Undergraduate Texts in Mathematics

Undergraduate Texts in Mathematics

Yiannis Moschovakis

Notes on Set Theory

Second Edition

With 48 Figures

Springer

Yiannis Moschovakis
Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90095-1555
USA
ynm@math.ucla.edu

Dedicated to the memory of Nikos Kritikos

PREFACE

What this book is about. The *theory of sets* is a vibrant, exciting mathematical theory, with its own basic notions, fundamental results and deep open problems, and with significant applications to other mathematical theories. At the same time, *axiomatic set theory* is often viewed as a *foundation of mathematics*: it is alleged that all mathematical objects are sets, and their properties can be derived from the relatively few and elegant axioms about sets. Nothing so simple-minded can be quite true, but there is little doubt that in standard, current mathematical practice, "making a notion precise" is essentially synonymous with "defining it in set theory". Set theory is the official language of mathematics, just as mathematics is the official language of science.

Like most authors of elementary, introductory books about sets, I have tried to do justice to both aspects of the subject.

From straight set theory, these Notes cover the basic facts about "abstract sets", including the Axiom of Choice, transfinite recursion, and cardinal and ordinal numbers. Somewhat less common is the inclusion of a chapter on "pointsets" which focuses on results of interest to analysts and introduces the reader to the Continuum Problem, central to set theory from the very beginning. There is also some novelty in the approach to cardinal numbers, which are brought in very early (following Cantor, but somewhat deviously), so that the basic formulas of cardinal arithmetic can be taught as quickly as possible. Appendix **A** gives a more detailed "construction" of the real numbers than is common nowadays, which in addition claims some novelty of approach and detail. Appendix **B** is a somewhat eccentric, mathematical introduction to the study of *natural models* of various set theoretic principles, including Aczel's Antifoundation. It assumes no knowledge of logic, but should drive the serious reader to study it.

About set theory as a foundation of mathematics, there are two aspects of these Notes which are somewhat uncommon. First, I have taken seriously this business about "everything being a set" (which of course it is not) and have tried to make sense of it in terms of the notion of *faithful representation* of mathematical objects by *structured sets*. An old idea, but perhaps this is the first textbook which takes it seriously, tries to explain it, and applies it consistently. Those who favor category theory will recognize some of its basic notions in places, shamelessly folded into a traditional set theoretical

approach to the foundations where categories are never mentioned. Second, *computation theory* is viewed as part of the mathematics "to be founded" and the relevant set theoretic results have been included, along with several examples. The ambition was to explain what every young mathematician or theoretical computer scientist needs to know about sets.

The book includes several historical remarks and quotations which in some places give it an undeserved scholarly gloss. All the quotations (and most of the comments) are from papers reprinted in the following two, marvellous and easily accessible source books, which should be perused by all students of set theory:

Georg Cantor, *Contributions to the founding of the theory of transfinite numbers*, translated and with an Introduction by Philip E. B. Jourdain, Dover Publications, New York.

Jean van Heijenoort, *From Frege to Gödel*, Harvard University Press, Cambridge, 1967.

How to use it. About half of this book can be covered in a Quarter (ten weeks), somewhat more in a longer Semester. Chapters **1 – 6** cover the beginnings of the subject and they are written in a leisurely manner, so that the serious student can read through them alone, with little help. The trick to using the Notes successfully in a class is to cover these beginnings very quickly: skip the introductory Chapter **1**, which mostly sets notation; spend about a week on Chapter **2**, which explains Cantor's basic ideas; and then proceed with all deliberate speed through Chapters **3 – 6**, so that the theory of well ordered sets in Chapter **7** can be reached no later than the sixth week, preferably the fifth. Beginning with Chapter **7**, the results are harder and the presentation is more compact. How much of the "real" set theory in Chapters **7 – 12** can be covered depends, of course, on the students, the length of the course, and what is passed over. If the class is populated by future computer scientists, for example, then Chapter **6** on Fixed Points should be covered in full, with its problems, but Chapter **10** on Baire Space might be omitted, sad as that sounds. For budding young analysts, at the other extreme, Chapter **6** can be cut off after **6.27** (and this too is sad), but at least part of Chapter **10** should be attempted. Additional material which can be left out, if time is short, includes the detailed development of addition and multiplication on the natural numbers in Chapter **5**, and some of the less central applications of the Axiom of Choice in Chapter **9**. The Appendices are quite unlikely to be taught in a course (I devote just one lecture to explain the idea of the construction of the reals in Appendix **A**), though I would like to think that they might be suitable for undergraduate Honors Seminars, or individual reading courses.

Since elementary courses in set theory are not offered regularly and they are seldom long enough to cover all the basics, I have tried to make these Notes accessible to the serious student who is studying the subject on their own. There are numerous, simple Exercises strewn throughout the text, which test understanding of new notions immediately after they are introduced. In class I present about half of them, as examples, and I assign some of the rest

for easy homework. The Problems at the end of each chapter vary widely in difficulty, some of them covering additional material. The hardest problems are marked with an asterisk (*).

Acknowledgments. I am grateful to the Mathematics Department of the University of Athens for the opportunity to teach there in Fall 1990, when I wrote the first draft of these Notes, and especially to Prof. A. Tsarpalias who usually teaches that Set Theory course and used a second draft in Fall 1991; and to Dimitra Kitsiou and Stratos Paschos for struggling with PCs and laser printers at the Athens Polytechnic in 1990 to produce the first "hard copy" version. I am grateful to my friends and colleagues at UCLA and Caltech (hotbeds of activity in set theory) from whom I have absorbed what I know of the subject, over many years of interaction. I am especially grateful to my wife Joan Moschovakis and my student Darren Kessner for reading large parts of the preliminary edition, doing the problems and discovering a host of errors; and to Larry Moss who taught out of the preliminary edition in the Spring Term of 1993, found the remaining host of errors and wrote out solutions to many of the problems.

The book was written more-or-less simultaneously in Greek and English, by the magic of bilingual LaTeX and in true reflection of my life. I have dedicated it to Prof. Nikos Kritikos (a student of Caratheodory), in fond memory of many unforgettable hours he spent with me back in 1973, patiently teaching me how to speak and write mathematics in my native tongue, but also much about the love of science and the nature of scholarship. In this connection, I am also greatly indebted to Takis Koufopoulos, who read critically the preliminary Greek version, corrected a host of errors and made numerous suggestions which (I believe) improved substantially the language of the final Greek draft.

Palaion Phaliron, Greece   November 1993

About the 2nd edition. Perhaps the most important changes I have made are in small things, which (I hope) will make it easier to teach and learn from this book: simplifying proofs, streamlining notation and terminology, adding a few diagrams, rephrasing results (especially those justifying *definition by recursion*) to ease their applications, and, most significantly, correcting errors, typographical and other. For spotting these errors and making numerous, useful suggestions over the years, I am grateful to Serge Bozon, Joel Hamkins, Peter Hinman, Aki Kanamori, Joan Moschovakis, Larry Moss, Thanassis Tsarpalias and many, many students.

The more substantial changes include:

— A proof of *Suslin's Theorem* in Chapter **10**, which has also been significantly massaged.

— A better exposition of ordinal theory in Chapter **12** and the addition of some material, including the basic facts about ordinal arithmetic.

— The last chapter, a compilation of solutions to the Exercises in the main part of the book – in response to popular demand. This eliminates the most obvious, easy homework assignments, and so I have added some easy problems.

I am grateful to Thanos Tsouanas, who copy-edited the manuscript and caught the worst of my mistakes.

