\boldsymbol{H}[\operatorname{pred}(a)]$ is a function from $\operatorname{pred}(a)$ to $\boldsymbol{D}$, and in particular $\boldsymbol{H}[\operatorname{pred}(a)] = (\pi_1 \circ \boldsymbol{H}) \upharpoonright \operatorname{pred}(a)$. So $\boldsymbol{H}[\operatorname{pred}(a)] \in \boldsymbol{E}$, and $\boldsymbol{H}(a) = (a, \boldsymbol{F}(\boldsymbol{H}[\operatorname{pred}(a)]))$; in particular, $\pi_0(\boldsymbol{H}(a)) = a$. Then

$$(\pi_1 \circ \boldsymbol{H})(a) = \boldsymbol{F}((\pi_1 \circ \boldsymbol{H}) \upharpoonright \operatorname{pred}(a)).$$

Thus the function $\pi_1 \circ \boldsymbol{H}$ is a function $\boldsymbol{G}$ as in (46); it is unique, since $\boldsymbol{H}$ can be recovered from $\pi_1 \circ \boldsymbol{H}$ as $x \mapsto (x, \pi_1 \circ \boldsymbol{H}(x))$.

Now suppose conversely that $\boldsymbol{G}$ exists uniquely as in (46) whenever $\boldsymbol{F} \colon \boldsymbol{E} \to \boldsymbol{D}$, and now $\boldsymbol{K} \colon \mathscr{P}(\boldsymbol{D}) \to \boldsymbol{D}$. Then we may let $\boldsymbol{F}$ be $x \mapsto \boldsymbol{K}(\operatorname{rng}(x))$ and obtain $\boldsymbol{G}$ uniquely from $\boldsymbol{C}$ to $\boldsymbol{D}$ so that $\boldsymbol{G}(a) = \boldsymbol{F}(\boldsymbol{G} \upharpoonright \operatorname{pred}(a)) = \boldsymbol{K}(\boldsymbol{G}[\operatorname{pred}(a)])$. □

4.6.4. 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 $0^{(\boldsymbol{C},<)}$). If $\boldsymbol{D}$ is a proper initial segment of $\boldsymbol{C}$, then $\boldsymbol{D}$ is the section $\operatorname{pred}(a)$, where $a = \min(\boldsymbol{C} \smallsetminus \boldsymbol{D})$. Every element a of $\boldsymbol{C}$ that is not a greatest element has a **successor**, which we may denote by

$$a^+$$

(or more precisely $a^{+^{(\boldsymbol{C},<)}}$); it is $\min(\boldsymbol{C} \smallsetminus (\operatorname{pred}(a) \cup \{a\}))$. Immediately,

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

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) $b < a \Rightarrow b^+ < a$.

By distinguishing between 0, successors, and limits, we obtain alternative formulations of induction and recursion:

4.6.5. Theorem. *Suppose $(\boldsymbol{C}, <)$ is a well-ordered class, and $\boldsymbol{D}$ is a subclass such that:*

(i) $0 \in \boldsymbol{D}$ *(that is,* $0^{(\boldsymbol{C},<)} \in \boldsymbol{D}$*), if* $\boldsymbol{C} \neq \varnothing$*;*
(ii) $a \in \boldsymbol{D} \Rightarrow a^+ \in \boldsymbol{D}$*, if a is not the greatest element of $\boldsymbol{C}$;*
(iii) if a is a limit of $\boldsymbol{C}$ and $\operatorname{pred}(a) \subseteq \boldsymbol{D}$*, then $a \in \boldsymbol{D}$.*

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

Proof. Exercise 4.17. □

4.6.6. Theorem. *Suppose $(\boldsymbol{C},<)$ is a well-ordered class, $\boldsymbol{D}$ is a class, $b \in \boldsymbol{D}$, $\boldsymbol{F}\colon \boldsymbol{D} \to \boldsymbol{D}$, and $\boldsymbol{G}\colon \mathscr{P}(\boldsymbol{D}) \to \boldsymbol{D}$. Then there is a unique function $\boldsymbol{H}$ from $\boldsymbol{C}$ to $\boldsymbol{D}$ such that*

(i) *$\boldsymbol{H}(0) = b$, if $\boldsymbol{C} \neq \varnothing$;*
(ii) *$\boldsymbol{H}(a^+) = \boldsymbol{F}(\boldsymbol{H}(a))$, if a is not the greatest element of $\boldsymbol{C}$;*
(iii) *$\boldsymbol{H}(d) = \boldsymbol{G}(\boldsymbol{H}[\operatorname{pred}(d)])$, if d is a limit.*

Proof. Exercise 4.18, using Theorem 4.6.3. □

4.6.7. An **embedding**of an ordered class $(\boldsymbol{C},\boldsymbol{R})$ in another one, $(\boldsymbol{D},\boldsymbol{S})$, is an injection $\boldsymbol{F}$ of the *class* $\boldsymbol{C}$ in $\boldsymbol{D}$ (¶ 3.6.7) such that $a\ \boldsymbol{R}\ b \Leftrightarrow a\ \boldsymbol{S}\ b$. A surjective embedding is an **isomorphism.**

4.6.8. Theorem. *Of any two well-ordered classes, one is uniquely isomorphic to a unique initial segment of the other.*

Proof. Let $(\boldsymbol{C},\boldsymbol{R})$ and $(\boldsymbol{D},\boldsymbol{S})$ be well-ordered classes. There is at most one way that $(\boldsymbol{C},\boldsymbol{R})$ can be isomorphic to an initial segment of $(\boldsymbol{D},\boldsymbol{S})$. Indeed, let $\boldsymbol{F}$ and $\boldsymbol{G}$ be two such isomorphisms. If they agree *on* $\operatorname{pred}(a)$, then they must agree *at* a (Exercise 4.20). By induction, the functions agree on $\boldsymbol{C}$.

Now let $\boldsymbol{C}^*$ comprise the elements a of $\boldsymbol{C}$ for which there is a function f_a such that, for all b in $\operatorname{pred}(a) \cup \{a\}$, the restriction $f_a \restriction (\operatorname{pred}(b) \cup \{b\})$ is an isomorphism from $(\operatorname{pred}(b) \cup \{b\}, \boldsymbol{R})$ to an initial segment of $(\boldsymbol{D},\boldsymbol{S})$. If $a \in \boldsymbol{C}^*$, and $b\ \boldsymbol{R}\ a$, then the function $f_a \restriction (\operatorname{pred}(b) \cup \{b\})$ shows $b \in \boldsymbol{C}^*$. Thus $\boldsymbol{C}^*$ is an initial segment of $\boldsymbol{C}$. Conversely, if a and b are in $\boldsymbol{C}^*$, and $b\ \boldsymbol{R}\ a$, then the uniqueness result above shows $f_b = f_a \restriction (\operatorname{pred}(a) \cup \{a\})$. Therefore the union $\bigcup\{f_x\colon x \in \boldsymbol{C}^*\}$ is the function $x \mapsto f_x(x)$ on $\boldsymbol{C}^*$, and this function is an isomorphism from $(\boldsymbol{C}^*, \boldsymbol{R})$ to an initial segment $(\boldsymbol{D}^*, \boldsymbol{S})$ of $(\boldsymbol{D},\boldsymbol{S})$.

Let this function be called $\boldsymbol{F}$, and suppose $\operatorname{pred}(a) \subseteq \boldsymbol{C}^*$ and $\boldsymbol{F}[\operatorname{pred}(a)] = \operatorname{pred}(c)$ for some a in $\boldsymbol{C}$ and c in $\boldsymbol{D}$. Then the function $\boldsymbol{F} \restriction \operatorname{pred}(a) \cup \{(a,c)\}$ shows that $a \in \boldsymbol{C}^*$. We can conclude that either $\boldsymbol{C}^* = \boldsymbol{C}$ or else $\boldsymbol{D}^* = \boldsymbol{D}$. In the former case, $\boldsymbol{F}$ is an isomorphism from $(\boldsymbol{C},\boldsymbol{R})$ to an initial segment of $(\boldsymbol{D},\boldsymbol{S})$; in the latter case, $\breve{\boldsymbol{F}}$ is an isomorphism from $(\boldsymbol{D},\boldsymbol{S})$ to an initial segment of $(\boldsymbol{C},\boldsymbol{R})$. □

4.7. Construction of countably infinite classes

4.7.1. Lemma. *Suppose $(\boldsymbol{C}, i, \boldsymbol{F})$ is an infinitary structure, and $\boldsymbol{A}$ is a subclass of $\boldsymbol{C}$ whose every element is either i or $\boldsymbol{F}(b)$ for some b in $\boldsymbol{A}$. Then there is at most one binary relation $<$ on $\boldsymbol{A}$ such that*

(i) *$a < \boldsymbol{F}(a)$ whenever $a \in \boldsymbol{C}$ and $\boldsymbol{F}(a) \in \boldsymbol{A}$;*
(ii) *$\boldsymbol{A}$ is well-ordered by $<$.*

Moreover, if $<$ is such a relation and $a \in \boldsymbol{A}$, then $\boldsymbol{F}(a)$ is the least element of $\{x\colon x \in \boldsymbol{A}\ \&\ a < x\}$, if this class is non-empty.

Proof. Suppose $<$ is a strict ordering of $\boldsymbol{A}$ such that (i) holds. Suppose there is some a in $\boldsymbol{A}$ such that $\boldsymbol{F}(a)$ is not the least element of $\{x\colon x \in \boldsymbol{A}\ \&\ a < x\}$, although this class is non-empty. If there is no *least* such a, then $\boldsymbol{A}$ is not well-ordered by $<$. So suppose a is least. If there is no least element of $\{x\colon x \in \boldsymbol{A} \mathbin{\&} a < x\}$, then $\boldsymbol{A}$ is not well-ordered by $<$. So suppose b is the least element of the class. Then

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

If $b = i$, then $\boldsymbol{A}$ has no least element, so $\boldsymbol{A}$ is not well-ordered by $<$. So suppose $b \neq i$. Then $b = \boldsymbol{F}(c)$ for some c in $\boldsymbol{A}$, and then $c < b$, but $c \neq a$. Also $c \nless a$, by the leastness of a; and $a \nless c$, by the leastness of b. Therefore $<$ is not a total ordering of $\boldsymbol{A}$.

Now suppose $\boldsymbol{R}$ is a binary relation $<$ on $\boldsymbol{A}$ satisfying (i) and (ii). We have shown $\boldsymbol{F}(a)$ is the least element of $\{x\colon x \in \boldsymbol{A} \mathbin{\&} a < x\}$, assuming this is nonempty. Suppose $\boldsymbol{S}$ is a binary relation on $\boldsymbol{A}$, distinct from $\boldsymbol{R}$, but such that

$$\{(c, \boldsymbol{F}(x))\colon x \in \boldsymbol{C} \mathbin{\&} \boldsymbol{F}(x) \in \boldsymbol{A}\} \subseteq \boldsymbol{S}.$$

Let a be least such that $a\boldsymbol{R} \neq a\boldsymbol{S}$, and let b be the least element of $a\boldsymbol{R} \vartriangle a\boldsymbol{S}$, always with respect to $\boldsymbol{R}$. There are two cases.

1. Suppose $a \mathrel{\boldsymbol{R}} b$, that is, $a < b$, but $\neg(a \mathrel{\boldsymbol{S}} b)$. Then $b \neq i$, so $b = \boldsymbol{F}(c)$ for some c in $\boldsymbol{A}$. Therefore $c < b$ and $c \mathrel{\boldsymbol{S}} b$, so $a \neq c$ and hence $a < c$, that is, $a \mathrel{\boldsymbol{R}} c$. By leastness of b, we have $a \mathrel{\boldsymbol{S}} c$, so $\boldsymbol{S}$ is not transitive.

2. Suppose $\neg(a \mathrel{\boldsymbol{R}} b)$, but $a \mathrel{\boldsymbol{S}} b$. Then $b \leqslant a$. If $b = a$, then $\boldsymbol{S}$ is not irreflexive. If $b < a$, then $b\boldsymbol{R} = b\boldsymbol{S}$, so $b \mathrel{\boldsymbol{S}} a$, so $\boldsymbol{S}$ is not an ordering.

This establishes uniqueness of $<$. □

4.7.2. Axiom (Binary union). *The union of two sets is a set:*

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

4.7.3. Theorem. *A relation from a set to another is a set.*

Proof. $a \times b \subseteq \mathscr{P}(\mathscr{P}(a \cup b))$ by ¶ 3.4.4. □

4.7.4. Theorem. *Let $(\boldsymbol{C}, i, \boldsymbol{F})$ be an infinitary structure. Then $\boldsymbol{C}$ has a subclass $\boldsymbol{D}$ such that*

(i) $(\boldsymbol{D}, i, \boldsymbol{F})$ is an arithmetic structure;
(ii) $\boldsymbol{D}$ is well-ordered by some unique $<$ such that $a < \boldsymbol{F}(a)$ for all a in $\boldsymbol{D}$.

Proof. Let $\boldsymbol{E}$ comprise all subsets a of $\boldsymbol{C}$ such that

(i) every element of a is either i or $\boldsymbol{F}(b)$ for some b in a, and
(ii) a is well-ordered by some $<$ such that $b < \boldsymbol{F}(b)$ whenever $\boldsymbol{F}(b) \in a$.

Here $<$ is a set by Theorem 4.7.3; by Lemma 4.7.1, it is unique. We have $\{i\} \in \boldsymbol{E}$. Suppose $a \in \boldsymbol{E}$, and r is the associated ordering $<$. If $b \in a$, but $\boldsymbol{F}(b) \notin a$, then $a \cup \{\boldsymbol{F}(b)\}$ is well-ordered by $r \cup \{(x, \boldsymbol{F}(b))\colon x \in a\}$, and this shows $a \cup \{\boldsymbol{F}(b)\} \in \boldsymbol{E}$. So $\bigcup \boldsymbol{E}$ contains i and is closed under $\boldsymbol{F}$; hence $(\bigcup \boldsymbol{E}, i, \boldsymbol{F})$ is an infinitary structure.

We shall show that $\bigcup \boldsymbol{E}$ is the desired class $\boldsymbol{D}$. To do so, suppose a and c are elements of $\boldsymbol{E}$. By Theorem 4.6.8, we may assume that there is an isomorphism g from $(a, <)$ onto an initial segment of $(c, <)$. Then $g(i) = i$, and if $g(b) = b$, and $\boldsymbol{F}(b) \in a$, then $g(\boldsymbol{F}(b)) = \boldsymbol{F}(b)$ by Lemma 4.7.1. Since $(a, <)$ has no limits, we have $g = \mathrm{id}_a$ by induction, by Theorem 4.6.5. That is, $(a, <)$ is an initial segment of $(c, <)$. If we now write r_a for the ordering of a, then $\bigcup \boldsymbol{E}$ is well-ordered by $\bigcup\{r_x\colon x \in \boldsymbol{E}\}$; this ordering is unique by 4.7.1. Since every element of $\bigcup \boldsymbol{E}$ is either i or $\boldsymbol{F}(b)$ for some b in $\bigcup \boldsymbol{E}$, we can conclude by Theorem 4.6.5 that $\bigcup \boldsymbol{E}$ is the desired class $\boldsymbol{D}$. □

4.7.5. In Theorem 4.7.4, the unique ordering $<$ may be called the ordering **associated** with the arithmetic structure. Since an infinitary structure does exist by Theorem 4.2.2, we can use Theorem 4.7.4 to obtain a particular arithmetic structure, say $(\boldsymbol{C}, i, \boldsymbol{F})$. By Theorem 4.6.3, there is a function $\boldsymbol{G}$ on $\boldsymbol{C}$ such that

$$\boldsymbol{G}(a) = \boldsymbol{G}[\,\mathrm{pred}(a)\,].$$

We denote $\mathrm{rng}(\boldsymbol{G})$ by

$$\omega.$$

4.7.6. Theorem. *The definition of* ω *is independent of the choice of infinitary structure* $(\boldsymbol{C}, i, \boldsymbol{F})$, *and* $(\omega, \varnothing, x \mapsto x \cup \{x\})$ *is an arithmetic structure, whose associated ordering is both membership* ($\in$) *and proper inclusion* ($\subset$).

Proof. Exercise 4.21. □

4.7.7. The elements of ω are the **natural numbers.** If n is one of them, we denote the **successor** $n \cup \{n\}$ of n by

$$n'.$$

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

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

In general, a natural number n is the set that can be denoted by

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

Exercises

4.1. Show that (30) does not generally stand for a sentence of the logic of sets.

4.2. Prove that (31) stands for a sentence of the logic of sets.

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

4.4. Show that $\preccurlyeq$ is transitive and reflexive, but $\prec$ is transitive and irreflexive.

4.5. Complete the proof of the Schroeder–Bernstein Theorem.

4.6. Is the Schroeder–Bernstein Theorem true for classes in general?

4.7. Establish the implications (36) and (37) in the proof of Cantor's Theorem.

4.8. Why does the proof of Cantor's Theorem (¶ 4.1.5) require $\{x \in a : x \notin f(x)\}$ to be a set?

4.9. If $a \neq b$, so that one of $a \smallsetminus b$ and $b \smallsetminus a$ is nonempty. If $a \smallsetminus b \neq \varnothing$, show that $a \smallsetminus b \in \mathscr{P}(a) \smallsetminus \mathscr{P}(b)$ and hence $\mathscr{P}(a) \neq \mathscr{P}(b)$.

4.10. Find a set a of minimal cardinality such that (a, i, f) is recursive for some i in a and f in ${}^{a}a$.

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

4.12. Prove the claim in the proof of Theorem 4.4.2.

4.13. Derive the Pairing Axiom from the Augmentation Axiom and the Separation Scheme.

4.14. Concerning the corollary (4.4.6) to the Recursion Theorem:

(i) prove it by means of the Theorem;

(ii) prove it directly by considering the class of sets a whose every element is either (i, j) or else $(\boldsymbol{F}(b), \boldsymbol{G}(c))$ for some element (b, c) of a.

4.15. Prove that the inverse of an isomorphism of iterative structures is an isomorphism (¶ 4.4.7) and that being isomorphic is an equivalence-relation.

4.16. Show that every subset of an ordered set that admits strong induction admits strong induction.

4.17. Prove Theorem 4.6.5.

4.18. Prove Theorem 4.6.6.

4.19. Suppose $(\boldsymbol{A}, \boldsymbol{R})$ and $(\boldsymbol{B}, \boldsymbol{S})$ are strict total orders, $\boldsymbol{H}\colon \boldsymbol{A} \to \boldsymbol{B}$, and $c\ \boldsymbol{R}\ d \Rightarrow \boldsymbol{H}(c)\ \boldsymbol{S}\ \boldsymbol{H}(d)$. Show that $\boldsymbol{H}$ is an isomorphism from $(\boldsymbol{A}, \boldsymbol{R})$ to $(\boldsymbol{H}[\,\boldsymbol{A}\,], \boldsymbol{S})$.

4.20. Let $\boldsymbol{F}$ and $\boldsymbol{G}$ be isomorphisms from a well-ordered class $(\boldsymbol{C}, \boldsymbol{R})$ to initial segments of a well-ordered class $(\boldsymbol{D}, \boldsymbol{S})$. Assuming that $\boldsymbol{F}$ and $\boldsymbol{G}$ agree on $\operatorname{pred}(a)$ for some a in $\boldsymbol{C}$, show that $\boldsymbol{F}(a) = \boldsymbol{G}(a)$.

4.21. Prove Theorem 4.7.6.

CHAPTER 5

The natural numbers

5.1. Structures

5.1.1. If $n \in \omega$, then the class ${}^n\boldsymbol{C}$ may also be denoted by

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

although some cases of this notation will get a different interpretation in ¶ 7.4.1. An element f of $\boldsymbol{C}^n$ can be written also as $x \mapsto f_x$. Then we have

$$\boldsymbol{C}^0 = \{0\} = 1, \tag{47}$$

as well as

$$x \mapsto x_0 \colon \boldsymbol{C}^1 \rightarrowtail\!\!\!\!\twoheadrightarrow \boldsymbol{C}, \tag{48}$$

$$x \mapsto (x_0, x_1) \colon \boldsymbol{C}^2 \rightarrowtail\!\!\!\!\twoheadrightarrow \boldsymbol{C} \times \boldsymbol{C}. \tag{49}$$

Note in particular that (47) still holds if $\boldsymbol{C} = \varnothing$, while $\varnothing^n = \varnothing$ if $n > 0$. 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 5.1).

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

$$(a_i \colon i \in n), \qquad (a_0, \dots, a_{n-1}), \qquad \vec{a}.$$

If the free variables of a formula φ are among the variables x_i, where $i \in n$, then we can write φ as $\varphi(x_0, \dots, x_{n-1})$ 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 subclass of $\boldsymbol{C}^n$ can be called an **n-ary relation on $\boldsymbol{C}$**: such a relation is

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

for some n-ary formula φ. By (48), a 1-ary relation on $\boldsymbol{C}$ can be considered as a subclass of $\boldsymbol{C}$; by (49), 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 (47) and (48), 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 operation $x \mapsto x'$ of succession on ω) 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 class equipped with some relations and operations on it. The class is then the **universe** of the structure. Examples include:

(i) orders (¶ 3.5.3), and in particular well-ordered sets (¶ 3.5.7);

(ii) iterative structures (¶ 4.2.5), and in particular recursive structures (¶ 4.2.8), infinitary structures, and arithmetic structures (¶ 4.4.1);

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

(iv) $(\mathscr{P}(a), \cap, \cup, x \mapsto x^{\mathrm{c}}, \varnothing, a)$.

We can understand a structure formally as a class (Exercise 5.2), 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. (See Appendix B.) So the structure $(\boldsymbol{C}, \dots)$ may be denoted by $\mathfrak{C}$, and $(a, \dots)$ by $\mathfrak{a}$.

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 $\mathfrak{C}$.

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

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

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, and $\boldsymbol{H} \colon \boldsymbol{C} \to \boldsymbol{D}$. For each n in ω, we can understand $\boldsymbol{H}$ also as the function $\vec{x} \mapsto (\boldsymbol{H}(x_i) \colon i \in n)$ from $\boldsymbol{C}^n$ to $\boldsymbol{D}^n$. If

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

(ii) $\vec{a} \in R^{\mathfrak{C}} \Rightarrow \boldsymbol{H}(\vec{a}) \in R^{\mathfrak{D}}$ for all $\vec{a}$ in $\boldsymbol{C}^n$ 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

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

The homomorphisms defined in ¶ 4.3.1 are a special case.

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. This is a special case of the embeddings defined in ¶¶ 3.6.7 and 4.6.7. In the general cas, if $\boldsymbol{C} \subseteq \boldsymbol{D}$, and $\mathrm{id}_{\boldsymbol{C}}$ is an embedding, then $\mathfrak{C}$ is a **substructure** 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 5.3).

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 $\mathfrak{D}$.

By Theorem 4.4.8, all arithmetic structures are isomorphic. We know of an arithmetic structure, namely $(\omega, \varnothing, x \mapsto x')$. However, throughout this chapter, let

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

denote an arbitrary arithmetic structure; also, let 0^{+} be denoted by

$$1.$$

We shall think of $\mathbb{N}$ as the class of natural numbers, when the understanding of natural numbers as ordinals in the sense of ¶ 6.1.3 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 recursive structures

5.2.1. Theorem and Definition. *On $\mathbb{N}$, there is a unique binary operation, called **addition** and denoted by $+$, such that*

(i) $m + 0 = m$,
(ii) $m + n^{+} = (m + n)^{+}$.

Proof. Exercise 5.4. □

5.2.2. It is sometimes suggested that the validity of the definition of addition given in the theorem is an easy consequence of induction. The intended argument seems to run as follows:

> We define $x + 0$ as x, for all x in $\mathbb{N}$. If $x + n$ has been defined for all x in $\mathbb{N}$, then we let $x + n^{+} = (x + n)^{+}$. Therefore, by induction, $x + y$ is defined for all x and y in $\mathbb{N}$.

Indeed, Peano [**16**] seems to have had in mind such an argument. Landau [**11**] reports having accepted such an argument until he was forced to recognize its wrongness. But Dedekind [**3**, II.IX (130)] has had the truth all along, even though he refers to definition by recursion as definition by induction. Proof by induction is not a method of definition; it is a method of proving that two classes are equal. The proposed 'proof' here of the definition of addition would have to start out in the following fashion:

> Let $\boldsymbol{C}$ be the subclass of $\mathbb{N}$ comprising all y for which the operation $x \mapsto x + y$ is defined on $\mathbb{N}$ so that $\forall x\ x + 0 = x$ and $\forall x\ \forall z\ x + z^{+} = (x + z)^{+}$.

But this definition of $\boldsymbol{C}$ makes no sense. Nonetheless, it *is* possible to define addition on every recursive structure $(\boldsymbol{A}, i, \boldsymbol{S})$, as we shall now see.

5.2.3. Lemma. *if $\boldsymbol{F}\colon \boldsymbol{B} \to \boldsymbol{C}$ and $\boldsymbol{G}\colon \boldsymbol{C} \to \boldsymbol{C}$, then there is a unique function $\boldsymbol{H}$ from $\boldsymbol{B} \times \mathbb{N}$ to $\boldsymbol{C}$ such that*

(i) $a\,\boldsymbol{H}\,0 = \boldsymbol{F}(a)$,
(ii) $a\,\boldsymbol{H}\,n^{+} = \boldsymbol{G}(a\,\boldsymbol{H}\,n)$.

Informally,

$$a\,\boldsymbol{H}\,n = \underbrace{\boldsymbol{G} \circ \cdots \circ \boldsymbol{G}}_{n} \circ \boldsymbol{F}(a).$$

Proof. Let the operation $(x, y) \mapsto (x, \boldsymbol{G}(y))$ on $\boldsymbol{B} \times \boldsymbol{C}$ be denoted by $\boldsymbol{K}$. By the Recursion Theorem (4.4.5), there is a unique subclass $\boldsymbol{R}$ of $\mathbb{N} \times (\boldsymbol{B} \times \boldsymbol{C})$ such that

(i) $0\boldsymbol{R} = \boldsymbol{F}$,
(ii) $n \in \mathbb{N} \Rightarrow n^{+}\boldsymbol{R} = \boldsymbol{K}[n\boldsymbol{R}] = \{(x, \boldsymbol{G}(y)) \colon n \mathrel{\boldsymbol{R}} (x, y)\}$.

By induction, if $n \in \mathbb{N}$, then $n\boldsymbol{R}$ is a function from $\boldsymbol{B}$ to $\boldsymbol{C}$. Indeed, $0\boldsymbol{R}$ is such a function (namely $\boldsymbol{F}$), and if $n\boldsymbol{R}$ is such a function, then $n^{+}\boldsymbol{R} = \boldsymbol{G} \circ (n\boldsymbol{R})$. We can now define the binary function $\boldsymbol{H}$ as $(x, y) \mapsto y\boldsymbol{R}(x)$. Then

(i) $a \mathrel{\boldsymbol{H}} 0 = 0\boldsymbol{R}(a) = \boldsymbol{F}(a)$,
(ii) $a \mathrel{\boldsymbol{H}} n^{+} = \boldsymbol{G} \circ (n\boldsymbol{R})(a) = \boldsymbol{G}(a \mathrel{\boldsymbol{H}} n)$.

So $\boldsymbol{H}$ is as desired. To see that $\boldsymbol{H}$ is unique, note that $\boldsymbol{R}$ determines $\boldsymbol{H}$, and conversely. Indeed,

$$\boldsymbol{H} = \{((x, y), z) \colon y \mathrel{\boldsymbol{R}} (x, z)\}, \qquad \boldsymbol{R} = \{(y, (x, z)) \colon x \mathrel{\boldsymbol{H}} y = z\}.$$

Since $\boldsymbol{R}$ uniquely satisfies the given conditions, so does $\boldsymbol{H}$. □

5.2.4. Theorem and Definition. *Suppose $(\boldsymbol{A}, i, \boldsymbol{S})$ is recursive. Then there is a unique binary operation of* ***addition*** *on $\boldsymbol{A}$ given by*

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

Proof. By Lemma 5.2.3, there is a unique function $\boldsymbol{H}$ from $\boldsymbol{A} \times \mathbb{N}$ to $\boldsymbol{A}$ such that

(i) $a \mathrel{\boldsymbol{H}} 0 = a$,
(ii) $a \mathrel{\boldsymbol{H}} n^{+} = \boldsymbol{S}(a \mathrel{\boldsymbol{H}} n)$.

So $\boldsymbol{H}$ is recursively defined in its second argument. We shall show that $\boldsymbol{H}$ is also recursively definable in its first argument, that is,

(i) $i \mathrel{\boldsymbol{H}} n = \boldsymbol{F}(n)$,
(ii) $\boldsymbol{S}(a) \mathrel{\boldsymbol{H}} n = \boldsymbol{S}(a \mathrel{\boldsymbol{H}} n)$,

where $\boldsymbol{F}$ is just the function $x \mapsto i \mathrel{\boldsymbol{H}} x$. We use induction. We have (ii) when $n = 0$, since $\boldsymbol{S}(a) \mathrel{\boldsymbol{H}} 0 = \boldsymbol{S}(a) = \boldsymbol{S}(a \mathrel{\boldsymbol{H}} 0)$. Suppose we have (ii) when $n = k$. Then

$$\begin{aligned}
\boldsymbol{S}(a) \mathrel{\boldsymbol{H}} k^{+} &= \boldsymbol{S}(\boldsymbol{S}(a) \mathrel{\boldsymbol{H}} k) && \text{[by definition of } \boldsymbol{H}\text{]}\\
&= \boldsymbol{S}(\boldsymbol{S}(a \mathrel{\boldsymbol{H}} k)) && \text{[by inductive hypothesis]}\\
&= \boldsymbol{S}(a \mathrel{\boldsymbol{H}} k^{+}). && \text{[by definition of } \boldsymbol{H}\text{]}
\end{aligned}$$

Thus (ii) holds when $n = k^{+}$. This completes the induction.

Now (i) and (ii) hold. In particular, the function $x \mapsto x \mathrel{\boldsymbol{H}} n$ on $\boldsymbol{A}$ is uniquely determined by $\boldsymbol{F}(n)$, by Theorem 4.3.3. That is,

$$\boldsymbol{F}(m) = \boldsymbol{F}(n) \Leftrightarrow \forall x \; x \mathrel{\boldsymbol{H}} m = x \mathrel{\boldsymbol{H}} n.$$

Moreover, $\boldsymbol{F} \colon \mathbb{N} \twoheadrightarrow \boldsymbol{A}$, by induction, since $\boldsymbol{F}(0) = i \mathrel{\boldsymbol{H}} 0 = i$, while if $\boldsymbol{F}(n) = a$, then

$$\boldsymbol{F}(n^{+}) = i \mathrel{\boldsymbol{H}} n^{+} = \boldsymbol{S}(i \mathrel{\boldsymbol{H}} n) = \boldsymbol{S}(\boldsymbol{F}(n)) = \boldsymbol{S}(a).$$

Now we can define addition on $\boldsymbol{A}$ by

$$a + b = c \Leftrightarrow \exists x \; (\boldsymbol{F}(x) = b \;\&\; a \mathrel{\boldsymbol{H}} x = c).$$

Since $\boldsymbol{F}(0) = i$, we have $a + i = a \mathrel{\boldsymbol{H}} 0 = a$. Also, if $b = \boldsymbol{F}(n)$, so that $\boldsymbol{S}(b) = \boldsymbol{F}(n^{+})$, then

$$a + \boldsymbol{S}(b) = a \mathrel{\boldsymbol{H}} n^{+} = \boldsymbol{S}(a \mathrel{\boldsymbol{H}} n) = \boldsymbol{S}(a + b).$$

Thus + is as desired; it is unique by Theorem 4.3.3. □

5.2.5. Lemma. *Suppose $(\boldsymbol{A}, i, \boldsymbol{S})$ is recursive. 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 Lemma 5.2.5, we cannot establish (ii) by induction on b.

5.2.7. If we know that $\boldsymbol{A}$ is a *set,* so that operations on $\boldsymbol{A}$ are sets, then we can prove Theorem 5.2.4 as follows. For each a in $\boldsymbol{A}$, by Corollary 4.3.4, there is at *most* one operation $x \mapsto x + a$ on $\boldsymbol{A}$ such that

(i) $i + a = a$,

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

This much does not require $\boldsymbol{A}$ to be a set; but the next observation does. By induction, for each a in $\boldsymbol{A}$, there is at *least* one such operation $x \mapsto x + a$; for we can let $x \mapsto x + 0$ be $\mathrm{id}_{\boldsymbol{A}}$, and if $x \mapsto x + b$ has been defined, then let $x + \boldsymbol{S}(b) = \boldsymbol{S}(x + b)$. This gives addition as desired. But again, if we do not know that $\boldsymbol{A}$ is a set, then it seems we have to follow a more roundabout route, as above.

5.2.8. Theorem. *Suppose $(\boldsymbol{A}, i, \boldsymbol{S})$ is recursive. For all a, b, and c in $\boldsymbol{A}$,*

(i) $\boldsymbol{S}(a) = a + \boldsymbol{S}(i)$*; in particular,* $n^+ = n + 1$ *in* $\mathbb{N}$*;*

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

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

Moreover, if $\boldsymbol{S}$ is injective, then

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

Proof. Exercise 5.5. □

5.3. Multiplication on recursive structures

5.3.1. Theorem and Definition. *On $\mathbb{N}$, there is a unique binary operation, called* ***multiplication*** *and denoted by $\cdot$ or by juxtaposition ($(x, y) \mapsto xy$), such that*

(i) $m \cdot 0 = 0$,

(ii) $m \cdot (n + 1) = m \cdot n + m$.

Proof. Exercise 5.7. □

5.3.2. Lemma. *If $\boldsymbol{F} \colon \boldsymbol{B} \to \boldsymbol{C}$ and $\boldsymbol{G} \colon \boldsymbol{C} \times \boldsymbol{B} \to \boldsymbol{C}$, then there is a unique function $\boldsymbol{H}$ from $\boldsymbol{B} \times \mathbb{N}$ to $\boldsymbol{C}$ such that*

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

(ii) $a\,\boldsymbol{H}(n + 1) = (a\,\boldsymbol{H}\,n)\,\boldsymbol{G}\,a$.

Proof. Let the operation $(x, y) \mapsto (x, y\,\boldsymbol{G}\,x)$ on $\boldsymbol{B} \times \boldsymbol{C}$ be denoted by $\boldsymbol{K}$. By the Recursion Theorem, 4.4.5, there is a unique subclass $\boldsymbol{R}$ of $\mathbb{N} \times (\boldsymbol{B} \times \boldsymbol{C})$ such that

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

(ii) $n \in \mathbb{N} \Rightarrow (n+1)\boldsymbol{R} = \boldsymbol{K}[n\boldsymbol{R}] = \{(x, y\,\boldsymbol{G}\,x)) \colon n\;\boldsymbol{R}\;(x,y)\}$.