Palaion Phaliron, Greece   July 2005

CONTENTS

INTRODUCTION

Mathematicians have always used sets, e.g., the ancient Greek geometers defined a circle as the set of points at a fixed distance r from a fixed point C, its center. But the systematic study of sets began only at the end of the 19th century with the work of the great German mathematician Georg Cantor, who created a rigorous theory of the concept of *completed infinite* by which we can compare infinite sets as to size. For example, let

$$\mathbb{N} = \{0, 1, \dots\} = \text{the set of natural numbers,}$$
$$\mathbb{Z} = \{\dots, -1, 0, 1, \dots\} = \text{the set of rational integers,}$$
$$\mathbb{Q} = \text{the set of rational numbers (fractions),}$$
$$\mathbb{R} = \text{the points of a straight line,}$$

where we also identify $\mathbb{R}$ with the set of real numbers, each point associated with its (positive or negative) coordinate with respect to a fixed origin and direction. Cantor asked if these four sets "have the same (infinite) number of elements", or if one of them is "more numerous" than the others. Before we make precise and answer this question in the next chapter, we review here some basic, well-known facts about sets and functions, primarily to explain the notation we will be using.

What are sets, anyway? The question is like "what are points", which Euclid answered with

a point is that which has no parts.

This is not a rigorous mathematical definition, a reduction of the concept of "point" to other concepts which we already understand, but just an intuitive description which suggests that a point is some thing which has no extension in space. Like that of point, the concept of set is fundamental and cannot be reduced to other, simpler concepts. Cantor described it as follows:

By a set we are to understand any collection into a whole of definite and separate objects of our intuition or our thought.

Vague as it is, this description implies two basic properties of sets.

1. Every set A has **elements** or **members**. We write

$$x \in A \iff \text{the object } x \text{ is a member of (or belongs to) } A.$$

2. A set is determined by its members, i.e., if A, B are sets, then[1]

$$A = B \iff A \text{ and } B \text{ have the same members} \qquad (1\text{-}1)$$
$$\iff (\forall x)[x \in A \iff x \in B].$$

This last is the **Extensionality Property**. For example, the set of students in this class will not change if we all switch places, lie down or move to another classroom; this set is completely determined by *who we are*, not our posture or the places where we happen to be.

Somewhat peculiar is the **empty set** $\emptyset$ which has no members. The extensionality property implies that *there is only one empty set*.

If A and B are sets, we write

$$A \subseteq B \iff (\forall x)[x \in A \implies x \in B],$$

and if $A \subseteq B$, we call A a **subset** of B, so that for every B,

$$\emptyset \subseteq B, \quad B \subseteq B.$$

A **proper subset** of B is a subset distinct from B,

$$A \subsetneqq B \iff [A \subseteq B \,\&\, A \neq B].$$

From the extensionality property it follows that for all sets A, B,

$$A = B \iff A \subseteq B \,\&\, B \subseteq A.$$

We have already used several different notations to define specific sets and we need still more, e.g.,

$$A = \{a_1, a_2, \ldots, a_n\}$$

is the (finite) set with members the objects $a_1, a_2, \ldots, a_n$. If P is a condition which specifies some property of objects, then

$$A = \{x \mid P(x)\}$$

is the set of all objects which satisfy the condition P, so that for all x,

$$x \in A \iff P(x).$$

For example, if

$$P(x) \iff x \in \mathbb{N} \,\&\, x \text{ is even,}$$

then $\{x \mid P(x)\}$ is the set of all even. natural numbers. We use a variant of this notation when we are only interested in "collecting into a whole" members of a given set A which satisfy a certain condition:

$$\{x \in A \mid P(x)\} =_{\mathrm{df}} \{x \mid x \in A \,\&\, P(x)\},$$

---

[1]We will use systematically, as abbreviations, the logical symbols

$$\&: \text{and}, \ \vee: \text{or}, \ \neg: \text{not}, \ \implies: \text{implies}, \ \iff: \text{if and only if,}$$

$$\forall: \text{for all}, \ \exists: \text{there exists}, \ \exists!: \text{there exists exactly one.}$$

The symbols $=_{\mathrm{df}}$ and $\iff_{\mathrm{df}}$ are read "equal by definition" and "equivalent by definition".
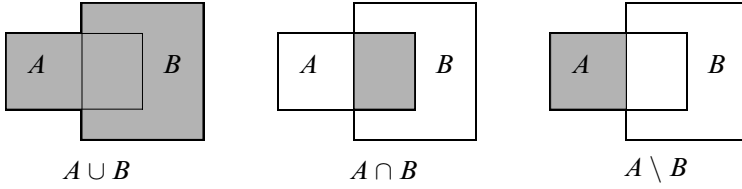
FIGURE **1.1**. The Boolean operations.

so that, for example, $\{x \in \mathbb{N} \mid x > 0\}$ is the set of all non-zero natural numbers, while $\{x \in \mathbb{R} \mid x > 0\}$ is the set of all positive real numbers.

For any two sets[2] A, B,

$$A \cup B = \{x \mid x \in A \vee x \in B\} \quad \text{(the union of } A, B\text{),}$$
$$A \cap B = \{x \in A \mid x \in B\} \quad \text{(the intersection of } A, B\text{),}$$
$$A \setminus B = \{x \in A \mid x \notin B\} \quad \text{(the difference of } A, B\text{).}$$

These "Boolean operations" are illustrated in the so-called *Venn diagrams* of Figure **1.1**, in which sets are represented by regions in the plane. The union and the intersection of infinite sequences of sets are defined in the same way,

$$\bigcup_{n=0}^{\infty} A_n = A_0 \cup A_1 \cup \cdots = \{x \mid (\exists n \in \mathbb{N})[x \in A_n]\},$$
$$\bigcap_{n=0}^{\infty} A_n = A_0 \cap A_1 \cap \cdots = \{x \mid (\forall n \in \mathbb{N})[x \in A_n]\}.$$

Two sets are **disjoint** if their intersection is empty,

$$A \text{ is disjoint from } B \iff A \cap B = \emptyset.$$

We will use the notations

$$f : X \to Y \text{ or } A \xrightarrow{f} B$$

to indicate that f is a **function** which associates with each member x of the set X, the **domain** of f some member $f(x)$ of the **range** Y of f. Functions are also called **mappings**, **operations**, **transformations** and many other things. Sometimes it is convenient to use the abbreviated notation $(x \mapsto f(x))$ which makes it possible to talk about a function without officially naming it. For example,

$$(x \mapsto x^2 + 1)$$

is the function on the real numbers which assigns to each real its square increased by 1; if we call it f, then it is defined by the formula

$$f(x) = x^2 + 1 \quad (x \in \mathbb{R})$$

---

[2]In "mathematical English", when we say "for any two objects x, y", we do not mean that necessarily $x \neq y$, e.g., the assertion that "for any two numbers x, y, $(x + y)^2 = x^2 + 2xy + y^2$" implies that "for every number x, $(x + x)^2 = x^2 + 2xx + x^2$".

so that $f(0) = 1$, $f(2) = 5$, etc. But we can say "all the values of $(x \mapsto x^2 + 1)$ are positive reals" without necessarily fixing a name for it, like f.

Two functions are **equal** if they have the same domain and they assign the same value to every member of their common domain,

$$f = g \iff (\forall x \in X)[f(x) = g(x)] \quad (f : X \to Y, \, g : X \to Z, x \in X).$$

In connection with functions we will also use the notations

$$f : X \rightarrowtail Y \iff_{\text{df}} f \text{ is an } \textbf{injection} \text{ (one-to-one)}$$
$$\iff (\forall x, x' \in X)[f(x) = f(x') \implies x = x'],$$

$$f : X \twoheadrightarrow Y \iff_{\text{df}} f \text{ is a } \textbf{surjection} \text{ (onto)}$$
$$\iff (\forall y \in Y)(\exists x \in X)[f(x) = y],$$

$$f : X \rightarrowtail\!\!\!\to Y \iff_{\text{df}} f \text{ is a } \textbf{bijection} \text{ or a } \textbf{correspondence}$$
$$\iff (\forall y \in Y)(\exists ! x \in X)[f(x) = y].$$

For every $f : X \to Y$ and $A \subseteq X$, the set

$$f[A] =_{\text{df}} \{f(x) \mid x \in A\}$$

is the **image** of A under f, and if $B \subseteq Y$, then

$$f^{-1}[B] =_{\text{df}} \{x \in X \mid f(x) \in B\}$$

is the **pre-image** of B by f.

If f is a bijection, then we can define the **inverse function** $f^{-1} : Y \to X$ by the condition

$$f^{-1}(y) = x \iff f(x) = y,$$

and then the inverse image $f^{-1}[B]$ (as we defined it above) is precisely the image of B under f^{-1}.

The **composition**

$$h =_{\text{df}} gf : X \to Z$$

of two functions

$$X \xrightarrow{f} Y \xrightarrow{g} Z$$

is defined by

$$h(x) = g(f(x)) \quad (x \in X).$$

It is easy to prove many basic properties of sets and functions using only these definitions and the extensionality property. For example,

$$A \cup B = B \cup A,$$

because, for any x,

$$x \in A \cup B \iff x \in A \text{ or } x \in B$$
$$\iff x \in B \text{ or } x \in A$$
$$\iff x \in B \cup A.$$

In some cases, the logic of the argument gets a bit complex and it is easier to prove an identity $U = V$ by verifying separately the two implications

$x \in U \Longrightarrow x \in V$ and $x \in V \Longrightarrow x \in U$. For example, if $f : X \to Y$ and $A, B \subseteq X$, then

$$f[A \cup B] = f[A] \cup f[B].$$

To prove this, we show first that

$$x \in f[A \cup B] \Longrightarrow x \in f[A] \cup f[B];$$

this holds because if $x \in f[A \cup B]$, then there is some $y \in A \cup B$ such that $x = f(y)$; and if $y \in A$, then $x = f(y) \in f[A] \subseteq f[A] \cup f[B]$, while if $y \in B$, then $x = f(y) \in f[B] \subseteq f[A] \cup f[B]$. Next we show the converse implication, that

$$x \in f[A] \cup f[B] \Longrightarrow x \in f[A \cup B];$$

this holds because if $x \in f[A]$, then $x = f(y)$ for some $y \in A \subseteq A \cup B$, and so $x \in f[A \cup B]$, while if $x \in f[B]$, then $x = f(y)$ for some $y \in B \subseteq A \cup B$, and so, again, $x \in f[A \cup B]$.

Problems for Chapter 1

x1.1. For any three sets A, B, C,

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \setminus (A \cap B) = A \setminus B.$$

x1.2. (De Morgan's laws) For any three sets A, B, C,

$$C \setminus (A \cup B) = (C \setminus A) \cap (C \setminus B),$$
$$C \setminus (A \cap B) = (C \setminus A) \cup (C \setminus B).$$

x1.3. (De Morgan's laws for sequences) For any set C and any sequence of sets $\{A_n\}_n = A_0, A_1, \ldots,$

$$C \setminus (\bigcup_n A_n) = \bigcap_n (C \setminus A_n),$$
$$C \setminus (\bigcap_n A_n) = \bigcup_n (C \setminus A_n).$$

x1.4. For every injection $f : X \rightarrowtail Y$, and all $A, B \subseteq X$,

$$f[A \cap B] = f[A] \cap f[B],$$
$$f[A \setminus B] = f[A] \setminus f[B].$$

Show also that these identities do not always hold if f is not an injection.

x1.5. For every $f : X \to Y$, and all $A, B \subseteq Y$,

$$f^{-1}[A \cup B] = f^{-1}[A] \cup f^{-1}[B],$$
$$f^{-1}[A \cap B] = f^{-1}[A] \cap f^{-1}[B],$$
$$f^{-1}[A \setminus B] = f^{-1}[A] \setminus f^{-1}[B].$$

x1.6. For every $f : X \to Y$ and all sequences of sets $A_n \subseteq X$, $B_n \subseteq Y$,

$$f^{-1}[\bigcup_{n=0}^{\infty} B_n] = \bigcup_{n=0}^{\infty} f^{-1}[B_n],$$

$$f^{-1}[\bigcap_{n=0}^{\infty} B_n] = \bigcap_{n=0}^{\infty} f^{-1}[B_n],$$

$$f[\bigcup_{n=0}^{\infty} A_n] = \bigcup_{n=0}^{\infty} f[A_n].$$

x1.7. For every injection $f : X \rightarrowtail Y$ and every sequence of sets $A_n \subseteq X$,

$$f[\bigcap_{n=0}^{\infty} A_n] = \bigcap_{n=0}^{\infty} f[A_n].$$

x1.8. The composition of injections is an injection, the composition of surjections is a surjection, and hence the composition of bijections is a bijection.

EQUINUMEROSITY

After these preliminaries, we can formulate the fundamental definitions of Cantor about the size or cardinality of sets.

2.1. Definition. Two sets A, B are **equinumerous** or **equal in cardinality** if there exists a (one-to-one) correspondence between their elements, in symbols

$$A =_c B \iff_{\text{df}} (\exists f)[f : A \rightarrowtail\!\!\!\rightarrow B].$$

This definition of equinumerosity stems from our intuitions about finite sets, e.g., we can be sure that a shoe store offers for sale the same number of left and right shoes without knowing exactly what that number is: the correspondence of each left shoe with the right shoe in the same pair establishes the equinumerosity of these two sets. The radical element in Cantor's definition is the proposal to accept the existence of such a correspondence as the characteristic property of equinumerosity for all sets, despite the fact that its application to infinite sets leads to conclusions which had been viewed as counterintuitive. A finite set, for example, cannot be equinumerous with one of its proper subsets, while the set of natural numbers $\mathbb{N}$ is equinumerous with $\mathbb{N} \setminus \{0\}$ via the correspondence $(x \mapsto x + 1)$,

$$\{0, 1, 2, \dots\} =_c \{1, 2, 3, \dots\}.$$

In the real numbers, also,

$$(0, 1) =_c (0, 2)$$

via the correspondence $(x \mapsto 2x)$, where as usual, for any two reals $\alpha < \beta$

$$(\alpha, \beta) = \{x \in \mathbb{R} \mid \alpha < x < \beta\}.$$

We will use the analogous notation for the closed and half-closed intervals $[\alpha, \beta]$, $[\alpha, \beta)$, etc.