By induction, if $n \in \mathbb{N}$, then $n\boldsymbol{R}$ is a function from $\boldsymbol{B}$ to $\boldsymbol{C}$. Indeed, $0\boldsymbol{R}$ is such a function (namely $\boldsymbol{F}$), and if $n\boldsymbol{R}$ is such a function, then $(n+1)\boldsymbol{R}$ is $x \mapsto n\boldsymbol{R}(x)\,\boldsymbol{G}\,x$. We can now define the binary function $\boldsymbol{H}$ as $(x,y) \mapsto y\boldsymbol{R}(x)$ on $\boldsymbol{A} \times \mathbb{N}$. Then

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

(ii) $a\,\boldsymbol{H}(n+1) = (n+1)\boldsymbol{R}(a) = n\boldsymbol{R}(a)\,\boldsymbol{G}\,a = (a\,\boldsymbol{H}\,n)\,\boldsymbol{G}\,a$.

So $\boldsymbol{H}$ is as desired; its uniqueness follows from that of $\boldsymbol{R}$, as in the proof of Lemma 5.2.3. □

5.3.3. Theorem and Definition. *Suppose $(\boldsymbol{A}, i, \boldsymbol{S})$ is recursive. Then there is a unique binary operation of* ***multiplication*** *on $\boldsymbol{A}$ given by*

(i) $a \cdot 0 = 0$,

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

Proof. We follow the pattern of the proof of Theorem 5.2.4. By Lemma 5.3.2, there is a unique function $\boldsymbol{H}$ from $\boldsymbol{A} \times \mathbb{N}$ into $\boldsymbol{A}$ such that

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

(ii) $a\,\boldsymbol{H}(n+1) = a\,\boldsymbol{H}\,n + a$.

By induction, $i\,\boldsymbol{H}\,n = i$ for all n in $\mathbb{N}$; indeed, this is given when $n = 0$, and if it holds when $n = m$, then $i\,\boldsymbol{H}\,m^+ = i\,\boldsymbol{H}\,m + i = i\,\boldsymbol{H}\,m = i$. Let $\boldsymbol{F}$ be the unique homomorphism from $(\mathbb{N}, 0, {}^+)$ into $(\boldsymbol{A}, i, \boldsymbol{S})$; by induction, it is surjective. The equation

$$\boldsymbol{S}(a)\,\boldsymbol{H}\,n = a\,\boldsymbol{H}\,n + \boldsymbol{F}(n) \tag{50}$$

holds when $n = 0$, since $\boldsymbol{S}(a)\,\boldsymbol{H}\,0 = i = i + i = a\,\boldsymbol{H}\,0 + \boldsymbol{F}(0)$. Suppose (50) holds for some n in $\mathbb{N}$. Then

$$\begin{aligned}
\boldsymbol{S}(a)\,\boldsymbol{H}(n+1) &= \boldsymbol{S}(a)\,\boldsymbol{H}\,n + \boldsymbol{S}(a) && [\text{by definition of } \boldsymbol{H}]\\
&= (a\,\boldsymbol{H}\,n + \boldsymbol{F}(n)) + \boldsymbol{S}(a) && [\text{by inductive hypothesis}]\\
&= a\,\boldsymbol{H}\,n + (\boldsymbol{F}(n) + \boldsymbol{S}(a)) && [\text{by associativity of } +]\\
&= a\,\boldsymbol{H}\,n + \boldsymbol{S}(\boldsymbol{F}(n) + a) && [\text{by definition of } +]\\
&= a\,\boldsymbol{H}\,n + (\boldsymbol{S}(\boldsymbol{F}(n)) + a) && [\text{by Lemma 5.2.5}]\\
&= a\,\boldsymbol{H}\,n + (\boldsymbol{F}(n+1) + a) && [\text{because } \boldsymbol{F} \text{ is a homomorphism}]\\
&= a\,\boldsymbol{H}\,n + (a + \boldsymbol{F}(n+1)) && [\text{by commutativity of } +]\\
&= (a\,\boldsymbol{H}\,n + a) + \boldsymbol{F}(n+1) && [\text{by associativity of } +]\\
&= a\,\boldsymbol{H}(n+1) + \boldsymbol{F}(n+1). && [\text{by definition of } \boldsymbol{H}]
\end{aligned}$$

So (50) holds for all n in $\mathbb{N}$. Therefore each of the operations $x \mapsto x\,\boldsymbol{H}\,n$ is recursively defined by

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

(ii) $\boldsymbol{S}(a)\,\boldsymbol{H}\,n = a\,\boldsymbol{H}\,n + \boldsymbol{F}(n)$.

In particular,

$$\boldsymbol{F}(m) = \boldsymbol{F}(n) \Leftrightarrow \forall x\; x\,\boldsymbol{H}\,m = x\,\boldsymbol{H}\,n.$$

Now we can define multiplication on $\boldsymbol{A}$ by

$$a \cdot b = c \Leftrightarrow \exists x\; (\boldsymbol{F}(x) = b \;\&\; a\,\boldsymbol{H}\,x = c).$$

Then $a \cdot i = a \boldsymbol{H}\, 0 = i$. Also, if $b = \boldsymbol{F}(n)$, so that $\boldsymbol{S}(b) = \boldsymbol{F}(n+1)$, then

$$a \cdot \boldsymbol{S}(b) = a \boldsymbol{H}(n+1) = a \boldsymbol{H}\, n + a = a \cdot b + a.$$

Thus $\cdot$ is as desired; it is unique by Theorem 4.3.3. □

5.3.4. LEMMA. *Suppose $(\boldsymbol{A}, i, \boldsymbol{S})$ is recursive. 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 5.8. □

5.3.5. THEOREM. *Suppose $(\boldsymbol{A}, i, \boldsymbol{S})$ is recursive. 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.***

PROOF. Exercise 5.9. □

5.4. Exponentiation

5.4.1. THEOREM AND DEFINITION. *On $\mathbb{N}$, there is a unique binary operation, called* ***exponentiation*** *and denoted by $(x, y) \mapsto x^y$, such that*

(i) $m^0 = 1$,
(ii) $m^{n+1} = m^n \cdot m$.

More generally, if $(\boldsymbol{A}, i, \boldsymbol{S})$ is recursive, there is a unique function from $\boldsymbol{A} \times \mathbb{N}$ into $\boldsymbol{A}$, also called exponentiation and denoted the same way, such that, for all a in $\boldsymbol{A}$ and m in $\mathbb{N}$,

(i) $a^0 = \boldsymbol{S}(i)$,
(ii) $a^{m+1} = a^m \cdot a$.

PROOF. Exercise 5.11. □

5.4.2. An **endomorphism** of a structure is a homomorphism from the structure into itself. The endomorphisms of $\mathfrak{C}$ compose a family (¶ 3.6.3), which we may denote by

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

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

5.4.3. THEOREM. *Suppose $(\boldsymbol{A}, i, \boldsymbol{S})$ is a recursive structure. For all a in $\boldsymbol{A}$, and m and n in $\mathbb{N}$,*

(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, $x \mapsto x^m$ 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 5.12. □

5.5. The ordering of the natural numbers

5.5.1. In the ordering of $\mathbb{N}$ guaranteed by Theorem 4.7.4, we have

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

This formula suggests that the ordering can be obtained by a kind of recursion.

5.5.2. 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) $\boldsymbol{G}(a+1) = a\,\boldsymbol{F}\,\boldsymbol{G}(a)$.

Proof. Let $\boldsymbol{F}_1$ be the operation $(x, y) \mapsto (x+1, x\,\boldsymbol{F}\,y)$ on $\mathbb{N} \times \boldsymbol{C}$. Then $(\mathbb{N} \times \boldsymbol{C}, (0, i), \boldsymbol{F}_1)$ is an iterative structure, so by the corollary (¶ 4.4.6) to the Recursion Theorem, there is a unique homomorphism $\boldsymbol{H}$ to this structure from $(\mathbb{N}, 0, x \mapsto x^+)$. Let π_0 and π_1 be as in ¶ 3.6.5. 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

$$\boldsymbol{H}(b) = (b, \boldsymbol{G}(b)). \tag{51}$$

Then

$$\boldsymbol{H}(b+1) = \boldsymbol{F}_1(\boldsymbol{H}(b)) = \boldsymbol{F}_1(b, \boldsymbol{G}(b)) = (b+1, b\,\boldsymbol{F}\,\boldsymbol{G}(b)).$$

Hence $\pi_0(\boldsymbol{H}(b+1)) = b+1$. This computation establishes the claim. In particular, the inductive hypothesis (51) has been proved correct. Hence the computation also establishes that $\boldsymbol{G}(b^+) = \pi_1(\boldsymbol{H}(b+1)) = b\,\boldsymbol{F}\,\boldsymbol{G}(b)$. 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 $\boldsymbol{H}$ is $x \mapsto (x, \boldsymbol{G}(x))$. □

5.5.3. Theorem. *In $\mathbb{N}$, for all a, b, and c,*

(i) $0 \leqslant a$,
(ii) $a < b \Leftrightarrow a + c < b + c$,
(iii) $a < b \Leftrightarrow a \cdot (c+1) < b \cdot (c+1)$,
(iv) $a = b \Leftrightarrow a \cdot (c+1) = b \cdot (c+1)$,
(v) $a \leqslant b \Leftrightarrow \exists x\ a + x = b$.

Proof. Exercise 5.15. □

5.5.4. Theorem 5.5.2 allows to make some standard definitions:
(i) The operation $x \mapsto x!$ on $\mathbb{N}$ is defined recursively by

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

(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:

$$\sum_{k=0}^{0} \boldsymbol{F}(k) = \boldsymbol{F}(0), \qquad \prod_{k=0}^{0} \boldsymbol{F}(k) = \boldsymbol{F}(0),$$
$$\sum_{k=0}^{n+1} \boldsymbol{F}(k) = \sum_{k=0}^{n} \boldsymbol{F}(k) + \boldsymbol{F}(n+1), \qquad \prod_{k=0}^{n+1} \boldsymbol{F}(k) = \prod_{k=0}^{n} \boldsymbol{F}(k) \cdot \boldsymbol{F}(n+1).$$

5.6. The integers and the rational numbers

5.6.1. In $\mathbb{N}$, if $a \leqslant b$, then (Theorem 5.5.3) there is a solution to the equation

$$a + x = b;$$

this solution is unique (Theorem 5.2.8 (iv)) and can be denoted by

$$b - a.$$

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 5.18).

5.6.2. 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 family $(\mathbb{N} \times \mathbb{N})/\boldsymbol{E}$ is denoted by

$$\mathbb{Z};$$

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

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

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 by

$$b - a.$$

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) $0 < a \mathbin{\&} 0 < b \Rightarrow 0 < a + b \mathbin{\&} 0 < a \cdot b$.

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 5.19. 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

$$\begin{gathered}
(a_0 + c_0, b_0 + d_0)\ \boldsymbol{E}\ (a_1 + c_1, b_1 + d_1),\\
(a_0 \cdot c_0 + b_0 \cdot d_0, b_0 \cdot c_0 + a_0 \cdot d_0)\ \boldsymbol{E}\ (a_1 \cdot c_1 + b_1 \cdot d_1, b_1 \cdot c_1 + a_1 \cdot d_1), \qquad (52)\\
b_0 + c_0 < a_0 + d_0 \Leftrightarrow b_1 + c_1 < a_1 + d_1.
\end{gathered}$$

For (52), consider $(a_1 \cdot c_0 + b_1 \cdot d_0, b_1 \cdot c_0 + a_1 \cdot d_0)$. □

5.6.3. 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 5.5.3 (iv) and can be denoted by

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

or by b/a; also a is a **divisor** of b. If b/a and d/c exist in $\mathbb{Z}$, then $b/a = d/c \Leftrightarrow a \cdot d = b \cdot c$ (Exercise 5.20).

5.6.4. 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 family $((\mathbb{Z} \smallsetminus \{0\}) \times \mathbb{Z})/\boldsymbol{E}$ is denoted by

$$\mathbb{Q};$$

*its elements are the **rational numbers.** Let the $\boldsymbol{E}$-class $(a, b)\boldsymbol{E}$ be denoted by*

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

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:

$$\begin{aligned}
\frac{b}{a} + \frac{d}{c} &= \frac{a \cdot d + b \cdot c}{ac},\\
-\frac{b}{a} &= \frac{-b}{a},\\
\left(\frac{b}{a}\right)^{-1} &= \frac{a}{b} \qquad (\textit{if } b \neq 0),\\
\frac{b}{a} \cdot \frac{d}{c} &= \frac{b \cdot d}{a \cdot c},\\
\frac{b}{a} < \frac{d}{c} &\Leftrightarrow a \cdot b \cdot c \cdot c < a \cdot a \cdot c \cdot d.
\end{aligned}$$

*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 5.21. □

Exercises

5.1. Write down bijections from $(\boldsymbol{C} \times \boldsymbol{C}) \times \boldsymbol{C}$ to $\boldsymbol{C} \times (\boldsymbol{C} \times \boldsymbol{C})$ and $\boldsymbol{C}^3$.

5.2. 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 (17) holds in this case. Define an ordered triple $(\boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E})$ of classes.

5.3. Prove that an iterative structure admits induction if and only if every embedding into it is a surjection onto the universe (¶ 5.1.6).

5.4. Prove Theorem 5.2.1 by means of the corollary (¶ 4.4.6) to the Recursion Theorem.

5.5. Prove Theorem 5.2.8.

5.6. Find a recursive structure on which addition is not cancellative.

5.7. Prove Theorem 5.3.1.

5.8. Prove Lemma 5.3.4.

5.9. Prove Theorem 5.3.5.

5.10. Find a recursive structure $(\boldsymbol{A}, i, \boldsymbol{S})$ where $\boldsymbol{A}$ has elements a, b, and c, all different from i, such that $a \cdot c = b \cdot c$, but $a \neq b$.

5.11. Prove Theorem 5.4.1.

5.12. Prove Theorem 5.4.3.

5.13. Find a recursive structure in which exponentiation cannot be defined as a binary operation.

5.14. Show that there is a unique binary operation $(x, y) \mapsto \binom{x}{y}$ on $\mathbb{N}$ such that $\binom{a}{0} = 1$ and $\binom{0}{b+1} = 0$ and $\binom{a+1}{b+1} = \binom{a}{b} + \binom{a}{b+1}$. Can such an operation be defined on any structures that admit induction, but are *not* arithmetic?

5.15. Prove Theorem 5.5.3.

5.16. Prove $1 + a \cdot b \leqslant (1 + a)^b$ in $\mathbb{N}$.

5.17. Prove $3 < a \Rightarrow a^2 < 2^a$ in $\mathbb{N}$.

5.18. Prove the claims in ¶ 5.6.1.

5.19. Prove Theorem 5.6.2. $a \neq 0$.

5.20. Prove the claims in ¶ 5.6.3.

5.21. Prove Theorem 5.6.4.

CHAPTER 6

Ordinality

6.1. Ordinals

6.1.1. A class is called **transitive** if it *includes* each of its elements. That is, the class $\boldsymbol{C}$ is transitive if (and only if) $a \in \boldsymbol{C} \Rightarrow a \subseteq \boldsymbol{C}$, or equivalently

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

Compare with ¶ 2.4.6: transitivity of *classes* must be distinguished from transitivity of *relations* on classes. Consider for example the relation $\in$ of membership:

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

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

Extending an operation on ω (¶ 4.7.7) to all of $\mathbf{V}$, we say that the **(set-theoretic) successor** of a set a is a', where

$$a' = a \cup \{a\}.$$

6.1.2. Lemma. *Every transitive set includes the set-theoretic successor of each of its elements. The set-theoretic 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. □

6.1.3. A set is an **ordinal (number)** if it is transitive and also well-ordered by membership. The class of ordinals is denoted by

$$\mathbf{ON}.$$

We shall generally denote arbitrary elements of this class by letters from the beginning of the Greek alphabet: α, β, γ, δ, and so on.

6.1.4. Theorem. **ON** *contains* 0 *and is closed under set-theoretic succession. In particular,* $\omega \subseteq \mathbf{ON}$.

Proof. Vacuously, 0 is transitive and well-ordered by membership. Suppose $\alpha \in \mathbf{ON}$. Then α' is transitive by Lemma 6.1.2. Also, since α is well-ordered by membership, so is α'; indeed, α is the $\in$-greatest member. □

6.1.5. Theorem. **ON** *is transitive and well-ordered by membership. With respect to this ordering, the least element is* 0*, and succession is set-theoretic succession* $(\alpha^+ = \alpha')$. *On* **ON** *and on any ordinal, membership is proper inclusion.*

PROOF. For transitivity of $\mathbf{ON}$, supposing $\alpha \in \mathbf{ON}$, we show $\alpha \subseteq \mathbf{ON}$. To do this, say $\beta \in \alpha$; we show $\beta \in \mathbf{ON}$. But $\beta \subseteq \alpha$, since α is transitive; therefore β, like α, is well-ordered by membership. To show β is transitive, suppose $\gamma \in \beta$. Then $\gamma \in \alpha$, so $\gamma \subseteq \alpha$. Suppose $\delta \in \gamma$; then $\delta \in \alpha$. Since membership is transitive on α, from $\delta \in \gamma$ and $\gamma \in \beta$, we can conclude $\delta \in \beta$. Thus $\gamma \subseteq \beta$. Therefore β is transitive, so it is an ordinal. Thus $\mathbf{ON}$ is transitive.