2.2. Proposition. *For all sets A, B, C,*

$$A =_c A,$$
$$\text{if } A =_c B, \text{ then } B =_c A,$$
$$\text{if } (A =_c B \,\&\, B =_c C), \text{ then } A =_c C.$$

PROOF. To show the third implication as an example, suppose that the bijections $f : A \rightarrowtail\!\!\!\rightarrow B$ and $g : B \rightarrowtail\!\!\!\rightarrow C$ witness the equinumerosities of the hypothesis; their composition $gf : A \rightarrowtail\!\!\!\rightarrow C$ then witnesses that $A =_c C$. ⊣
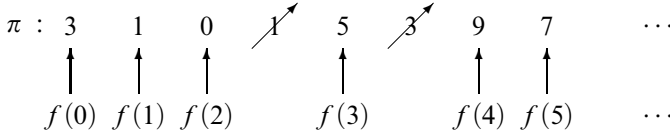
FIGURE **2.1**. Deleting repetitions.

2.3. Definition. The set A is **less than or equal to** B **in size** if it is equinumerous with some subset of B, in symbols:

$$A \leq_c B \iff (\exists C)[C \subseteq B \,\&\, A =_c C].$$

2.4. Proposition. $A \leq_c B \iff (\exists f)[f : A \rightarrowtail B]$.

PROOF. If $A =_c C \subseteq B$ and $f : A \twoheadrightarrow C$ witnesses this equinumerosity, then f is an injection from A into B. Conversely, if there exists an injection $f : A \rightarrowtail B$, then the same f is a bijection of A with its image $f[A]$, so that $A =_c f[A] \subseteq B$ and so $A \leq_c B$ by the definition. $\dashv$

2.5. Exercise. *For all sets A, B, C,*

$$A \leq_c A,$$
$$\text{if } (A \leq_c B \,\&\, B \leq_c C), \text{ then } A \leq_c C.$$

2.6. Definition. A set A is **finite** if there exists some natural number n such that

$$A =_c \{i \in \mathbb{N} \mid i < n\} = \{0, 1, \dots, n-1\},$$

otherwise A is **infinite**. (Thus the empty set is finite, since $\emptyset = \{i \in \mathbb{N} \mid i < 0\}$.)

A set A is **countable** if it is finite or equinumerous with the set of natural numbers $\mathbb{N}$, otherwise it is **uncountable**. Countable sets are also called **denumerable**, and correspondingly, uncountable sets are **non-denumerable**.

2.7. Proposition. *The following are equivalent for every set A:*
   (1) *A is countable.*
   (2) *$A \leq_c \mathbb{N}$.*
   (3) *Either $A = \emptyset$, or A has an **enumeration**, a surjection $\pi : \mathbb{N} \to A$, so that*

$$A = \pi[\mathbb{N}] = \{\pi(0), \pi(1), \pi(2), \dots \}.$$

PROOF. We give what is known as a "round robin proof".
   (1) $\implies$ (2). If A is countable, then either $A =_c \{i \in \mathbb{N} \mid i < n\}$ for some n or $A =_c \mathbb{N}$, so that, in either case, $A =_c C$ for some $C \subseteq \mathbb{N}$ and hence $A \leq_c \mathbb{N}$.
   (2) $\implies$ (3). Suppose $A \neq \emptyset$, choose some $x_0 \in A$, and assume by (2) that $f : A \rightarrowtail \mathbb{N}$. For each $i \in \mathbb{N}$, let

$$\pi(i) = \begin{cases} x_0, & \text{if } i \notin f[A], \\ f^{-1}(i), & \text{otherwise, i.e., if } i \in f[A]. \end{cases}$$

The definition works (because f is an injection, and so $f^{-1}(i)$ is uniquely determined in the second case), and it defines a surjection $\pi : \mathbb{N} \twoheadrightarrow A$, because $x_0 \in A$ and for every $x \in A$, $x = \pi(f(x))$.

(3) $\implies$ (1). If A is finite then (1) is automatically true, so assume that A is infinite but it has an enumeration $\pi : \mathbb{N} \twoheadrightarrow A$. We must find another enumeration $f : \mathbb{N} \twoheadrightarrow A$ which is *without repetitions*, so that it is in fact a bijection of $\mathbb{N}$ with A, and hence $A =_c \mathbb{N}$. The proof is suggested by Figure **2.1**: we simply delete the repetitions from the given enumeration π of A. To get a precise definition of f by recursion, notice that because A is not finite, for every finite sequence $a_0, \dots, a_n$ of members of A there exists some m such that $\pi(m) \notin \{a_0, \dots, a_n\}$. Set

$$f(0) = \pi(0),$$
$$m_n = \text{the least } m \text{ such that } \pi(m) \notin \{f(0), \dots, f(n)\},$$
$$f(n+1) = \pi(m_n).$$

It is obvious that f is an injection, so it is enough to verify that every $x \in A$ is a value of f, i.e., that *for every* $n \in \mathbb{N}$, $\pi(n) \in f[\mathbb{N}]$. This is immediate for 0, since $\pi(0) = f(0)$. If $x = \pi(n+1)$ for some n and $x \in \{f(0), \dots, f(n)\}$, then $x = f(i)$ for some $i \leq n$; and if $x \notin \{f(0), \dots, f(n)\}$, then $m_n = n+1$ and $f(n+1) = \pi(m_n) = x$ by the definition. $\dashv$

2.8. Exercise. *If A is countable and there exists an injection $f : B \rightarrowtail A$, then B is also countable; in particular, every subset of a countable set is countable.*

2.9. Exercise. *If A is countable and there exists a surjection $f : A \twoheadrightarrow B$, then B is also countable.*

The next, simple theorem is one of the most basic results of set theory.

2.10. Theorem (Cantor). *For each sequence $A_0, A_1, \dots$ of countable sets, the union*

$$A = \bigcup_{n=0}^{\infty} A_n = A_0 \cup A_1 \cup \dots$$

is also a countable set.

In particular, the union $A \cup B$ of two countable sets is countable.

PROOF. The second claim follows by applying the first to the sequence

$$A, B, B, \cdots$$

For the first, it is enough (why?) to consider the special case where none of the A_n is empty, in which case we can find for each A_n an enumeration $\pi^n : \mathbb{N} \twoheadrightarrow A_n$. If we let

$$a_i^n = \pi^n(i)$$

to simplify the notation, then for each n

$$A_n = \{a_0^n, a_1^n, \dots\},$$

and we can construct from these enumerations a table of elements which lists all the members of the union A. This is pictured in Figure **2.2**, and the arrows
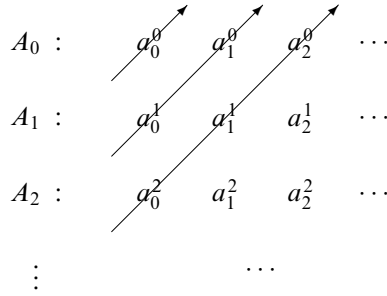
FIGURE **2.2**. Cantor's first diagonal method.

in that picture show how to enumerate the union:

$$A = \{a_0^0, a_0^1, a_1^0, a_0^2, a_1^1, \dots\}. \qquad \dashv$$

2.11. Corollary. *The set of rational* (*positive and negative*) *integers*

$$\mathbb{Z} = \{\dots -2, -1, 0, 1, 2, \dots\}$$

is countable.