Since the members of $\mathbf{ON}$ are transitive, membership is a transitive relation on $\mathbf{ON}$. Indeed, if $\mathbf{ON}$ contains α, β, and γ, and $\alpha \in \beta$ and $\beta \in \gamma$, then $\beta \subseteq \gamma$, so $\alpha \in \gamma$. Since membership is irreflexive on an ordinal, it is so on $\mathbf{ON}$. That is, if α is in $\mathbf{ON}$, and $\beta \in \alpha$, then $\beta \notin \beta$, so $\beta \neq \alpha$. In particular, $\alpha \notin \alpha$. Thus membership strictly orders $\mathbf{ON}$.

Since 0 has no members, it is the least member of $\mathbf{ON}$. Also, every element of α' is either α itself or an element of α; so α' is minimal among ordinals that contain α.

Suppose α and β are arbitrary elements of $\mathbf{ON}$. By Theorem 4.6.8, we may assume there is an isomorphism f from α to an initial segment of β. By induction then, α must *be* an initial segment of β. Indeed, say $\gamma \in \alpha$, so $\operatorname{pred}(\gamma) = \gamma$. If $f \restriction \gamma = \mathrm{id}_\gamma$, then $f(\gamma) = \gamma$. So either $\alpha = \beta$ or else α is a section $\operatorname{pred}(\delta)$ of β; but in the latter case, $\alpha = \delta$, so $\alpha \in \beta$. So membership is a strict total ordering of $\mathbf{ON}$.

Every section of $\mathbf{ON}$ is an element of $\mathbf{ON}$; in particular, it is a set. If d is a subset of $\mathbf{ON}$ with an element α, then either α is the least element of d, or else the least element of d is the least element of $d \cap \alpha$. So $\mathbf{ON}$ is well-ordered by $\in$.

Finally, on $\mathbf{ON}$ and its members, membership implies *inclusion;* and this inclusion is proper because membership is irreflexive *on* $\mathbf{ON}$. Conversely, if the ordinal β properly includes the ordinal α, then α is a section of β, by transitivity of the latter; hence $\alpha \in \beta$ as just shown. □

6.1.6. COROLLARY (Burali-Forti Paradox). $\mathbf{ON}$ *is not a set, so it is not an ordinal.*

PROOF. Exercise 6.3. □

6.1.7. If ω is a proper class, then it is just $\mathbf{ON}$ (Exercise 6.4); if it is a set, then it is a limit in $\mathbf{ON}$ (¶ 4.6.4). In any case, ω is the class of ordinals that neither *are* limits nor *contain* limits (Exercise 6.5).

6.2. Order-types

6.2.1. By definition (¶ 6.1.3), every ordinal α is well-ordered by membership. We may then understand α to denote the structure $(\alpha, \in)$. Likewise, since $\mathbf{ON}$ is also well-ordered by membership (¶ 6.1.5), we may take $\mathbf{ON}$ to denote the structure $(\mathbf{ON}, \in)$. Finally, since for ordinals, $\alpha \in \beta \Leftrightarrow \alpha \subset \beta$ (¶ 6.1.5), we may use $\in$ and $\subset$ interchangeably in $\mathbf{ON}$; 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:*

$$\exists x\ x = \bigcup b.$$

6.2.4. Theorem. *The union of a set of ordinals is an ordinal, which is the supremum of the set:*

$$b \subset \mathbf{ON} \Rightarrow \bigcup b = \sup(b).$$

Proof. Let b be a set of ordinals. Ordinals are sets of ordinals, by transitivity of $\mathbf{ON}$ (¶ 6.1.5), so $\bigcup b$ is a subclass of $\mathbf{ON}$. Hence $\bigcup b$ is well-ordered by proper inclusion. Also $\bigcup b$ is a set by the Union Axiom. Let $\beta \in \bigcup b$. Then some element α of b contains β. But then α is transitive, so $\beta \subseteq \alpha \subseteq \bigcup b$. Therefore $\bigcup b$ is an ordinal.

Finally, if $\alpha \in b$, then $\alpha \subseteq \bigcup b$; so $\bigcup b$ is an upper bound of b. If $\beta < \bigcup b$, then β belongs to an element of b; that is, β is less than that element, so β is not an upper bound of b. □

6.2.5. Theorem. *If b is a set of ordinals, then $\bigcup\{x' \colon x \in b\}$ is the strict upper bound of b.*

Proof. Exercise 6.7. □

6.2.6. Theorem and Definition. *Every well-ordered set $\mathfrak{b}$ is uniquely isomorphic to a unique ordinal, denoted by*

$$\operatorname{ord}(\mathfrak{b})$$

and called the ***order-type*** *or the* ***ordinality*** *of $\mathfrak{b}$. Every well-ordered proper class is uniquely isomorphic to* $\mathbf{ON}$.

Proof. Theorem 4.6.8 and its proof show that every well-ordered set is isomorphic to an initial segment of $\mathbf{ON}$, since the latter is closed under succession. The initial segment must then be proper, since it is a set. Therefore it is an ordinal, as in the proof of Theorem 6.1.5.

Similarly, of two well-ordered proper classes, one is isomorphic to an initial segment of the other. But *proper* initial segments of well-ordered classes must be sets; therefore all well-ordered proper classes are isomorphic. □

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

$$(a, b) < (c, d) \Leftrightarrow b < d \vee (b = d \mathbin{\&} a < c).$$

This is a *total* ordering of $\boldsymbol{C} \times \boldsymbol{D}$ (Exercise 6.8; see also Figure 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

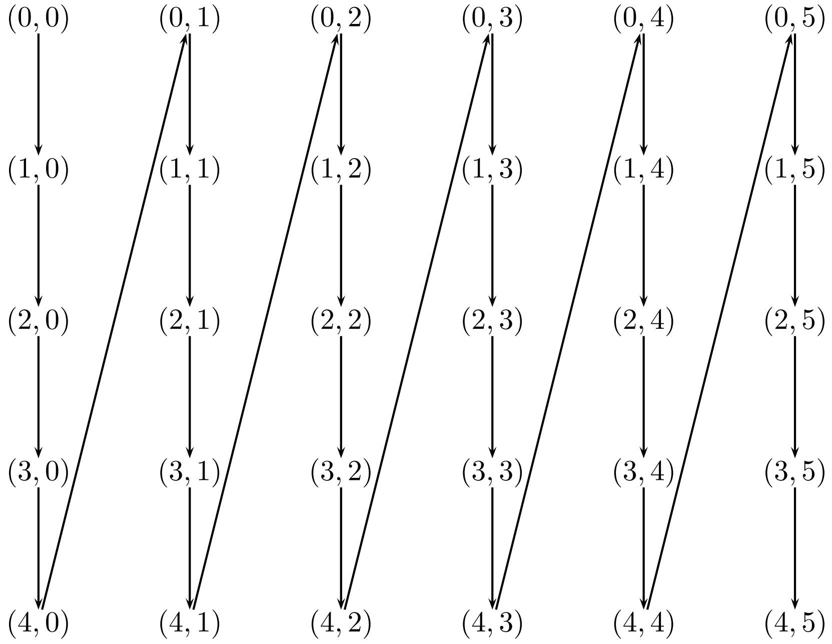


Figure 6.1. The lexicographic ordering of 5×6

$\delta \in \beta$; 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 ¶ 5.2.4 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 6.9. In (v), the solution is $\operatorname{ord}(\alpha \smallsetminus \beta)$; it is unique by (iv). □

6.3.4. Theorem. *Ordinal addition can be defined recursively by the method given by Theorem 4.6.6:*

(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 ¶ 5.2.1.

Proof. Exercise 6.10. □

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 Definition 5.3.1 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) the system $\alpha \cdot x + y = \beta \ \&\ y < \alpha$ *has a unique ordinal solution if* $0 < \alpha$*.*

Proof. Exercise 6.11. 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 can be defined recursively by the method given by Theorem 4.6.6:*

(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 ¶ 5.3.1.*

Proof. Exercise 6.12. □

6.4.4. Theorem. $0 < n < \omega \Rightarrow n \cdot \omega = \omega$.

Proof. Exercise 6.13. (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:

$$(1 + 1) \cdot \omega = 2 \cdot \omega = \omega < \omega + \omega = 1 \cdot \omega + 1 \cdot \omega;$$

also the ordering in Theorem 6.4.2 (v) cannot be made strict.

6.5. Ordinal exponentiation

6.5.1. The binary operation of **(ordinal) exponentiation** on **ON**, denoted by

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

is such that $0^0 = 1$ and $0^\beta = 0$ when $\beta \neq 0$, and, when $0 < \alpha$, is defined by the method given by Theorem 4.6.6:

(i) $\alpha^0 = 1$,
(ii) $\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$,
(iii) $\alpha^\beta = \bigcup\{\alpha^x \colon x \in \beta\}$ when β is a limit.

In particular, on ω, this operation is as given in ¶ 5.4.1.

6.5.2. Theorem. *If* $1 < \alpha$ *and* $\beta < \gamma$, *then* $\alpha^\beta < \alpha^\gamma$.

Proof. Exercise 6.14. □

6.5.3. 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.

6.5.4. Lemma. *The following operations on* **ON** *are normal:*

(i) $x \mapsto \alpha + x$;
(ii) $x \mapsto \alpha \cdot x$, *when* $0 < \alpha$;
(iii) $x \mapsto \alpha^x$, *when* $1 < \alpha$.

Proof. By Theorems 6.3.3 (iv), 6.4.2 (iv), and 6.5.2, the operations are endomorphisms. Hence they are normal by Theorems 6.3.4 (iii) and 6.4.3 (iii) and ¶ 6.5.1. □

6.5.5. 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 $\beta < \gamma < \alpha$ 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 $\boldsymbol{F}$. □

6.5.6. Theorem. *For all ordinals α, β, and γ,*

(i) $1^\alpha = 1$,
(ii) $\alpha^1 = \alpha$,
(iii) $\alpha^{\beta+\gamma} = \alpha^\beta \cdot \alpha^\gamma$,
(iv) $\alpha^{\beta\cdot\gamma} = (\alpha^\beta)^\gamma$.

Proof. Exercise 6.16. For (iii), by Lemma 6.5.5, if $1 < \alpha$ and γ is a limit, one has by induction

$$\begin{aligned}\alpha^\beta \cdot \alpha^\gamma &= \alpha^\beta \cdot \sup(\{\alpha^x \colon x \in \gamma\})\\ &= \sup(\{\alpha^\beta \cdot \alpha^x \colon x \in \gamma\})\\ &= \sup(\{\alpha^{\beta+x} \colon x \in \gamma\})\\ &= \alpha^{\sup(\{\beta+x\colon x\in\gamma\})}\\ &= \alpha^{\beta+\sup(\{x\colon x\in\gamma\})} = \alpha^{\beta+\gamma}.\end{aligned}$$

Likewise for (iv). □

6.5.7. We now have the following initial segment of $\mathbf{ON}$:

$$\{0,1,\dots;\omega,\omega+1,\dots,\omega\cdot 2,\dots;\omega^2,\omega^2+1,\dots;\omega^2+\omega,\dots;$$
$$\omega^2\cdot 2,\dots;\omega^3,\dots;\omega^\omega,\dots;\omega^{\omega\cdot 2},\dots;\omega^{\omega^2},\dots;\omega^{\omega^\omega},\dots;\omega^{\omega^{\omega^\omega}},\dots\}.$$

This set is closed under the operations that we have defined so far. The supremum of the set is a solution of the equation

$$\omega^x = x \tag{53}$$

(Exercise 6.17); in fact it is the least solution and so is denoted by

$$\epsilon_0.$$

6.6. Base omega

6.6.1. Theorem. $\alpha \leqslant \omega^\alpha$.

Proof. The claim holds by definition of exponentiation (¶ 6.5.1) when $\alpha = 0$. Suppose it holds when $\alpha = \beta$. Then

$$\omega^{\beta+1} = \omega^\beta \cdot \omega^1 = \omega^\beta \cdot \omega > \omega^\beta \cdot 2 = \omega^\beta + \omega^\beta \geqslant \omega^\beta + \omega^0 = \omega^\beta + 1 \geqslant \beta + 1.$$

Finally, if β is a limit, and the claim holds when $\alpha < \beta$, then it holds when $\alpha = \beta$ by normality. □

6.6.2. Theorem. *Suppose $\alpha > 0$. There are, uniquely, a natural number n, an $(n+1)$-tuple $(\beta_0, \dots, \beta_n)$ of ordinals, and an $(n+1)$-tuple $(c_0, \dots, c_n)$ of non-zero natural numbers, such that*

$$\beta_0 > \dots > \beta_n, \qquad \alpha = \omega^{\beta_0} \cdot c_0 + \dots + \omega^{\beta_n} \cdot c_n.$$

Proof. Uniqueness follows from observing

$$\omega^{\beta_0} \leqslant \omega^{\beta_0} \cdot c_0 + \dots + \omega^{\beta_n} \cdot c_n < \omega^{\beta_0+1}.$$

For existence, we proceed recursively. Assuming $(\beta_0, \dots, \beta_{k-1})$ and $(c_0, \dots, c_{k-1})$ have been defined, we consider the inequality

$$\sum_{i<k} \omega^{\beta_i} \cdot c_i + \omega^x \leqslant \alpha.$$

If this has no solution, then $n = k - 1$, and we are done. If there is a solution, let β_k be the supremum of the set of solutions; then this is a solution by normality. Now let c_k be the maximal solution of

$$\sum_{i<k} \omega^{\beta_i} \cdot c_i + \omega^{\beta_k} \cdot x \leqslant \alpha.$$

By induction, we have $\beta_0 > \beta_1 > \cdots$, so the construction must terminate in finitely many steps. □

6.6.3. Lemma. *If $\alpha > 0$, then $1 + \omega^\alpha = \omega^\alpha$.*

Proof. Use induction. The claim holds when $\alpha = 1$ by Theorem 6.3.8. Suppose it holds when $\alpha = \beta$. Then

$$\begin{aligned} \omega^{\beta+1} \leqslant 1 + \omega^{\beta+1} &\leqslant \omega^\beta + \omega^{\beta+1} && \text{[Theorem 6.3.6]} \\ &= \omega^\beta \cdot 1 + \omega^\beta \cdot \omega && \text{[Theorems 6.4.2 and 6.5.6]} \\ &= \omega^\beta \cdot (1 + \omega) = \omega^\beta \cdot \omega = \omega^{\beta+1}, \end{aligned}$$

so the claim holds when $\alpha = \beta + 1$. The claim holds at limits by normality of exponentiation. □

6.6.4. Theorem. *If $\beta < \alpha$, and $n \in \omega$, then $\omega^\beta \cdot n + \omega^\alpha = \omega^\alpha$.*

Proof. Exercise 6.18. □

6.6.5. Theorem. *If $\beta + \alpha = \alpha$, then $\beta \leqslant \alpha$ and*

$$(\alpha + \beta) \cdot \gamma = \begin{cases} \alpha \cdot \gamma + \beta, & \text{if } \gamma \text{ is a successor;} \\ \alpha \cdot \gamma, & \text{if } \gamma \text{ is a limit or } 0. \end{cases}$$

Proof. For the first claim, note that if $\alpha < \beta$, then $\alpha < \beta \leqslant \beta + \alpha$. For the second claim, use induction. The claim holds when $\gamma = 0$ by definition of multiplication. Suppose it holds when $\gamma = \delta$; then

$$(\alpha + \beta) \cdot (\delta + 1) = (\alpha + \beta) \cdot \delta + \alpha + \beta = \alpha \cdot \delta + \alpha + \beta = \alpha \cdot (\delta + 1) + \beta,$$

so the claim holds when $\gamma = \delta + 1$. Finally, since

$$\alpha \cdot \delta + \beta \leqslant \alpha \cdot \delta + \alpha = \alpha \cdot (\delta + 1),$$

by normality the claim holds at limits. □

6.6.6. In short, ordinals can be written in **base** ω, and when they are so written, we can add and multiply them. For example,

$$\begin{aligned}
&(\omega^{\omega+1} \cdot 3 + \omega^6 \cdot 4 + 1) \cdot (\omega^{\omega^2} \cdot 2 + 3) \\
&= (\omega^{\omega+1} \cdot 3 + \omega^6 \cdot 4 + 1) \cdot (\omega^{\omega^2} \cdot 2) + (\omega^{\omega+1} \cdot 3 + \omega^6 \cdot 4 + 1) \cdot 3 \\
&= (\omega^{\omega+1} \cdot 3) \cdot (\omega^{\omega^2} \cdot 2) + (\omega^{\omega+1} \cdot 3) \cdot 3 + \omega^6 \cdot 4 + 1 \\
&= \omega^{\omega+1} \cdot (3 \cdot \omega^{\omega^2}) \cdot 2 + \omega^{\omega+1} \cdot (3 \cdot 3) + \omega^6 \cdot 4 + 1 \\
&= (\omega^{\omega+1} \cdot \omega^{\omega^2}) \cdot 2 + \omega^{\omega+1} \cdot 9 + \omega^6 \cdot 4 + 1 \\
&= (\omega^{\omega+1+\omega^2}) \cdot 2 + \omega^{\omega+1} \cdot 9 + \omega^6 \cdot 4 + 1 \\
&= (\omega^{\omega^2}) \cdot 2 + \omega^{\omega+1} \cdot 9 + \omega^6 \cdot 4 + 1.
\end{aligned}$$

Exercises

6.1. Prove that

$$a' = \bigcap\{x\colon a \in x \mathbin{\&} a \subseteq x\}. \tag{54}$$

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

6.3. Prove the Burali-Forti Paradox (¶ 6.1.6).

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

6.5. Show that ω is the class of ordinals that neither are limits nor contain limits.

6.6. Prove that an ordinal α is 0 or a limit if and only if $\alpha = \bigcup \alpha$.

6.7. Prove Theorem 6.2.5.

6.8. Prove that the lexicographic ordering is total (¶ 6.3.1).

6.9. Prove Theorem 6.3.3.

6.10. Prove Theorem 6.3.4.

6.11. Prove Theorem 6.4.2.

6.12. Prove Theorem 6.4.3.

6.13. Prove Theorem 6.4.4.

6.14. Prove Lemma 6.5.2.

6.15. Prove $1 < \alpha \Rightarrow \beta \leqslant \alpha^\beta$.

6.16. Prove Theorem 6.5.6.

6.17. Defining $(a_n\colon n \in \omega)$ by $a_0 = 0$ and $a_{n+1} = \omega^{a_n}$, show that $\sup\{a_n\colon n \in \omega\}$ is closed under the operations of addition, multiplication, and exponentiation, and is the least solution of (53).