PROOF. $\mathbb{Z} = \mathbb{N} \cup \{-1, -2, \dots, \}$ and the set of negative integers is countable via the correspondence $(x \mapsto -(x+1))$. $\qquad \dashv$

2.12. Corollary. *The set $\mathbb{Q}$ of rational numbers is countable.*

PROOF. The set $\mathbb{Q}^+$ of non-negative rationals is countable because

$$\mathbb{Q}^+ = \bigcup_{n=1}^\infty \{\frac{m}{n} \mid m \in \mathbb{N}\}$$

and each $\{\frac{m}{n} \mid m \in \mathbb{N}\}$ is countable via the enumeration $(m \mapsto \frac{m}{n})$. The set $\mathbb{Q}^-$ of negative rationals is countable by the same method, and then the union $\mathbb{Q}^+ \cup \mathbb{Q}^-$ is countable. $\qquad \dashv$

This corollary was Cantor's first significant result in the program of classification of infinite sets by their size, and it was considered somewhat "paradoxical" because $\mathbb{Q}$ appears to be so much larger than $\mathbb{N}$. Immediately afterwards, Cantor showed the existence of uncountable sets.

2.13. Theorem (Cantor). *The set of infinite, binary sequences*

$$\Delta = \{(a_0, a_1, \dots,) \mid (\forall i)[a_i = 0 \lor a_i = 1]\}$$

is uncountable.

PROOF. Suppose (towards a contradiction) that Δ is countable, so there exists an enumeration

$$\Delta = \{\alpha_0, \alpha_1, \dots\},$$

where for each n,

$$\alpha_n = (a_0^n, a_1^n, \dots)$$

$$\alpha_0 : \qquad a_0^0 \qquad a_1^0 \qquad a_2^0 \qquad \cdots$$

$$\alpha_1 : \qquad a_0^1 \qquad a_1^1 \qquad a_2^1 \qquad \cdots$$

$$\alpha_2 : \qquad a_0^2 \qquad a_1^2 \qquad a_2^2 \qquad \cdots$$

$$\vdots \qquad\qquad \cdots$$
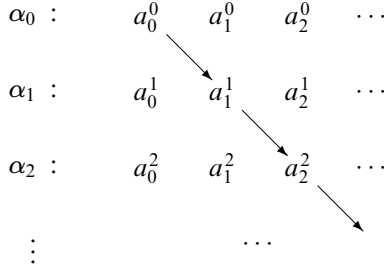
FIGURE **2.3**. Cantor's second diagonal method.

is a sequence of 0's and 1's.[3] We construct a table with these sequences as before, and then we define the sequence β by interchanging 0 and 1 in the "diagonal" sequence $a_0^0, a_1^1, \ldots$:

$$\beta(n) = 1 - a_n^n.$$

It is obvious that for each α_n, $\beta \neq \alpha_n$, since

$$\beta(n) = 1 - \alpha_n(n) \neq \alpha_n(n),$$

so that the sequence $\alpha_0, \alpha_1, \ldots$ does not enumerate the entire Δ, contrary to our hypothesis. $\dashv$

2.14. Corollary (Cantor). *The set $\mathbb{R}$ of real numbers is uncountable.*

PROOF. We define first a sequence of sets $C_0, C_1, \ldots$, of real numbers which satisfy the following conditions:

1. $C_0 = [0, 1]$.
2. Each C_n is a union of 2^n closed intervals and

$$C_0 \supseteq C_1 \supseteq \cdots C_n \supseteq C_{n+1} \supseteq \cdots .$$

3. C_{n+1} is constructed by removing the (open) middle third of each interval in C_n, i.e., by replacing each $[a, b]$ in C_n by the two closed intervals

$$L[a, b] = [a, a + \frac{1}{3}(b - a)],$$

$$R[a, b] = [a + \frac{2}{3}(b - a), b].$$

With each binary sequence $\delta \in \Delta$ we associate now a sequence of closed intervals,

$$F_0^\delta, F_1^\delta, \ldots ,$$

---

[3]To prove a proposition θ by the method of *reduction to a contradiction*, we assume its negation $\neg\theta$ and derive from that assumption something which violates known facts, a contradiction, something *absurd:* we conclude that θ cannot be false, so it must be true. Typically we will begin such arguments with the code-phrase *towards a contradiction*, which alerts the reader that the supposition which follows is the negation of what we intend to prove.

FIGURE **2.4**. The first four stages of the Cantor set construction.

by the following recursion:

$$F_0^\delta = \mathcal{C}_0 = [0, 1],$$

$$F_{n+1}^\delta = \begin{cases} LF_n^\delta, & \text{if } \delta(n) = 0, \\ RF_n^\delta, & \text{if } \delta(n) = 1. \end{cases}$$

By induction, for each n, F_n^δ is one of the closed intervals of $\mathcal{C}_n$ of length 3^{-n} and obviously

$$F_0^\delta \supseteq F_1^\delta \supseteq \cdots,$$

so by the fundamental **completeness property** of the real numbers the intersection of this sequence is not empty; in fact, it contains exactly one real number, call it

$$f(\delta) = \text{ the unique element in the intersection } \bigcap_{n=0}^{\infty} F_n^\delta.$$

The function f maps the uncountable set Δ into the set

$$\mathcal{C} = \bigcap_{n=0}^{\infty} \mathcal{C}_n,$$

the so-called **Cantor set**, so to complete the proof it is enough to verify that f is one-to-one. But if n is the least number for which $\delta(n) \neq \varepsilon(n)$ and (for example) $\delta(n) = 0$, we have $F_n^\delta = F_n^\varepsilon$ from the choice of n,

$$f(\delta) \in F_{n+1}^\delta = LF_n^\delta, \ f(\varepsilon) \in F_{n+1}^\varepsilon = RF_n^\delta, \text{ and } LF_n^\delta \cap RF_n^\delta = \emptyset,$$

so that indeed f is an injection.   ⊣

The basic mathematical ingredient of this proof is the appeal to the completeness property of the real numbers, which we will study carefully in Appendix **A**. Some use of a special property of the reals is necessary: the rest of Cantor's construction relies solely on arithmetical properties of numbers which are also true of the rationals, so if we could avoid using completeness we would also prove that $\mathbb{Q}$ is uncountable, contradicting Corollary **2.12**.

The fundamental importance of this theorem was instantly apparent, the more so because Cantor used it immediately in a significant application to the theory of algebraic numbers. Before we prove this corollary we need some definitions and lemmas.

2.15. Definition. For any two sets A, B, the set of **ordered pairs** of members of A and members of B is denoted by

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

In the same way, for each $n \geq 2$,

$$A_1 \times \cdots \times A_n = \{(x_1, \ldots, x_n) \mid x_1 \in A_1, \ldots, x_n \in A_n\},$$
$$A^n = \{(x_1, \ldots, x_n) \mid x_1, \ldots, x_n \in A\}.$$

We call $A_1 \times \cdots \times A_n$ the **Cartesian product** of $A_1, \ldots, A_n$.

2.16. Lemma. (1) *If $A_1, \ldots, A_n$ are all countable, so is their Cartesian product $A_1 \times \cdots \times A_n$.*
   (2) *For every countable set A, each A^n ($n \geq 2$) and the union*

$$\bigcup_{n=2}^{\infty} A^n = \{(x_1, \ldots, x_n) \mid n \geq 2, x_1, \ldots, x_n \in A\}$$

are all countable.

   PROOF. (1) If some A_i is empty, then the product is empty (by the definition) and hence countable. Otherwise, in the case of two sets A, B, we have some enumeration

$$B = \{b_0, b_1, \ldots\}$$

of B, obviously

$$A \times B = \bigcup_{n=0}^{\infty} (A \times \{b_n\}),$$

and each $A \times \{b_n\}$ is equinumerous with A (and hence countable) via the correspondence $(x \mapsto (x, b_n))$. This gives the result for $n = 2$. To prove the proposition for all $n \geq 2$, notice that

$$A_1 \times \cdots \times A_n \times A_{n+1} =_c (A_1 \times \cdots \times A_n) \times A_{n+1}$$

via the bijection

$$f(a_1, \ldots, a_n, a_{n+1}) = ((a_1, \ldots, a_n), a_{n+1}).$$

Thus, if every product of $n \geq 2$ countable factors is countable, so is every product of $n + 1$ countable factors, and so (1) follows by induction.
   (2) Each A^n is countable by (1), and then $\bigcup_{n=2}^{\infty} A^n$ is also countable by another appeal to Theorem **2.10**.   ⊣

2.17. Definition. A real number α is **algebraic** if it is a root of some polynomial

$$P(x) = a_0 + a_1 x + \cdots + a_n x^n$$

with integer coefficients $a_0, \ldots, a_n \in Z$ ($n \geq 1, a_n \neq 0$), i.e., if

$$P(\alpha) = 0.$$

   Typical examples of algebraic numbers are $\sqrt{2}$, $(1 + \sqrt{2})^2$ (why?) but also the real root of the equation $x^5 + x + 1 = 0$ which exists (why?) but cannot be expressed in terms of radicals, by a classical theorem of Abel. The basic fact (from algebra) about algebraic numbers is that *a polynomial of degree $n \geq 1$ has at most n real roots*; this is all we need for the next result.

2.18. Corollary. *The set K of algebraic real numbers is countable* (Cantor), *and hence there exist real numbers which are not algebraic* (Liouville).

Proof. The set Π of all polynomials with integer coefficients is countable, because each such polynomial is determined by the sequence of its coefficients, so that Π can be injected into the countable set $\bigcup_{n=2}^{\infty} Z^n$. For each polynomial $P(x)$, the set of its roots

$$\Lambda(P(x)) = \{\alpha \mid P(\alpha) = 0\}$$

is finite and hence countable. It follows that the set of algebraic numbers K is the union of a sequence of countable sets and hence it is countable.   ⊣

This first application of the (then) new theory of sets was instrumental in ensuring its quick and favorable acceptance by the mathematicians of the period, particularly since the earlier proof of Liouville (that there exist non-algebraic numbers) was quite intricate. Cantor showed something stronger, that "almost all" real numbers are not algebraic, and he did it with a much simpler proof which used just the fact that a polynomial of degree n cannot have more than n real roots, the completeness of $\mathbb{R}$, and, of course, the new method of *counting the members of infinite sets*.

So far we have shown the existence of only two "orders of infinity", that of $\mathbb{N}$—the countable, infinite sets—and that of $\mathbb{R}$. There are many others.

2.19. Definition. The **powerset** $\mathcal{P}(A)$ of a set A is the set of all its subsets,

$$\mathcal{P}(A) = \{X \mid X \text{ is a set and } X \subseteq A\}.$$

2.20. Exercise. *For all sets A, B,*

$$A =_c B \Longrightarrow \mathcal{P}(A) =_c \mathcal{P}(B).$$

2.21. Theorem (Cantor). *For every set A,*

$$A <_c \mathcal{P}(A),$$

i.e., $A \leq_c \mathcal{P}(A)$ but $A \neq_c \mathcal{P}(A)$; in fact there is no surjection $\pi : A \twoheadrightarrow \mathcal{P}(A)$.

Proof. That $A \leq_c \mathcal{P}(A)$ follows from the fact that the function

$$(x \mapsto \{x\})$$

which associates with each member x of A its **singleton** $\{x\}$ is an injection. (Careful here: the singleton $\{x\}$ is a set with just the one member x and it is not the same object as x, which is probably not a set to begin with!)

To complete the proof, we assume (towards a contradiction) that there exists a surjection

$$\pi : A \twoheadrightarrow \mathcal{P}(A),$$

and we define the set

$$B = \{x \in A \mid x \notin \pi(x)\},$$

so that for every $x \in A$,

$$x \in B \iff x \notin \pi(x). \tag{2-1}$$

Now B is a subset of A and π is a surjection, so there must exist some $b \in A$ such that $B = \pi(b)$; and setting $x = b$ and $\pi(b) = B$ in (2-1), we get

$$b \in B \iff b \notin B$$

which is absurd.   ⊣

So there are many orders of infinity, and specifically (at least) those of the sets

$$\mathbb{N} <_c \mathcal{P}(\mathbb{N}) <_c \mathcal{P}(\mathcal{P}(\mathbb{N})) <_c \cdots .$$

If we name these sets by the recursion

$$\begin{aligned} T_0 &= \mathbb{N}, \\ T_{n+1} &= \mathcal{P}(T_n), \end{aligned} \tag{2-2}$$

then their union $T_\infty = \bigcup_{n=0}^{\infty} T_n$ has a larger cardinality than each T_n, Problem **x2.8**. The classification and study of these orders of infinity is one of the central problems of set theory.

Somewhat more general than powersets are *function spaces*.

2.22. Definition. For any two sets A, B,

$$\begin{aligned} (A \to B) &=_{\mathrm{df}} \{f \mid f : A \to B\} \\ &= \quad \text{the set of all functions from } A \text{ to } B. \end{aligned}$$

2.23. Exercise. *If $A_1 =_c A_2$ and $B_1 =_c B_2$, then $(A_1 \to B_1) =_c (A_2 \to B_2)$.*

Function spaces are "generalizations" of powersets because each subset $X \subseteq A$ can be represented by its *characteristic function* $c_X : A \to \{0, 1\}$,

$$c_X(t) = \begin{cases} 1, & \text{if } t \in A \cap X, \\ 0, & \text{if } t \in A \setminus X, \end{cases} \quad (t \in A). \tag{2-3}$$

We can recover X from c_X,

$$X = \{t \in A \mid c_X(t) = 1\},$$

and so the mapping $(X \mapsto c_X)$ is a correspondence of $\mathcal{P}(A)$ with $(A \to \{0, 1\})$. Thus

$$(A \to \{0, 1\}) =_c \mathcal{P}(A) >_c A, \tag{2-4}$$

and the function space operation also leads to large, uncountable sets. The next obvious problem is to compare for size these uncountable sets, starting with the two simplest ones, $\mathcal{P}(\mathbb{N})$ and the set $\mathbb{R}$ of real numbers.

2.24. Lemma. $\mathcal{P}(\mathbb{N}) \leq_c \mathbb{R}$.

PROOF. It is enough to prove that $\mathcal{P}(\mathbb{N}) \leq_c \Delta$, since we have already shown that $\Delta \leq_c \mathbb{R}$. This follows immediately from (2-4), as $\Delta = (\mathbb{N} \to \{0, 1\})$.   ⊣
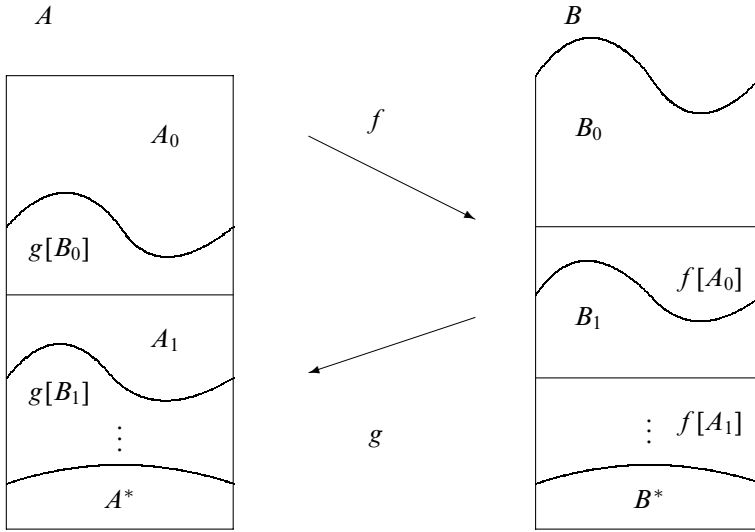
Figure **2.5**. Proof of the Schröder-Bernstein Theorem.