6.18. Prove Theorem 6.6.4.

6.19. Compute the following in base ω:

(i) $(\omega + 1) + (\omega + 2)$;
(ii) $(\omega^{\omega^\omega} + \omega^2 \cdot 3 + 4) + (\omega^{\omega^2} + \omega^\omega \cdot 5 + 1)$;
(iii) $(\omega + 1) \cdot (\omega + 2)$;
(iv) $(\omega^2 \cdot 3 + 2) \cdot (\omega^\omega + 5)$

6.20. Compute the following in base ω:

(i) $(\omega + n)^2$;
(ii) $(\omega + n)^3$;
(iii) $(\omega + n)^m$;
(iv) $(\omega + n)^\omega$;
(v) $(\omega + n)^{\omega+m}$.

CHAPTER 7

Cardinality

7.1. Finite sets

7.1.1. A set a is **finite** if it belongs to every subset of $\mathscr{P}(a)$ that

(i) contains $\varnothing$, and

(ii) if closed under the operation $x \mapsto x \cup \{b\}$, for every b in a.

(If one wants a word, one may refer to such a subset of $\mathscr{P}(a)$ as **inductive.**) We shall investigate the relation between finite sets and infinite sets as defined in ¶ 4.2.1.

7.1.2. Lemma. *$\varnothing$ is finite, and if a is finite, then so is $a \cup \{b\}$.*

Proof. Since $\varnothing$ belongs to every subset of $\mathscr{P}(\varnothing)$ that contains $\varnothing$, it is finite.

Suppose a finite, and c is subset of $\mathscr{P}(a \cup \{b\})$ that contains $\varnothing$ and is closed under $x \mapsto x \cup \{d\}$ for every d in $a \cup \{b\}$. Then $c \cap \mathscr{P}(a)$ contains $\varnothing$ and is closed under $x \mapsto x \cup \{d\}$ whenever $d \in a$. Therefore $a \in c \cap \mathscr{P}(a)$, since a is finite; hence $a \in c$, so $a \cup \{b\} \in c$. Therefore $a \cup \{b\}$ is finite. □

7.1.3. Theorem (Induction). *A class of finite sets that contains $\varnothing$ and is closed under each operation $x \mapsto x \cup \{b\}$ contains all finite sets.*

Proof. Let $\boldsymbol{C}$ be such a class, and let a be a finite set. Then $a \in \boldsymbol{C} \cap \mathscr{P}(a)$ (since this set is inductive in the sense of ¶ 7.1.1). □

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 $a \cup \{b\} \subseteq \operatorname{dom}(\boldsymbol{F})$. Then $\boldsymbol{F}[\,a\,]$ is finite, and $\boldsymbol{F}[\,a \cup \{b\}\,] = \boldsymbol{F}[\,a\,] \cup \{\boldsymbol{F}(b)\}$, so this is also finite by Lemma 7.1.2. 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, $f\colon a' \rightarrowtail\!\!\!\!\twoheadrightarrow b' \Leftrightarrow g\colon a \rightarrowtail\!\!\!\!\twoheadrightarrow b$.

Proof. If f is a bijection, then

$$\begin{aligned}\operatorname{rng}(g) &= \{f(x)\colon x \in a \mathrel{\&} f(x) \neq b\} \cup \{f(a)\colon x \in a \mathrel{\&} f(x) = b\}\\ &= f[\,a \smallsetminus \{f^{-1}(b)\}\,] \cup f[\,\{a\} \smallsetminus \{f^{-1}(b)\}\,]\\ &= f[\,a' \smallsetminus \{f^{-1}(b)\}\,]\\ &= f[\,a'\,] \smallsetminus \{b\}\\ &= b,\end{aligned}$$

so g is surjective. If g is a bijection, then

$$\operatorname{rng}(f) = \operatorname{rng}(g) \cup f[\,f^{-1}[\,\{b\}\,]\,] = b \cup \{b\} = b',$$

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 $a \notin \operatorname{dom}(g)$, 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 infinite.*

Proof. By Theorem 7.1.5, it is enough to show that no natural number is 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

$$a \in \omega \mathrel{\&} b \in \omega \mathrel{\&} a + b \approx a \Rightarrow b = 0.$$

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

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

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 cardinals compose the class denoted by

$$\mathbf{CN};$$

this is a subclass 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 7.2).

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 $\operatorname{ord}(c,<) = \beta$, and $\gamma < \beta$, then $\operatorname{ord}(d,<) = \gamma$ for some section d of c. This shows that $\{\operatorname{ord}(\mathfrak{c})\colon \mathfrak{c} \in b\}$ 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 an ordinal α such that $\kappa < \alpha$, but $\kappa \not\approx \alpha$. Therefore $\kappa < \operatorname{card}(\alpha)$. 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 an ordinal, namely $\sup(a)$, by Theorem 6.2.4. Hence, if κ is a cardinal less than $\sup(a)$, then $\kappa < \lambda$ for some λ in a, and therefore $\kappa \neq \operatorname{card}(\sup(a))$. Therefore $\sup(a)$ must be a cardinal (namely its own cardinality). □

7.2.5. The function

$$x \mapsto \aleph_x$$

from **ON** into **CN** is given (as in Theorem 4.6.6) 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.

Here $\aleph$ is *aleph,* the first letter of the Hebrew alphabet, as mentioned in ¶ 1.3.6. If $0 < \alpha$, then $\aleph_\alpha$ is called **uncountable.**

7.2.6. Theorem. *The function $x \mapsto \aleph_x$ is an isomorphism between* **ON** *and the class of infinite cardinals.*

Proof. Exercise 7.3. □

7.3. Cardinal addition and multiplication

7.3.1. The **(cardinal) sum** and **(cardinal) product** of two cardinals κ and λ are defined by

$$\begin{aligned}\kappa + \lambda &= \operatorname{card}((\kappa \times \{0\}) \cup (\lambda \times \{1\})),\\ \kappa \cdot \lambda &= \operatorname{card}(\kappa \times \lambda).\end{aligned}$$

If κ and λ are merely cardinalities, not ordinals, then

$$\kappa + \lambda = \operatorname{card}((a \times \{0\}) \cup (b \times \{1\})), \tag{55}$$

$$\kappa \cdot \lambda = \operatorname{card}(a \times b), \tag{56}$$

where $a \in \kappa$ and $b \in \lambda$ (Exercise 7.4). 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. □

7.3.3. LEMMA. *The class* **ON** *of ordinals becomes isomorphic to* **ON** × **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 Figure 7.1.)

PROOF. Exercise 7.8. 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

$$\gamma = \min\{y\colon (\beta, y) \in a\},$$
$$\beta = \min\{x\colon \exists y\ (x, y) \in a \;\&\; \max(x, y) = \alpha)\},$$
$$\alpha = \min\{\max(x, y)\colon (x, y) \in a\}.$$

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*

$$\kappa + \lambda = \kappa \cdot \lambda = \lambda.$$

In particular,

$$\aleph_\alpha + \aleph_\beta = \aleph_\alpha \cdot \aleph_\beta = \aleph_{\max(\alpha,\beta)}.$$

$$\begin{tikzcd}
(0,0) \arrow[r] & (0,1) \arrow[dl] & (0,2) \arrow[d] & (0,3) \arrow[d] & (0,4) \arrow[d] & (0,5) \arrow[d] \\
(1,0) \arrow[r] & (1,1) \arrow[ur] & (1,2) \arrow[dll] & (1,3) \arrow[d] & (1,4) \arrow[d] & (1,5) \arrow[d] \\
(2,0) \arrow[r] & (2,1) \arrow[r] & (2,2) \arrow[uur] & (2,3) \arrow[dlll] & (2,4) \arrow[d] & (2,5) \arrow[d] \\
(3,0) \arrow[r] & (3,1) \arrow[r] & (3,2) \arrow[r] & (3,3) \arrow[uuur] & (3,4) \arrow[dllll] & (3,5) \arrow[d] \\
(4,0) \arrow[r] & (4,1) \arrow[r] & (4,2) \arrow[r] & (4,3) \arrow[r] & (4,4) \arrow[uuuur] & (4,5)
\end{tikzcd}$$

FIGURE 7.1. **ON** × **ON**, well-ordered

PROOF. We shall show

$$\lambda \cdot \lambda = \lambda. \tag{57}$$

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 (57) 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 $\mathbf{ON} \times \mathbf{ON}$ onto $\mathbf{ON}$ guaranteed by Lemma 7.3.3. As $\alpha \times \alpha$ is always a section (namely, $\operatorname{pred}(0, \alpha)$) of $\mathbf{ON} \times \mathbf{ON}$, so $\boldsymbol{F}[\,\alpha \times \alpha\,]$ must be a section of $\mathbf{ON}$: that is, $\boldsymbol{F}[\,\alpha \times \alpha\,]$ is an ordinal. Suppose $\boldsymbol{F}[\,\mu \times \mu\,] = \alpha$. Then

$$\mu \leqslant \mu \cdot \mu = \operatorname{card}(\mu \times \mu) = \operatorname{card}(\alpha) \leqslant \alpha.$$

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 7.5), the successor δ of $\max(\beta, \gamma)$ is also less than μ. Hence

$$\nu \in \boldsymbol{F}[\,\delta \times \delta\,],$$
$$\nu \subseteq \boldsymbol{F}[\,\delta \times \delta\,],$$
$$\nu \leqslant \operatorname{card}(\delta \times \delta) = \operatorname{card}(\delta) \cdot \operatorname{card}(\delta)$$

(Exercise 7.6). 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) $\kappa^{\lambda+\mu} = \kappa^{\lambda} \cdot \kappa^{\mu}$,
(vi) $\kappa^{\lambda \cdot \mu} = (\kappa^{\lambda})^{\mu}$,
(vii) $\kappa \leqslant \mu \;\&\; \lambda \leqslant \nu \Rightarrow \kappa^{\lambda} \leqslant \mu^{\nu}$.

Proof. Exercise 7.9. □

7.4.3. Theorem. *For all sets a,*

$$\mathrm{card}(\mathscr{P}(a)) = 2^{\mathrm{card}(a)}; \tag{58}$$

hence for all cardinalities κ,

$$\kappa < 2^{\kappa}. \tag{59}$$

Proof. There is a bijection between ${}^{a}2$ and $\mathscr{P}(a)$ that takes the function f to the set $\{x \in a \colon f(x) = 1\}$. This establishes (58). Then (59) follows by Cantor's Theorem (¶ 4.1.5). □

7.4.4. Corollary. *If κ and λ are cardinals such that $2 \leqslant \kappa \leqslant 2^{\lambda}$ 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. The function

$$x \mapsto \beth_x$$

from **ON** into **CN** is given recursively (as in Theorem 4.6.6) by

(i) $\beth_0 = \omega$,
(ii) $\beth_{\alpha'} = 2^{\beth_{\alpha}}$,
(iii) $\beth_{\alpha} = \bigcup\{\beth_x \colon x \in \alpha\}$ if α is a limit.

Here $\beth$ is *beth,* the second letter of the Hebrew alphabet.

7.4.6. Theorem. *The function $x \mapsto \beth_x$ is an embedding of* **ON** *in* **CN**, *and*

$$\aleph_{\alpha} \leqslant \beth_{\alpha}. \tag{60}$$

Proof. Theorem 7.4.3 and induction. □

7.4.7. The letter $\beth$ is *beth,* the second letter of the Hebrew alphabet. The **Continuum Hypothesis**, or GCH, is that $\aleph_1 = \beth_1$; the **Generalized Continuum Hypothesis**is that $\aleph_\alpha = \beth_\alpha$ for all ordinals α. Here the **continuum** is the set of *real numbers,* defined in §7.7. We shall see in Chapter 8 that we *can* make these hypotheses without contradicting our other axioms about sets. We shall not see what is also the case, that these hypotheses are not *implied* by our axioms [**2**].

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

$$b \in \mathscr{P}(a) \smallsetminus \{\varnothing\} \Rightarrow f(b) \in b.$$

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 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 $\mathbf{ON}$ defined recursively by

$$\boldsymbol{G}(\alpha) = f(a \smallsetminus \boldsymbol{G}[\alpha]).$$

Then $\boldsymbol{G}^{-1}[a]$ is an initial segment of $\mathbf{ON}$, and $\boldsymbol{G} \restriction \boldsymbol{G}^{-1}[a]$ 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. *Assume a has a choice-function. If $(a,<)$ is an order such that every totally ordered subset of a has an upper bound in a, then a has a maximal element.*

Proof. Let f be the operation on $\mathscr{P}(a)$ taking each element b to 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 $\mathbf{ON}$ by

$$\boldsymbol{H}(\alpha) = g(f(a \cap \boldsymbol{H}[\alpha])).$$

In particular, if $\boldsymbol{H}[\alpha] \subset a$ and has a strict upper bound, then $\boldsymbol{H}(a)$ is such an upper bound. So $\boldsymbol{H}$ embeds $\boldsymbol{H}^{-1}[a]$ in a, and $a \cap \operatorname{rng}(\boldsymbol{H})$ 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. Theorem. *For a set a, let b be the set of choice-functions for subsets of a. Then b is ordered by inclusion, and every totally ordered subset c of b is bounded above by $\bigcup c$. If b has a maximal element f, then f is a choice-function for a.*

Proof. Suppose $f \in b$, but is a choice-function only for a proper subset d of a. Then $a \smallsetminus d$ has an element e, and $f \subset g$, where g is the choice-function for $d \cup \{e\}$ that agrees with f on $\mathscr{P}(d) \smallsetminus \{\varnothing\}$, but takes x to e if $e \in x$. Thus f is not a maximal element of b. □

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

7.5.7. The Axiom of Choice, or AC, is a completely new kind of axiom, since it asserts the existence of sets that we do not already have as classes. However, as with the Generalized Continuum Hypothesis (¶ 7.4.7), so we shall see in Chapter 8 that we *can* assume the Axiom of Choice without contradicting our other axioms. The Axiom of Choice is convenient for mathematics in that it allows many theorems to be proved. Often it does this through (the conclusion of) Theorem 7.5.4, which is often known as Zorn's Lemma (though it appears to be due to Hausdorff). Then Theorem 7.5.5 shows that assuming the Axiom of Choice is *equivalent* to assuming Zorn's Lemma (this equivalence 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. The ordinal power α^β is not defined as the order-type of $({}^\beta\alpha, <)$ for some $<$. Indeed, when $1 < \alpha$ and $\omega \leqslant \beta$, there is no obvious relation that well-orders ${}^\beta\alpha$. Thinking that

$${}^\beta\alpha = \overbrace{\alpha \times \alpha \times \cdots}^{\beta},$$

we can define the left lexicographic ordering $<_\ell$ on ${}^\beta\alpha$, where, if $\gamma = \min\{x \in \beta\colon f(x) \neq g(x)\}$, and $f(\gamma) < g(\gamma)$, then $f <_\ell g$. But then $<_\ell$ does not well-order ${}^\beta\alpha$, since $\{f_n\colon n \in \omega\}$ has no least element when

$$f_n(x) = \begin{cases}1, & \text{if } x = n;\\ 0, & \text{otherwise.}\end{cases}$$

There is also a right lexicographic ordering $<_{\mathrm{r}}$, where, if $f(\delta) \leqslant g(\delta)$ when $\gamma < \delta < \beta$, and $f(\gamma) < g(\gamma)$, then $f <_{\mathrm{r}} g$. But this ordering of ${}^{\beta}\alpha$ is not generally total (Exercise 7.14).

7.6.3. If $f \in {}^{\beta}\alpha$, let the **support** of f be the set $\{x \in \beta\colon f(x) \neq 0\}$; this can be denoted by

$$\operatorname{supp}(f).$$

7.6.4. Theorem. ${}^{\beta}\alpha = \operatorname{ord}(\{x \in {}^{\beta}\alpha\colon \operatorname{card}(\operatorname{supp}(x)) < \omega\}, <_{\mathrm{r}})$.

Proof. Exercise 7.15. See Figure 7.2. □

7.6.5. 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*

$$\operatorname{card}\Bigl(\bigcup\{a_x\colon x \in b\}\Bigr) = \operatorname{card}(b) \cdot \sup\{\operatorname{card}(a_x)\colon x \in b\}.$$

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. □

7.6.6. 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. 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 7.12.) Then we may let $f_c = g(\{x \in e\colon \operatorname{dom}(x) = a_c\})$.