2.25. Lemma. $\mathbb{R} \leq_c \mathcal{P}(\mathbb{N})$.

PROOF. It is enough to show that $\mathbb{R} \leq_c \mathcal{P}(\mathbb{Q})$, since the set of rationals $\mathbb{Q}$ is equinumerous with $\mathbb{N}$ and hence $\mathcal{P}(\mathbb{N}) =_c \mathcal{P}(\mathbb{Q})$. This follows from the fact that the function

$$x \mapsto \pi(x) = \{q \in \mathbb{Q} \mid q < x\} \subseteq \mathbb{Q}$$

is an injection, because if $x < y$ are distinct real numbers, then there exists some rational q between them, $x < q < y$ and $q \in \pi(y) \setminus \pi(x)$. ⊣

With these two simple Lemmas, the equinumerosity $\mathbb{R} =_c \mathcal{P}(\mathbb{N})$ will follow immediately from the following basic theorem.

2.26. Theorem (Schröder-Bernstein). *For any two sets A, B,*

$$\text{if } A \leq_c B \text{ and } B \leq_c A, \text{ then } A =_c B.$$

PROOF.[4] We assume that there exist injections

$$f : A \rightarrowtail B, \; g : B \rightarrowtail A,$$

and we define the sets A_n, B_n by the following recursive definitions:

$$A_0 = A, \qquad B_0 = B,$$
$$A_{n+1} = gf[A_n], \quad B_{n+1} = fg[B_n],$$

---

[4]A different proof of this theorem is outlined in Problems **x4.26**, **x4.27**.

where $fg[X] = \{f(g(x)) \mid x \in X\}$ and correspondingly for the function gf. By induction on n (easily)

$$A_n \supseteq g[B_n] \supseteq A_{n+1},$$
$$B_n \supseteq f[A_n] \supseteq B_{n+1},$$

so that we have the "chains of inclusions"

$$A_0 \supseteq g[B_0] \supseteq A_1 \supseteq g[B_1] \supseteq A_2 \cdots,$$
$$B_0 \supseteq f[A_0] \supseteq B_1 \supseteq f[A_1] \supseteq B_2 \cdots.$$

We also define the intersections

$$A^* = \bigcap_{n=0}^{\infty} A_n, \ B^* = \bigcap_{n=0}^{\infty} B_n,$$

so that

$$B^* = \bigcap_{n=0}^{\infty} B_n \supseteq \bigcap_{n=0}^{\infty} f[A_n] \supseteq \bigcap_{n=0}^{\infty} B_{n+1} = B^*$$

and since f is an injection, by Problem **x1.7**,

$$f[A^*] = f[\bigcap_{n=0}^{\infty} A_n] = \bigcap_{n=0}^{\infty} f[A_n] = B^*.$$

Thus f is a bijection of A^* with B^*. On the other hand,

$$A = A^* \cup (A_0 \setminus g[B_0]) \cup (g[B_0] \setminus A_1) \cup (A_1 \setminus g[B_1]) \cup (g[B_1] \setminus A_2) \ldots$$
$$B = B^* \cup (B_0 \setminus f[A_0]) \cup (f[A_0] \setminus B_1) \cup (B_1 \setminus f[A_1]) \cup (f[A_1] \setminus B_2) \ldots$$

and these sequences are separated, i.e., no set in them has any common element with any other. To finish the proof it is enough to check that for every n,

$$f[A_n \setminus g[B_n]] = f[A_n] \setminus B_{n+1},$$
$$g[B_n \setminus f[A_n]] = g[B_n] \setminus A_{n+1},$$

from which the first (for example) is true because f is an injection and so

$$f[A_n \setminus g[B_n]] = f[A_n] \setminus fg[B_n] = f[A_n] \setminus B_{n+1}.$$

Finally we have the bijection $\pi : A \rightarrowtail\!\!\!\rightarrow B$,

$$\pi(x) = \begin{cases} f(x), & \text{if } x \in A^* \text{ or } (\exists n)[x \in A_n \setminus g[B_n]], \\ g^{-1}(x), & \text{if } x \notin A^* \text{ and } (\exists n)[x \in g[B_n] \setminus A_{n+1}], \end{cases}$$

which verifies that $A =_c B$ and finishes the proof. ⊣

Using the Schröder-Bernstein Theorem we can establish easily several equinumerosities which are quite difficult to prove directly.

Problems for Chapter 2

x2.1. For any $\alpha < \beta$ where α, β are reals, ∞ or $-\infty$, construct bijections which prove the equinumerosities

$$(\alpha, \beta) =_c (0, 1) =_c \mathbb{R}.$$

****x2.2.** For any two real numbers $\alpha < \beta$, construct a bijection which proves the equinumerosity

$$[\alpha, \beta) =_c [\alpha, \beta] =_c \mathbb{R}.$$

x2.3. $\mathcal{P}(\mathbb{N}) =_c \mathbb{R} =_c \mathbb{R}^n$, for every $n \geq 2$.

x2.4. For any two sets A, B, $(A \rightarrow B) \leq_c \mathcal{P}(A \times B)$. HINT. Represent each $f : A \rightarrow B$ by its *graph*, the set

$$G_f = \{(x, y) \in A \times B \mid y = f(x)\}.$$

x2.5. $(\mathbb{N} \rightarrow \mathbb{N}) =_c \mathcal{P}(\mathbb{N})$.

*__x2.6.__ $(\mathbb{N} \rightarrow \mathbb{R}) =_c \mathbb{R}$.

*__x2.7.__ For any three sets A, B, C,

$$((A \times B) \rightarrow C) =_c (A \rightarrow (B \rightarrow C)).$$

HINT. For any $p : A \times B \rightarrow C$, define $\pi(p) = q : A \rightarrow (B \rightarrow C)$ by the formula

$$q(x)(y) = p(x, y).$$

x2.8. Using the definition (2-2), for every m,

$$T_m <_c T_\infty = \bigcup_{n=0}^{\infty} T_n.$$

You need to know something about continuous functions to do the last two problems.

*__x2.9.__ The set $C[0, 1]$ of all continuous, real functions on the closed interval $[0, 1]$ is equinumerous with $\mathbb{R}$.

*__x2.10.__ The set of all monotone real functions on the closed interval $[0, 1]$ is equinumerous with $\mathbb{R}$.

PARADOXES AND AXIOMS

In the preceding chapter we gave a brief exposition of the first, basic results of set theory, as it was created by Cantor and the pioneers who followed him in the last twenty five years of the 19th century. By the beginning of the 20th century, the theory had matured and justified itself with diverse and significant applications, especially in mathematical analysis. Perhaps its greatest success was the creation of an exceptionally beautiful and useful *transfinite arithmetic*, which introduces and studies the operations of addition, multiplication and exponentiation on infinite numbers. By 1900, there were still two fundamental problems about equinumerosity which remained unsolved. These have played a decisive role in the subsequent development of set theory and we will consider them carefully in the following chapters. Here we just state them, in the form of hypotheses.