7.6.7. Lemma. *If a is an infinite set, then*

$$\operatorname{card}(\{x \in \mathscr{P}(a)\colon \operatorname{card}(x) < \aleph_0\}) = \operatorname{card}(a).$$

Proof. Exercise 7.13. Use Lemma 7.6.5. □

7.6.8. 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

$$\operatorname{card}(\alpha^{\beta}) = \operatorname{card}(\{x \in {}^{\beta}\alpha\colon \operatorname{card}(\operatorname{supp}(x)) < \aleph_0\}) = \operatorname{card}\Bigl(\bigcup\{{}^{x}(\alpha \smallsetminus \{0\})\colon x \in b\}\Bigr).$$

If α is infinite and b is finite, then $\operatorname{card}({}^{b}(\alpha \smallsetminus \{0\})) = \operatorname{card}(\alpha)$ (Exercise 7.10). If β is infinite, then $\operatorname{card}(b) = \operatorname{card}(\beta)$ by Lemma 7.6.7. In either case, we are done by Lemma 7.6.5. □

$$\begin{array}{l}
(0,0,0,0,\dots) \\
(1,0,0,0,\dots) \\
(2,0,0,0,\dots) \\
\dots\dots\dots\dots \\
(0,1,0,0,\dots) \\
(1,1,0,0,\dots) \\
(2,1,0,0,\dots) \\
\dots\dots\dots\dots \\
(0,0,1,0,\dots) \\
(1,0,1,0,\dots) \\
\dots\dots\dots\dots \\
(0,1,1,0,\dots) \\
\dots\dots\dots\dots
\end{array}$$

FIGURE 7.2. The ordering of $\{x \in {}^{\beta}\alpha\colon \operatorname{card}(\operatorname{supp}(x)) < \omega\}$.

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 $\{x \in \mathbb{Q}\colon x^2 < 2 \lor x < 0\}$ 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 Figure 7.3. (If a is a cut of $\mathbb{Q}$, then the ordered pair $(a, \mathbb{Q} \smallsetminus a)$ 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 $\mathbb{R}$ so that $x \mapsto \operatorname{pred}(x)$ is an embedding of ordered fields. First,

$$\begin{aligned}
a + b &= \{x + y\colon x \in a \mathbin{\&} y \in b\}, \\
-a &= \{-x\colon x \in \mathbb{Q} \smallsetminus a \mathbin{\&} \exists y\ (y \in \mathbb{Q} \smallsetminus a \mathbin{\&} y < x)\}.
\end{aligned}$$

Multiplication is defined by cases. First, if $0 \in a \cap b$ (so that a and b are *positive* real numbers), then

$$a \cdot b = \{x \cdot y\colon x \in a \mathbin{\&} y \in b\}.$$

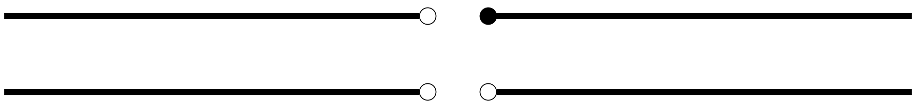


FIGURE 7.3. Cuts

If $a = 0^{\mathbb{R}}$ or $b = 0^{\mathbb{R}}$ (that is, if $a = \operatorname{pred}(0)$ or $b = \operatorname{pred}(0)$), 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 $x \mapsto \operatorname{pred}(x)$ is an embedding of ordered fields (Exercise 7.16).

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 7.17). 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 7.11) $\operatorname{card}(\mathbb{Z}) = \aleph_0$. We have a surjection of $\mathbb{Z} \times (\mathbb{Z} \smallsetminus \{0\})$ onto $\mathbb{Q}$, so $\operatorname{card}(\mathbb{Q}) = \aleph_0$. (See also Exercise 7.18.) Since $\mathbb{R} \subseteq \mathscr{P}(\mathbb{Q})$, we have

$$\operatorname{card}(\mathbb{R}) \leqslant 2^{\aleph_0}. \tag{61}$$

7.7.5. There is an embedding of ${}^{\omega}2$ into $\mathbb{R}$, so that $2^{\aleph_0} \leqslant \operatorname{card}(\mathbb{R})$; because of (61), by the Schroeder–Bernstein Theorem (¶ 4.1.3), we then have

$$\operatorname{card}(\mathbb{R}) = 2^{\aleph_0}.$$

To define the embedding, replace $\mathbb{Q}$ with its image in $\mathbb{R}$; that is, treat $\mathbb{Q}$ as a subfield of $\mathbb{R}$. Let $f\colon {}^{\omega}2 \to \mathbb{R}$, where

$$f(\sigma) = \sup\Bigl\{\sum_{k=0}^{x} \frac{2 \cdot \sigma(k)}{3^{k+1}} : x \in \omega\Bigr\}.$$

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 Figure 7.4.

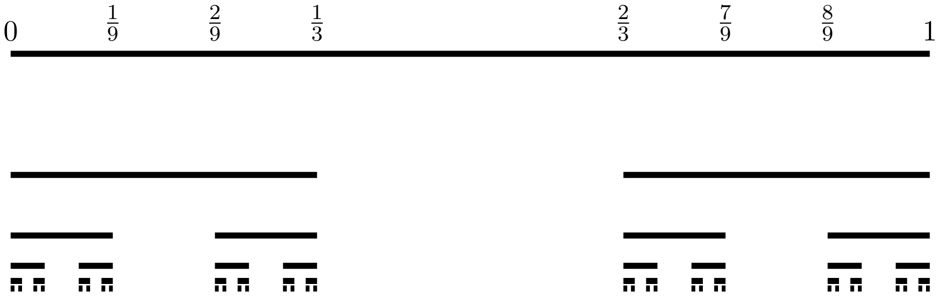


FIGURE 7.4. Towards the Cantor set

Exercises

7.1. If b is an inductive subset of $\mathscr{P}(a)$, and $c \subseteq a$, prove that $b \cap \mathscr{P}(c)$ is an inductive subset of $\mathscr{P}(c)$.

7.2. Prove that $\in$ coincides with $\prec$ on **CN**.

7.3. Prove Theorem 7.2.6.

7.4. Show that (55) and (56) are independent of the choice of a in κ and b in λ.

7.5. Prove that infinite cardinals are limit ordinals.

7.6. Show that $\operatorname{card}(a \times b) = \operatorname{card}(a) \cdot \operatorname{card}(b)$.

7.7. Prove Theorem 7.3.2 without using Theorem 7.3.4.

7.8. Prove Lemma 7.3.3.

7.9. Prove Theorem 7.4.2.

7.10. If κ is an infinite cardinal, and $1 < n < \omega$, prove $\kappa^n = \kappa$.

7.11. Use the Axiom of Choice to show that, if $f \colon a \twoheadrightarrow b$, then $b \preccurlyeq a$.

7.12. If a is a set, why is there a *set* of bijections between elements of a and their cardinals?

7.13. Prove Lemma 7.6.7.

7.14. Prove that the right lexicographic ordering of ${}^{\beta}\alpha$ defined in ¶ 7.6.2 is not generally total.

7.15. Prove Theorem 7.6.4.

7.16. Show that $x \mapsto \operatorname{pred}(x)$ is an embedding of $\mathbb{Q}$ in $\mathbb{R}$ as ordered fields.

7.17. Show that the union of a cut of $\mathbb{R}$ is the supremum of the cut.

7.18. 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. A binary relation $\boldsymbol{R}$ is **well-founded** on a class $\boldsymbol{C}$ if

(i) $\boldsymbol{R}a \cap \boldsymbol{C}$ (that is, $\{x\colon x \in \boldsymbol{C} \mathbin{\&} x \mathrel{\boldsymbol{R}} a\}$) is a set whenever $a \in \boldsymbol{C}$, and

(ii) $b \cap \boldsymbol{R}c$ (that is, $\{x\colon x \in b \mathbin{\&} x \mathrel{\boldsymbol{R}} c\}$) is empty for some c in b, whenever $b \subseteq \boldsymbol{C}$ and is nonempty.

So an *ordering* of a class is well-founded on the class if and only if each section of the class is a set and each non-empty subset of the class has a minimal element. In particular, a relation that well-orders a class is well-founded on the class. A class itself is **well-founded** if *membership* is well-founded on the class. If $\boldsymbol{R}$ is membership, then $\boldsymbol{R}a \cap \boldsymbol{C}$ is just $a \cap \boldsymbol{C}$, a set; and $b \cap \boldsymbol{R}c$ is just $b \cap c$; so $\boldsymbol{C}$ is well-founded if and only if

$$a \subseteq \boldsymbol{C} \mathbin{\&} a \neq \varnothing \Rightarrow \exists x\,(x \in a \mathbin{\&} a \cap x = \varnothing).$$

A class that is not well-founded is **ill-founded.**

8.1.2. Suppose there are *distinct* sets a and b such that $a = \{b\}$ and $b = \{a\}$. Then a and b are well-founded, but not transitive. However, $a \cup \bigcup a = \{b\} \cup b = \{a, b\}$, which is transitive, but ill-founded.

8.1.3. Theorem and Definition. *Given a set a, we recursively define the sequence $(\bigcup^n a\colon n \in \omega)$:*

(i) $\bigcup a = a$,

(ii) $\bigcup^{n+1} a = \bigcup \bigcup^n a$.

Now we define

$$\operatorname{tc}(a) = \bigcup^{\omega} a = \bigcup \{ \bigcup^{x} a\colon x \in \omega\};$$

this is the ***transitive closure*** *of a in the sense that it includes a and is transitive and is included in every class that includes a and is transitive.*

Proof. Exercise 8.2. □

8.1.4. Theorem. *If $\boldsymbol{C}$ is non-empty and well-founded, then $\boldsymbol{C} \cap a = \varnothing$ for some a in $\boldsymbol{C}$.*

Proof. Suppose $\boldsymbol{C}$ is well-founded, and $b \in \boldsymbol{C}$, but $\boldsymbol{C} \cap b \neq \varnothing$. Let $a = \operatorname{tc}(b) \cap \boldsymbol{C}$. Then $a \cap d = \varnothing$ for some d in a. But then $d \in \boldsymbol{C}$, and $d \in \operatorname{tc}(b)$, so $d \subseteq \operatorname{tc}(b)$, and therefore

$$\boldsymbol{C} \cap d = \boldsymbol{C} \cap \operatorname{tc}(b) \cap d = a \cap d = \varnothing,$$

as desired. □

8.1.5. The function $\mathbf{R}$ on $\mathbf{ON}$ is defined recursively by

(i) $\mathbf{R}(0) = \varnothing$,
(ii) $\mathbf{R}(\alpha + 1) = \mathscr{P}(\mathbf{R}(\alpha))$,
(iii) $\mathbf{R}(\alpha) = \bigcup \mathbf{R}[\,\alpha\,]$ if α is a limit.

Then we define

$$\mathbf{WF} = \bigcup \mathbf{R}[\,\mathbf{ON}\,].$$

8.1.6. 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 8.3.) Therefore, by induction, 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 a limit, then immediately $\mathbf{R}(\beta) \subseteq \mathbf{R}(\alpha)$. By induction, $\beta < \alpha \Rightarrow \mathbf{R}(\beta) \subseteq \mathbf{R}(\alpha)$. □

8.1.7. If $c \in \mathbf{WF}$, then the least ordinal α such that $c \in \mathbf{R}(\alpha)$ must be a successor, $\beta + 1$. Then β is the **rank** of c:

$$\operatorname{rank}(c) = \min\{x \in \mathbf{ON} \colon c \in \mathbf{R}(x + 1)\} = \min\{x \in \mathbf{ON} \colon c \subseteq \mathbf{R}(x)\}.$$

8.1.8. Lemma.

(i) Every subset of $\mathbf{WF}$ is an element of $\mathbf{WF}$.
(ii) $\mathbf{WF}$ and its members are well-founded.

Proof. Supposing $a \subseteq \mathbf{WF}$, let $\beta = \sup\{\operatorname{rank}(x) \colon x \in a\}$; then $a \subseteq \mathbf{R}(\beta + 1)$, so $a \in \mathbf{R}(\beta + 2)$.

Suppose $a \subseteq \mathbf{WF}$, and let b be an element of a of minimal rank. If $c \in b$, then $\operatorname{rank}(c) < \operatorname{rank}(b)$ (Exercise 8.5), so $c \notin a$. Thus $b \cap a = \varnothing$. So $\mathbf{WF}$ is well-founded; since it is transitive, its members are also well-founded. □

8.1.9. Theorem. *$\mathbf{WF}$ comprises the sets whose transitive closures are well-founded.*[1]

Proof. Suppose $a \in \mathbf{WF}$. Then $a \subseteq \mathbf{WF}$, so $\operatorname{tc}(a) \subseteq \mathbf{WF}$ by Theorem 8.1.3, and then $\operatorname{tc}(a) \in \mathbf{WF}$ by Lemma 8.1.8 (i), so $\operatorname{tc}(a)$ is well-founded by Lemma 8.1.8 (ii).

Now suppose $a \notin \mathbf{WF}$. Then $a \nsubseteq \mathbf{WF}$, so $\operatorname{tc}(a) \nsubseteq \mathbf{WF}$. Let $b \in \operatorname{tc}(a) \smallsetminus \mathbf{WF}$. Then $b \subseteq \operatorname{tc}(a)$, but $b \nsubseteq \mathbf{WF}$. Consequently

$$(\operatorname{tc}(a) \smallsetminus \mathbf{WF}) \cap b = \operatorname{tc}(a) \cap (b \smallsetminus \mathbf{WF}) = b \smallsetminus \mathbf{WF} \neq \varnothing.$$

Thus $\operatorname{tc}(a)$ is not well-founded. □

8.2. Families of classes

8.2.1. Axiom (Foundation). *All sets are well-founded:*

$$a \neq 0 \Rightarrow \exists y\ (y \in a \ \&\ y \cap a = \varnothing);$$

equivalently, $\mathbf{V} = \mathbf{WF}$ (Exercise 8.8).

8.2.2. The Foundation Axiom allows us to make the following definition, which will allow us to understand the family $\boldsymbol{C}/\boldsymbol{E}$ of ¶ 3.6.3 as a certain class. Indeed, along with the class-operations of ¶¶ 2.3.8 and 3.2.2, there is an operation that assigns to the class $\boldsymbol{C}$ a *set* $\tau(\boldsymbol{C})$ such that

[1] It is such sets that are called *well-founded* in **[6]**.

(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 $\alpha = \min\{x\colon \exists y\ (y \in \boldsymbol{C}\ \&\ \operatorname{rank}(y) = x)\}$, and let

$$\tau(\boldsymbol{C}) = \mathbf{R}(\alpha + 1) \cap \boldsymbol{C}.$$

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 $\boldsymbol{F}(a) = \tau(a\boldsymbol{E}) = \bigl\{x\colon a\ \boldsymbol{E}\ x\ \&\ \operatorname{rank}(x) = \min\{\operatorname{rank}(y)\colon a\ \boldsymbol{E}\ y\}\bigr\}$. □

8.3. Consistency

8.3.1. As axioms of set theory, Zermelo [**26**] proposed Extension, Pairing, Separation, Power Set, Union, Choice, and Infinity. Then Fraenkel (see [**6**, p. 50, n. 3]), and independently Skolem [**20**], proposed Replacement (which makes Separation redundant by Exercise 3.33); Skolem and more definitely von Neumann [**25**] proposed Foundation. The list of all of these axioms, besides Choice, is called ZF for Zermelo and Fraenkel. When Choice is added, the list is called ZFC. It was mentioned in ¶ 2.7.2 that we cannot prove the consistency of ZFC. Nonetheless, there are *relative* consistency results, as for example that *if* ZF is consistent, then so is ZFC. We shall prove this and similar results.

8.3.2. A class $\boldsymbol{M}$ serves as a truth-assignment in the sense of ¶ 2.5.1. If a and b are in $\boldsymbol{M}$, and $a \in b$, then we write $\boldsymbol{M} \models a \in b$; if $a \notin b$, then $\boldsymbol{M} \nvDash a \in b$. Recursively, we have

(i) $\boldsymbol{M} \models \neg\sigma$ if and only if $\boldsymbol{M} \nvDash \sigma$;
(ii) $\boldsymbol{M} \models \sigma \Rightarrow \tau$ if and only if $\boldsymbol{M} \nvDash \sigma$ or $\boldsymbol{M} \models \tau$;
(iii) $\boldsymbol{M} \models \exists x\ \varphi$ if and only if $\boldsymbol{M} \models \varphi(a)$ for some a in $\boldsymbol{M}$.

Again, $\boldsymbol{M} \models \sigma$ means σ is **true in $\boldsymbol{M}$**. If φ is an n-ary formula with constants from $\boldsymbol{M}$, then

$$\varphi^{\boldsymbol{M}} = \{\vec{x}\colon \vec{x} \in \boldsymbol{M}^n\ \&\ \boldsymbol{M} \models \varphi(\vec{x})\};$$

this is the class **defined by φ in $\boldsymbol{M}$**. If

$$\varphi^{\boldsymbol{M}} = \boldsymbol{M}^n \cap \varphi^{\mathbf{V}},$$

then φ is **absolute for $\boldsymbol{M}$**.

8.3.3. Theorem. *In $\mathbf{R}(\omega)$, the axioms of* ZFC *besides Infinity are true, but Infinity is false.*

Proof. Say a and b are in $\mathbf{R}(\omega)$.

1. Since $\{a\}$ and $\{b\}$ are in $\mathbf{R}(\omega)$, *equality* (that is, the formula $x = y$) is absolute for $\mathbf{R}(\omega)$. If $a \not\subseteq b$, then, since $\mathbf{R}(\omega)$ is transitive, it contains an element of $a \smallsetminus b$, so $\mathbf{R}(\omega) \models a \not\subseteq b$. This shows that *inclusion* (the formula $x \subseteq y$) is absolute for $\mathbf{R}(\omega)$; therefore Extension is true in $\mathbf{R}(\omega)$.

2. Since $\{a, b\} \in \mathbf{R}(\omega)$, Pairing is true in $\mathbf{R}(\omega)$.

3. Every element of $\mathbf{R}(\omega)$ is finite, and every finite subset of $\mathbf{R}(\omega)$ is an element. This shows that Replacement (hence Separation) is true in $\mathbf{R}(\omega)$.

4. For the same reason, Infinity is false in $\mathbf{R}(\omega)$.

5. We have $a \in \mathbf{R}(n+1)$ for some n in ω, so $a \subseteq \mathbf{R}(n)$, hence $\mathscr{P}(a) \subseteq \mathbf{R}(n+1)$, so $\mathscr{P}(a) \in \mathbf{R}(n+2)$. Thus Power Set is true in $\mathbf{R}(\omega)$.

6. If $c \in a$, then $c \in \mathbf{R}(n)$, so $c \subseteq \mathbf{R}(n-1)$. Hence $\bigcup a \subseteq \mathbf{R}(n-1)$, so $\bigcup a \in \mathbf{R}(n)$. This shows Union is true in $\mathbf{R}(\omega)$.

7. If c is an element of a of minimal rank, then $c \in \mathbf{R}(\omega)$ (by transitivity) and $a \cap c = \varnothing$; thus Foundation is true in $\mathbf{R}(\omega)$.

8. Finally, Choice is true in $\mathbf{R}(\omega)$ because a choice-function for a is a subset of $\mathscr{P}(a) \times a$, hence of $\mathscr{P}(\mathscr{P}(\mathscr{P}(a) \cup a))$ (¶ 3.4.4), hence of $\mathbf{R}(n+3)$, so it belongs to $\mathbf{R}(n+4)$. (See also Exercise 8.9.) □

8.3.4. Theorem. *In* $\mathbf{R}(\omega \cdot 2)$*, the axioms of* ZFC*, besides Replacement, are true, but Replacement is false.*

Proof. The proof of Theorem 8.3.3 shows that all axioms but Separation, Replacement, and Infinity are true in $\mathbf{R}(\omega \cdot 2)$. Moreover, Separation is true in $\mathbf{R}(\omega \cdot 2)$ since, if $a \in \mathbf{R}(\omega \cdot 2)$ and $b \subseteq a$, then $b \subseteq \mathbf{R}(\alpha)$ for some α in $\omega \cdot 2$, and then $b \in \mathbf{R}(\alpha+1)$, so $b \in \mathbf{R}(\omega \cdot 2)$. Infinity is true in $\mathbf{R}(\omega \cdot 2)$ since it contains ω. Replacement is not true in $\mathbf{R}(\omega \cdot 2)$ since $\{\mathbf{R}(\omega+n)\colon n \in \omega\}$ is the image of ω under a function definable in $\mathbf{R}(\omega \cdot 2)$, but is not an element of $\mathbf{R}(\omega \cdot 2)$. □

8.3.5. Theorem. *In* **WF***, the axioms of* ZFC *are true.*

Proof. By the proof of Theorem 8.3.4, it remains to show that Replacement is true in **WF**. If $a \in \mathbf{WF}$, and $\boldsymbol{F}\colon a \to \mathbf{WF}$, let $\beta = \sup\{\operatorname{rank}(\boldsymbol{F}(x))\colon x \in a\}$; then $\boldsymbol{F}[\,a\,] \in \mathbf{R}(\beta+2)$, so $\boldsymbol{F}[\,a\,] \in \mathbf{WF}$. □

8.3.6. Theorems 8.3.3 and 8.3.4 show that neither Infinity nor Replacement is deducible from the other axioms in ZFC, provided that these others are consistent. Since the definition of **WF** and the proof of Theorem 8.3.5 do not assume Foundation, we can conclude that, if the ZF axioms without Foundation are consistent, then ZF is consistent, and likewise for ZFC.

8.4. Constructible sets

8.4.1. If $a \subseteq b$, a relation on b is **definable over** a if the relation is φ^b (in the sense of ¶ 8.3.2) for some formula φ with constants from a. A relation on a is **definable,** simply, if it is definable over a. By Theorem 8.4.2 below, for all n in ω, the n-ary definable relations on a set a compose a set, which we shall denote by

$$\mathscr{D}_n(a).$$

Then $\mathscr{D}_n(a) \subseteq \mathscr{P}(a^n)$. However, if a is infinite, then $\operatorname{card}(\mathscr{D}_n(a)) = \operatorname{card}(a)$ (Exercise 8.10); so not every relation on a is definable.

8.4.2. Theorem. *For all n in ω, the n-ary relations on a set compose a set.*

Proof. Considering the recursive definition of formulas (¶ 2.2.1), we obtain the definable relations as follows. Given a set a, we let b be the set of all sequences $(c_n\colon n \in \omega)$, where $c_n \subseteq a^n$, and

(i) $\{\vec{x} \in a^n\colon x_i \in x_j\} \in c_n$ whenever $\{i,j\} \subseteq n$;

(ii) $\{\vec{x} \in a^n\colon x_i \in d\} \in c_n$ whenever $i \in n$ and $d \in a$;

(iii) $\{\vec{x} \in a^n \colon d \in x_i\} \in c_n$ whenever $i \in n$ and $d \in a$;
(iv) c_n is closed under the operations $x \mapsto a^n \smallsetminus x$ and $(x, y) \mapsto y \smallsetminus x$ on a^n;
(v) if f_n is the function $(x_0, \dots, x_n) \mapsto (x_0, \dots, x_{n-1})$ from a^{n+1} to a^n, and $e \in c_{n+1}$, then $f_n[\,e\,] \in c_n$.

Let b_k be the set of all c_k such that $(c_n \colon n \in \omega) \in b$ for some c_n where $n \neq k$. Then the definable k-ary relations on a compose the set $\bigcap b_k$. □

8.4.3. A function $x \mapsto \mathbf{L}(x)$ on $\mathbf{ON}$ is defined recursively as follows:
(i) $\mathbf{L}(0) = \varnothing$,
(ii) $\mathbf{L}(\alpha + 1) = \mathscr{D}_1(\mathbf{L}(\alpha))$,
(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 $x \mapsto \mathbf{L}(x)$). The elements of $\mathbf{L}$ are the **constructible sets.**

8.4.4. Lemma. *Each set $\mathbf{L}(\alpha)$ is transitive, so $\mathbf{L}$ is transitive. If $\beta < \alpha$, then $\mathbf{L}(\beta) \subseteq \mathbf{L}(\alpha)$.*

Proof. Suppose a is transitive. Then every element b of a is the subset of a defined by $x \in b$. Thus $a \subseteq \mathscr{D}_1(a)$. Since every element of $\mathscr{D}_1(a)$ is a subset of a, we conclude that $\mathscr{D}_1(a)$ is transitive. Since $\mathbf{L}(0)$ is trivially transitive, by induction the claim follows. □

8.4.5. Lemma (Tarski–Vaught Test). *Suppose $\boldsymbol{C} \subseteq \boldsymbol{D}$ and φ is an n-ary formula with constants from $\boldsymbol{C}$ such that, for every subformula of φ of the form $\exists x\ \psi(x, \vec{y})$, for all tuples $\vec{b}$ of elements of $\boldsymbol{C}$, if $\boldsymbol{D} \models \exists x\ \psi(x, \vec{b})$, then $\boldsymbol{D} \models \psi(a, \vec{b})$ for some a in $\boldsymbol{C}$. Then*

$$\varphi^{\boldsymbol{C}} = \boldsymbol{C}^n \cap \varphi^{\boldsymbol{D}}. \tag{62}$$

Proof. We prove by induction that (62) holds whenever φ is a subformula of the original formula. The claim holds easily when φ is atomic, and if it holds when φ is ψ or ρ, then it holds when φ is $\neg\chi$ or $\chi \Rightarrow \rho$. Suppose (62) holds when φ is $\psi(x, \vec{y})$, and $\vec{b}$ is from $\boldsymbol{C}$. The following statements are equivalent:
(i) $\boldsymbol{D} \models \exists x\ \psi(x, \vec{b})$;
(ii) $\boldsymbol{D} \models \psi(a, \vec{b})$ for some a in $\boldsymbol{C}$ (by assumption);
(iii) $\boldsymbol{C} \models \psi(a, \vec{b})$ for some a in $\boldsymbol{C}$ (by inductive hypothesis);
(iv) $\boldsymbol{C} \models \exists x\ \psi(x, \vec{b})$.

Thus (62) holds when φ is $\exists x\ \psi(x, \vec{y})$. Therefore it holds for all subformulas of the original formula. □

8.4.6. Lemma. *For every n in ω, for every n-ary formula φ with constants from $\mathbf{L}$, there is β such that*

$$\varphi^{\mathbf{L}(\beta)} = \mathbf{L}(\beta)^n \cap \varphi^{\mathbf{L}}.$$

Proof. For every ordinal α, let α^* be the least ordinal γ greater than α such that, if $\vec{b}$ is from $\mathbf{L}(\alpha)$, and $\exists x\ \psi(x, \vec{y})$ is a subformula of ψ, and $\mathbf{L} \models \exists x\ \psi(x, \vec{b})$, then $\mathbf{L} \models \psi(a, \vec{b})$ for some a in $\mathbf{L}(\gamma)$. Let α_0 be such that the constants of φ are from $\mathbf{L}(\alpha_0)$, and let $\alpha_{k+1} = {\alpha_k}^*$. By the Tarski–Vaught Test, it suffices to let $\beta = \sup\{\alpha_n \colon n \in \omega\}$, which is a limit. Indeed, if $\vec{b}$ is from $\mathbf{L}(\beta)$, then $\vec{b}$ is from $\mathbf{L}(\alpha_n)$ for some n, so if $\mathbf{L} \models \exists x\ \psi(x, \vec{b})$, where $\exists x\ \psi(x, \vec{y})$ is a subformula of φ, then $\mathbf{L} \models \psi(a, \vec{b})$ for some a in $\mathbf{L}(\alpha_{n+1})$ and hence in $\mathbf{L}(\beta)$. □

8.4.7. By analogy with ¶ 8.1.7, if $a \in \mathbf{L}$, we define

$$\operatorname{rank}_{\mathbf{L}}(a) = \min\{x \colon a \in \mathbf{L}(x+1)\}.$$

Note however that possibly $a \in \mathbf{L}$, and $a \subseteq \mathbf{L}(\beta)$, but a is not a *definable* subset of $\mathbf{L}(\beta)$, so $a \notin \mathbf{L}(\beta+1)$, and so $\beta < \operatorname{rank}_{\mathbf{L}}(a)$.

8.4.8. Theorem. ZF *is true in* $\mathbf{L}$.

Proof. As in the proof of Theorem 8.3.3, Foundation is true in $\mathbf{L}$; also equality and inclusion are absolute for $\mathbf{L}$, and therefore Extension is true in $\mathbf{L}$, and so is Pairing.

Suppose $a \in \mathbf{L}$. Let $\beta = \sup\{\operatorname{rank}_{\mathbf{L}}(x) \colon x \in \mathbf{L} \mathbin{\&} x \subseteq a\}$. Then $\mathbf{L} \cap \mathscr{P}(a) \in \mathbf{L}(\beta+2)$. Thus Power Set is true in $\mathbf{L}$.

Say $\operatorname{rank}_{\mathbf{L}}(a) = \gamma$. Then a is a subset of $\mathbf{L}(\gamma)$ defined by a formula φ. Since $\mathbf{L}(\gamma)$ is transitive, $\bigcup a$ is defined in $\mathbf{L}(\gamma)$ by $\exists y\ (\varphi(y) \mathbin{\&} x \in y)$. Therefore $\bigcup a \in \mathbf{L}(\gamma+1)$. Thus Union is true in $\mathbf{L}$.

If $b \in \mathbf{L}(\alpha)$, then $b \cup \{b\} \subseteq \mathbf{L}(\alpha+1)$, and so $b' \in \mathbf{L}(\alpha+1)$. Since $0 \in \mathbf{L}(1)$, we have $\omega \subseteq \mathbf{L}(\omega)$. Also $x \in \omega$ is absolute for $\mathbf{L}$. Therefore $\omega \in \mathbf{L}(\omega+1)$, so Infinity is true in $\mathbf{L}$.

Suppose $\varphi(x, y)$ defines in $\mathbf{L}$ a function $\boldsymbol{F}$, and $a \in \mathbf{L}$. Some $\mathbf{L}(\alpha)$ contains all constants in φ and elements of $\boldsymbol{F}[a]$. Then $\varphi(a, y)^{\mathbf{L}} \subseteq \mathbf{L}(\alpha)$. By Lemma 8.4.6, there is β such that $\varphi(a, y)^{\mathbf{L}} = \varphi(a, y)^{\mathbf{L}(\beta)}$. Thus $\boldsymbol{F}[a] \in \mathbf{L}(\beta+1)$. Therefore Replacement is true in $\mathbf{L}$. □

8.4.9. Theorem. *The Axiom of Choice is true in* $\mathbf{L}$. *Therefore* ZFC *is consistent (assuming* ZF *is consistent).*

Proof. There is a binary formula φ such that $\mathbf{L}$ is well-ordered by $\varphi^{\mathbf{L}}$. Indeed, because of the recursive construction of the sets $\mathscr{D}_n(a)$, there is a ternary formula ψ such that, if r well-orders $\mathbf{L}(\alpha)$, then $\psi(r, x, y)^{\mathbf{L}}$ well-orders $\mathbf{L}(\alpha+1)$. We define the ordering $<$ on $\mathbf{L}$ recursively as $\bigcup\{r_x \colon x \in \mathbf{ON}\}$ so that $a < b$ if $\operatorname{rank}_{\mathbf{L}}(a) < \operatorname{rank}_{\mathbf{L}}(b)$ or if $\operatorname{rank}_{\mathbf{L}}(a) = \operatorname{rank}_{\mathbf{L}}(b)$ and $\mathbf{L} \models \psi(r_{\operatorname{rank}_{\mathbf{L}}(a)}, a, b)$. This argument does not use the Axiom of Choice. □

8.5. The Generalized Continuum Hypothesis

8.5.1. Theorem (Löwenheim–Skolem). *For every set a and every formula φ with constants from a, there is a set b such that $a \subseteq b$, and*

$$\operatorname{card}(b) \leqslant \operatorname{card}(a) + \aleph_0,$$

and φ is absolute for b.

Proof. Suppose $a \subseteq c$. Let α be the supremum of the set of all ordinals

$$\min\{\operatorname{rank}(z) \colon z \in \psi(x, \vec{b})^{\mathbf{V}}\},$$

where $\exists x\ \psi(x, \vec{y})$ is a subformula of φ, and $\vec{b}$ is from c, and $\exists x\ \psi(x, \vec{b})$ is true. Fix a relation r that well-orders $\mathbf{R}(\alpha+1)$. Define $c(r)$ to comprise each element of c and the least element of each nonempty set $\mathbf{R}(\alpha+1) \cap \psi(x,\vec{b})^{\mathbf{V}}$, again where $\exists x\ \psi(x,\vec{y})$ is a subformula of φ, and $\vec{b}$ is from c. Note that $\operatorname{card}(c(r)) \leqslant \operatorname{card}(c) + \aleph_0$.

Now let $(a_n : n \in \omega)$ be such that $a_0 = a$ and $a_{n+1} = a_n(r)$ for some choice of r as above. Let $b = \bigcup\{a_n : n \in \omega\}$. Then b is as desired, by the Tarski–Vaught Test. □

8.5.2. A relation $\boldsymbol{R}$ on a class $\boldsymbol{C}$ is **extensional** if the Extension Axiom is true in the structure $(\boldsymbol{C},\boldsymbol{R})$ in the sense that

$$\boldsymbol{R}a = \boldsymbol{R}b \Rightarrow a = b.$$

8.5.3. Theorem (Mostowski Collapse). *Let $\boldsymbol{R}$ be a well-founded relation on $\boldsymbol{C}$. There is a unique function $\boldsymbol{F}$ on $\boldsymbol{C}$ given by*

$$\boldsymbol{F}(a) = \boldsymbol{F}[\,\boldsymbol{R}a\,].$$

Then $\boldsymbol{F}[\,\boldsymbol{C}\,]$ is transitive. If $\boldsymbol{R}$ is extensional, then $\boldsymbol{F}$ is an isomorphism from $(\boldsymbol{C},\boldsymbol{R})$ to $(\boldsymbol{F}[\,\boldsymbol{C}\,],\in)$.

Proof. We follow the proof of Theorem 4.6.2, though since $\boldsymbol{R}$ need not be transitive, we shall need the following definition. if $a \in \boldsymbol{C}$, then

$$\operatorname{cl}_0(a) = \{a\}, \quad \operatorname{cl}_{n+1}(a) = \{x : \exists y\ (y \in \operatorname{cl}_n(a)\ \&\ x\ \boldsymbol{R}\ y)\},$$

$$\operatorname{cl}(a) = \bigcup\{\operatorname{cl}_n(a) : n \in \omega\}.$$

The structure $(\boldsymbol{C},\boldsymbol{R})$ admits induction in the sense that, if $\boldsymbol{C}_0 \subseteq \boldsymbol{C}$ and, for all a in $\boldsymbol{C}$, we have $a \in \boldsymbol{C}_0$ whenever $\boldsymbol{R}a \subseteq \boldsymbol{C}_0$, then $\boldsymbol{C}_0 = \boldsymbol{C}$. Indeed, suppose $\boldsymbol{C}_0 \subset \boldsymbol{C}$, and $a \in \boldsymbol{C} \smallsetminus \boldsymbol{C}_0$. By definition (¶ 8.1.1), $(\boldsymbol{C} \smallsetminus \boldsymbol{C}_0) \cap \operatorname{cl}(a)$ has an element b such that

$$(\boldsymbol{C} \smallsetminus \boldsymbol{C}_0) \cap \operatorname{cl}(a) \cap \boldsymbol{R}b = \varnothing.$$

But $\boldsymbol{R}b \subseteq \operatorname{cl}(a)$ (since $b \in \operatorname{cl}_n(a)$ for some n, and then $\boldsymbol{R}b \subseteq \operatorname{cl}_{n+1}(a)$). Hence $(\boldsymbol{C} \smallsetminus \boldsymbol{C}_0) \cap \boldsymbol{R}b = \varnothing$, so $\boldsymbol{R}b \subseteq \boldsymbol{C}_0$.

Now we can show by induction that, for all a in $\boldsymbol{C}$, there is a unique function f_a on $\operatorname{cl}(a)$ such that

$$f_a(c) = f_a[\,\boldsymbol{R}c\,].$$

Indeed, suppose the claim holds when $a\ \boldsymbol{R}\ b$. If a and d are in $\boldsymbol{R}b$, then f_a and f_d must agree on $\operatorname{cl}(a) \cap \operatorname{cl}(d)$ since, if f_a and g disagree on $\operatorname{cl}(a) \cap \operatorname{cl}(d)$, then by well-foundedness this set has an element e such that f_a and g agree on $\boldsymbol{R}e$, but

$$g(e) \neq f_a(e) = f_a[\,\boldsymbol{R}e\,] = g[\,\boldsymbol{R}e\,].$$

Now we can define f_b on $\operatorname{cl}(b)$ so that, if $c \in \operatorname{cl}(b) \smallsetminus \{b\}$, then $f_b(c) = f_a(c)$, where a is such that $a\ \boldsymbol{R}\ b$ and $c \in \operatorname{cl}(a)$; and $f_b(b) = f_b[\,\boldsymbol{R}b\,]$.

The desired function $\boldsymbol{F}$ is now $\bigcup\{f_a : a \in \boldsymbol{C}\}$. Indeed, this is a function, since any two functions f_a agree as before on the intersection of their domains. Likewise, $\boldsymbol{F}$ itself is unique. Since $\boldsymbol{F}(a) = \boldsymbol{F}[\,\boldsymbol{R}a\,] \subseteq \boldsymbol{F}[\,\boldsymbol{C}\,]$, it follows that $\boldsymbol{F}[\,\boldsymbol{C}\,]$ is transitive.

We have $a\ \boldsymbol{R}\ b \Rightarrow \boldsymbol{F}(a) \in \boldsymbol{F}(b)$. If $\boldsymbol{F}$ is injective, then $\boldsymbol{F}(a) \in \boldsymbol{F}(b) \Rightarrow a\ \boldsymbol{R}\ b$, so $\boldsymbol{F}$ is an isomorphism. Suppose $\boldsymbol{F}$ is not injective. Let $\boldsymbol{C}_0$ comprise those a in $\boldsymbol{C}$ for which there is no distinct b such that $\boldsymbol{F}(a) = \boldsymbol{F}(b)$. As in the proof that $(\boldsymbol{C},\boldsymbol{R})$ admits induction, there is an element a of $\boldsymbol{C} \smallsetminus \boldsymbol{C}_0$ such that $\boldsymbol{R}a \subseteq \boldsymbol{C}_0$. Then $\boldsymbol{F}(a) = \boldsymbol{F}(b)$ for some distinct b. This means

$$\{\boldsymbol{F}(x) : x\ \boldsymbol{R}\ a\} = \{\boldsymbol{F}(y) : y\ \boldsymbol{R}\ b\}.$$

Since $\boldsymbol{R}a \subseteq \boldsymbol{C}_0$, we conclude $\boldsymbol{R}b = \boldsymbol{R}a$. Thus $\boldsymbol{R}$ is not extensional. □

8.5.4. Theorem. *The Generalized Continuum Hypothesis is true in* $\mathbf{L}$. *Thus* GCH *is consistent with* ZFC *(assuming* ZF *is consistent).*

Proof. Suppose $a \in \mathscr{P}(\mathbf{L}(\alpha)) \cap \mathbf{L}$, where α is infinite. We shall show

$$a \in \mathbf{L}(\operatorname{card}(\alpha)^+).$$

Since $\operatorname{card}(\mathbf{L}(\beta)) = \operatorname{card}(\beta)$ (in $\mathbf{V}$ and in $\mathbf{L}$), it will follow that $\operatorname{card}(\mathscr{P}(\kappa)) = \kappa^+$ in $\mathbf{L}$.

Apply the Löwenheim–Skolem Theorem to the set $\mathbf{L}(\alpha) \cup \{a\}$ and the conjunction of the Extension Axiom with the formula $a \in \mathbf{L}(x)$. We get a set b such that $\mathbf{L}(\alpha) \subseteq b$, $a \in b$, $\operatorname{card}(b) = \operatorname{card}(\alpha)$, $(b, \in)$ is extensional, and

$$b \models \exists x\ a \in \mathbf{L}(x).$$

By the Mostowski Collapse Theorem, we may assume further that b is *transitive.* In particular, an element β of b such that $b \models a \in \mathbf{L}(\beta)$ really is an ordinal. Then $a \in \mathbf{L}(\beta)$, but $\operatorname{card}(\beta) = \operatorname{card}(\alpha)$, so $a \in \mathbf{L}(\operatorname{card}(\alpha)^+)$. □

8.5.5. About a quarter century after Gödel proved that AC and GCH are consistent with ZF, Cohen (see [**2**]) proved the same of their negations.

Exercises

8.1. Prove that class is ill-founded if and only if there is a sequence $(a_n \colon n \in \omega)$ of elements of the class such that $a_{n+1} \in a_n$ in each case. (One direction requires the Axiom of Choice.)

8.2. Prove Theorem 8.1.3.

8.3. Show that both the power set of a transitive set and the union of a set of transitive sets are transitive.

8.4. Show $\beta < \alpha \Rightarrow \mathbf{R}(\beta) \in \mathbf{R}(\alpha)$.

8.5. Assuming $a \in \mathbf{WF}$ and $b \in a$, show that $b \in \mathbf{WF}$ and $\operatorname{rank}(b) < \operatorname{rank}(a)$.

8.6. Show that all subsets of $\mathbf{WF}$ are elements of $\mathbf{WF}$.

8.7. 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 $f(n+1) \in f(n) \smallsetminus \mathbf{WF}$.

8.8. Show that the two given formulations of the Foundation Axiom are equivalent.

8.9. Prove that the Axiom of Choice is true in $\mathbf{R}(\omega)$ without using the Axiom of Choice.

8.10. If a is infinite, show that $\operatorname{card}(\mathscr{D}_n(a)) = \operatorname{card}(a)$.

APPENDIX A

The Greek alphabet

capital	minuscule	transliteration	name
Α	α	a	alpha
Β	β	b	beta
Γ	γ	g	gamma
Δ	δ	d	delta
Ε	ϵ	e	epsilon
Ζ	ζ	z	zeta
Η	η	ê	eta
Θ	θ	th	theta
Ι	ι	i	iota
Κ	κ	k	kappa
Λ	λ	l	lambda
Μ	μ	m	mu
Ν	ν	n	nu
Ξ	ξ	x	xi
Ο	ο	o	omicron
Π	π	p	pi
Ρ	ρ	r	rho
Σ	σ, ς	s	sigma
Τ	τ	t	tau
Υ	υ	y, u	upsilon
Φ	ϕ	ph	phi
Χ	χ	ch	chi
Ψ	ψ	ps	psi
Ω	ω	ô	omega

The following remarks pertain to *ancient* Greek. The vowels are α, ϵ, η, ι, ο, υ, ω, where η is a long ϵ, and ω is a long ο; the other vowels (α, ι, υ) can be long or short. Some vowels may be given tonal accents (ά, ᾶ, ὰ). An initial vowel takes either a rough-breathing mark (as in ἁ) or a smooth-breathing mark (ἀ): the former mark is transliterated by a preceding h, and the latter can be ignored, as in ὑπϵρβολή hyperbolê *hyperbola,* ὀρθογώνιον orthogônion *rectangle.* Likewise, ῥ is transliterated as rh, as in ῥόμβος rhombos *rhombus.* A long vowel may have an iota subscript (ᾳ, ῃ, ῳ), especially in case-endings of nouns. Of the two forms of minuscule sigma, the ς appears at the ends of words; elsewhere, σ appears, as in βάσις basis *base.*

APPENDIX B

The German script

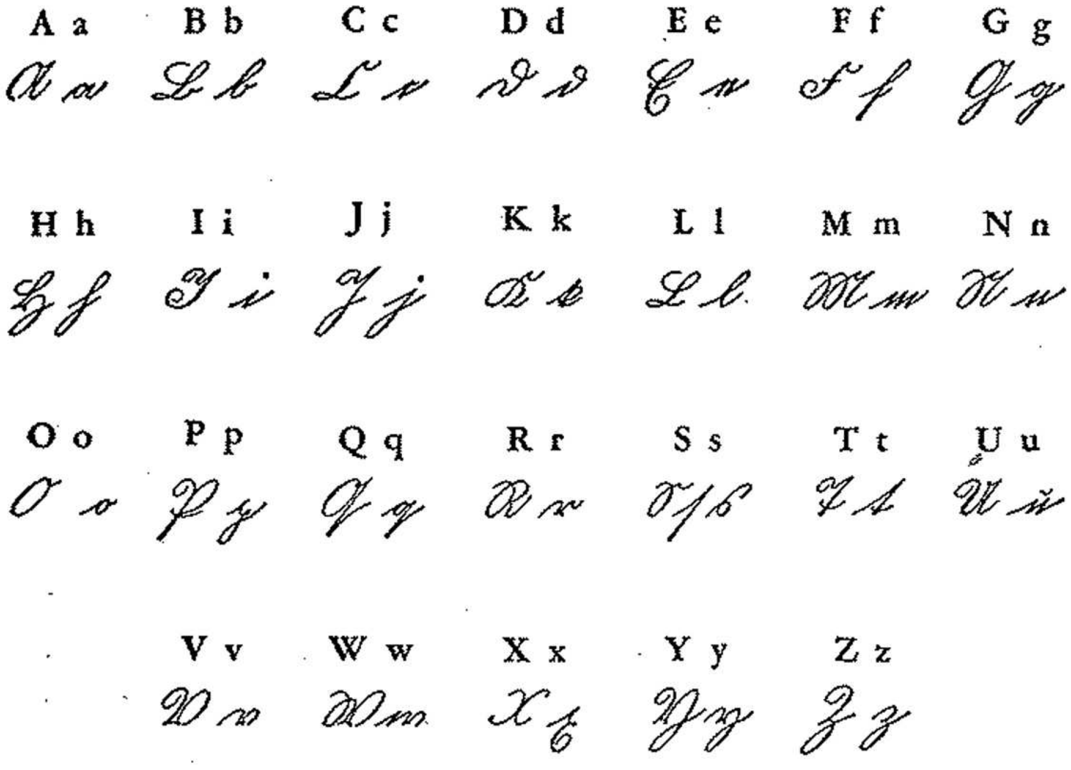
Writing in 1993, Wilfrid Hodges [**8**, Ch. 1, p. 21] observes

> Until about a dozen years ago, most model theorists named structures in horrible Fraktur lettering. Recent writers sometimes adopt a notation according to which all structures are named M, M', M^*, $\bar{M}$, M_0, M_i or occasionally N.

For Hodges, structures are A, B, C, and so forth; he refers to their universes as **domains** and denotes these by dom(A) and so forth. I still prefer the Fraktur letters:

$\mathfrak{A}$	$\mathfrak{B}$	$\mathfrak{C}$	$\mathfrak{D}$	$\mathfrak{E}$	$\mathfrak{F}$	$\mathfrak{G}$	$\mathfrak{H}$	$\mathfrak{I}$	$\mathfrak{a}$	$\mathfrak{b}$	$\mathfrak{c}$	$\mathfrak{d}$	$\mathfrak{e}$	$\mathfrak{f}$	$\mathfrak{g}$	$\mathfrak{h}$	$\mathfrak{i}$
$\mathfrak{J}$	$\mathfrak{K}$	$\mathfrak{L}$	$\mathfrak{M}$	$\mathfrak{N}$	$\mathfrak{O}$	$\mathfrak{P}$	$\mathfrak{Q}$	$\mathfrak{R}$	$\mathfrak{j}$	$\mathfrak{k}$	$\mathfrak{l}$	$\mathfrak{m}$	$\mathfrak{n}$	$\mathfrak{o}$	$\mathfrak{p}$	$\mathfrak{q}$	$\mathfrak{r}$
$\mathfrak{S}$	$\mathfrak{T}$	$\mathfrak{U}$	$\mathfrak{V}$	$\mathfrak{W}$	$\mathfrak{X}$	$\mathfrak{Y}$	$\mathfrak{Z}$		$\mathfrak{s}$	$\mathfrak{t}$	$\mathfrak{u}$	$\mathfrak{v}$	$\mathfrak{w}$	$\mathfrak{x}$	$\mathfrak{y}$	$\mathfrak{z}$	

A way to write these by hand is seen in a textbook of German from 1931 [**7**]:



Bibliography

[1] Alonzo Church. *Introduction to mathematical logic. Vol. I.* Princeton University Press, Princeton, N. J., 1956.

[2] Paul J. Cohen. *Set theory and the continuum hypothesis.* W. A. Benjamin, Inc., New York-Amsterdam, 1966.

[3] Richard Dedekind. *Essays on the theory of numbers. I: Continuity and irrational numbers. II: The nature and meaning of numbers.* authorized translation by Wooster Woodruff Beman. Dover Publications Inc., New York, 1963.

[4] Herbert B. Enderton. *A mathematical introduction to logic.* Academic Press, New York, 1972.

[5] Euclid. *Euclid's* Elements. Green Lion Press, Santa Fe, NM, 2002. All thirteen books complete in one volume, the Thomas L. Heath translation, edited by Dana Densmore.

[6] 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.

[7] Roe-Merrill S. Heffner. *Brief German Grammar.* D. C. Heath and Company, Boston, 1931.

[8] Wilfrid Hodges. *Model theory*, volume 42 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1993.

[9] Wilfrid Hodges. *Logic.* Penguin Books, London, UK, second edition, 2001. An introduction to elementary logic.

[10] Kenneth Kunen. *Set theory*, volume 102 of *Studies in Logic and the Foundations of Mathematics.* North-Holland Publishing Co., Amsterdam, 1983. An introduction to independence proofs, Reprint of the 1980 original.

[11] Edmund Landau. *Foundations of Analysis. The Arithmetic of Whole, Rational, Irrational and Complex Numbers.* Chelsea Publishing Company, New York, N.Y., third edition, 1966. translated by F. Steinhardt; first edition 1951; first German publication, 1929.

[12] Azriel Levy. *Basic set theory.* Dover Publications Inc., Mineola, NY, 2002. Reprint of the 1979 original [Springer, Berlin].

[13] Yiannis N. Moschovakis. *Notes on set theory.* Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1994.

[14] Murray et al., editors. *The Compact Edition of the Oxford English Dictionary.* Oxford University Press, 1973.

[15] Filiz Öktem. *Uygulamalı Latin Dili [Practical Latin Grammar].* Sosyal Yayınlar, İstanbul, 1996.

[16] Giuseppe Peano. The principles of arithmetic, presented by a new method (1889). In Jean van Heijenoort, editor, *From Frege to Gödel*, pages 83–97. Harvard University Press, 1976.

[17] Plato. *Republic.* Loeb Classical Library. Harvard University Press and William Heinemann Ltd., Cambridge, Massachusetts, and London, 1980. with an English Translation by Paul Shorey, in two volumes.

[18] Plato. *Republic.* Oxford University Press, Translated with an Introduction and Notes by Robin Waterfield 1998.

[19] Joseph R. Shoenfield. *Mathematical logic.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1967.

[20] Thoralf Skolem. Some remarks on axiomatized set theory (1922). In Jean van Heijenoort, editor, *From Frege to Gödel*, pages 290–301. Harvard University Press, 1976.

[21] Robert R. Stoll. *Set theory and logic.* Dover Publications Inc., New York, 1979. corrected reprint of the 1963 edition.

[22] 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).

[23] 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.

[24] Jean van Heijenoort, editor. *From Frege to Gödel.* Harvard University Press, Cambridge, MA, 2002. A source book in mathematical logic, 1879–1931, Reprint of the third printing of the 1967 original.

[25] John von Neumann. An axiomatization of set theory (1925). In Jean van Heijenoort, editor, *From Frege to Gödel*, pages 393–413. Harvard University Press, 1976.

[26] Ernst Zermelo. Investigations in the foundations of set theory I (1908a). In Jean van Heijenoort, editor, *From Frege to Gödel*, pages 199–215. Harvard University Press, 1976.

Index