3.1. Cardinal Comparability Hypothesis.[5] *For any two sets A, B, either $A \leq_c B$ or $B \leq_c A$.*

3.2. Continuum Hypothesis. *There is no set of real numbers X with cardinality intermediate between those of $\mathbb{N}$ and $\mathbb{R}$, i.e.,*

$$(\textbf{CH}) \qquad (\forall X \subseteq \mathbb{R})[X \leq_c \mathbb{N} \vee X =_c \mathbb{R}].$$

Since $\mathbb{R} =_c \mathcal{P}(\mathbb{N})$, **CH** is a special case of the **Generalized Continuum Hypothesis**, the statement that for every infinite set A,

$$(\textbf{GCH}) \qquad (\forall X \subseteq \mathcal{P}(A))[X \leq_c A \vee X =_c \mathcal{P}(A)].$$

If both of these hypotheses are true, then the natural numbers $\mathbb{N}$ and the reals $\mathbb{R}$ represent the two smallest "orders of infinity": every set is either countable, or equinumerous with $\mathbb{R}$, or strictly greater than $\mathbb{R}$ in cardinality.

In this beginning "naive" phase, set theory was developed on the basis of Cantor's definition of sets quoted in Chapter **1**, much as we proved its basic results in Chapter **2**. If we analyze carefully the proofs of those results, we

---

[5]Cantor announced the "theorem of comparability of cardinals" in 1895 and in 1899 he outlined a proposed proof of it in a letter to Dedekind, which was not, however, published until 1932. There were problems with that argument and it is probably closer to the truth to say that until 1900 (at least) the question of comparability of cardinals was still open.

will see that they are all based on the Extensionality Property (1-1) and the following simple assumption.

3.3. General Comprehension Principle. For each n-ary definite condition P, there is a set

$$A = \{\vec{x} \mid P(\vec{x})\}$$

whose members are precisely all the n-tuples of objects which satisfy $P(\vec{x})$, so that for all $\vec{x}$,

$$\vec{x} \in A \iff P(\vec{x}). \tag{3-1}$$

The extensionality principle implies that at most one set A can satisfy (3-1), and we call this A the **extension** of the condition P.

3.4. Definite conditions and operations. It is necessary to restrict the comprehension principle to definite conditions to avoid questions of vagueness which have nothing to do with science. We do not want to admit the "set"

$$A =_{\mathrm{df}} \{x \mid x \text{ is an honest politician}\},$$

because membership of some specific public figure in it may be a hotly debated topic. *An n-ary* **condition** *P is* **definite** *if for each n-tuple $\vec{x} = (x_1, \ldots, x_n)$ of objects, it is determined unambiguously whether $P(\vec{x})$ is true or false.* For example, the binary conditions P and S defined by

$$P(x, y) \iff_{\mathrm{df}} x \text{ is a parent of } y,$$
$$S(s, t) \iff_{\mathrm{df}} s \text{ and } t \text{ are siblings}$$
$$\iff (\exists x)[P(x, s) \,\&\, P(x, t)]$$

are both definite, assuming (for the example) that the laws of biology determine parenthood unambiguously. The General Comprehension Principle applies to them and we can form the sets of pairs

$$A =_{\mathrm{df}} \{(x, y) \mid x \text{ is a parent of } y\},$$
$$B =_{\mathrm{df}} \{(s, t) \mid s \text{ and } t \text{ are siblings}\}.$$

We do not demand of a definite condition that its truth value be *effectively determined*. For example, it is a famous open problem of number theory whether there exist infinitely many pairs of successive, odd primes, and the truth or falsity of the condition

$$G(n) \iff_{\mathrm{df}} n \in \mathbb{N} \,\&\, (\exists m > n)[m, m + 2 \text{ are both prime numbers}]$$

is not known for sufficiently large n. Still the condition G is unambiguous and we can use it to form the set of numbers

$$C =_{\mathrm{df}} \{n \in \mathbb{N} \mid (\exists m > n)[m, m + 2 \text{ are both prime numbers}]\}.$$

The *twin primes conjecture* asserts that $C = \mathbb{N}$, but if it is false, then C is some large, initial segment of the natural numbers.

In the same way, *an n-ary* **operation** *F is* **definite** *if it assigns to each n-tuple of objects $\vec{x}$ a unique, unambiguously determined object $w = F(\vec{x})$.* For example, assuming again that biology will not betray us, the operation

$$F(x) =_{\mathrm{df}} \begin{cases} \text{the father of } x, & \text{if } x \text{ is a human,} \\ x, & \text{otherwise,} \end{cases}$$

is definite. The silly consideration of cases here was put in to ensure that F determines a value for each argument x. In practice, we would define this operation by the simpler

$$F(x) =_{\mathrm{df}} \text{ the father of } x,$$

leaving it to the reader to supply some conventional, irrelevant value $F(x)$ for non-human x's. Again, definite operations need not be *effectively computable*, in fact the determination of the value $F(x)$ is sometimes the subject of judicial conflict in this specific case.

In addition to the General Comprehension Principle, we also assumed in the preceding chapter the existence of some specific sets, including the sets $\mathbb{N}$ and $\mathbb{R}$ of natural and real numbers, as well as the definiteness of some basic conditions from classical mathematics, e.g., the condition of "being a function",

$$\text{Function}(f, A, B) \iff f \text{ is a function from } A \text{ to } B.$$

This poses no problem as mathematicians have always made these assumptions, explicitly or implicitly.

The General Comprehension Principle has such strong intuitive appeal that the next theorem is called a "paradox".

3.5. Russell's paradox. *The General Comprehension Principle is not valid.*

PROOF. Notice first that if the General Comprehension Principle holds, then the set of all sets

$$V =_{\mathrm{df}} \{x \mid x \text{ is a set}\}$$

exists, and it has the peculiar property that it belongs to itself, $V \in V$. The common sets of everyday mathematics—sets of numbers, functions, etc.— surely do not belong to themselves, and so it is natural to consider them as members of a smaller, more natural universe of sets, by applying the General Comprehension Principle again,

$$R = \{x \mid x \text{ is a set and } x \notin x\}.$$

From the definition of R, however,

$$R \in R \iff R \notin R,$$

which is absurd. ⊣

When it is more than just a mistake, a "paradox" is simply a fact which runs counter to our intuitions, and set theorists already knew several such "paradoxes" before Russell announced this one in 1902, in a historic letter to the leading German philosopher and founder of mathematical logic Gottlob

Frege. These other paradoxes, however, were technical and affected only some of the most advanced parts of Cantor's theory. One could imagine that higher set theory had a systematic error built in, something like allowing a careless "division by 0" which would soon be discovered and disallowed, and then everything would be fixed. After all, contradictions and paradoxes had plagued the "infinitesimal calculus" of Newton and Leibnitz and they all went away after the rigorous foundation of the theory which was just being completed in the 1890s, without affecting the vital parts of the subject. Russell's paradox, however, was something else again: simple and brief, it affected directly the fundamental notion of set and the "obvious" principle of comprehension on which set theory had been built. It is not an exaggeration to say that Russell's paradox brought a foundational *crisis of doubt*, first to set theory and through it, later, to all of mathematics, which took over thirty years to overcome.

Some, like the French geometer Poincaré and the Dutch topologist and philosopher Brouwer, proposed radical solutions which essentially dismissed set theory (and much of classical mathematics along with it) as "pseudotheories", without objective content. From those who were reluctant to leave "Cantor's paradise", Russell first attempted to "rescue" set theory with his famous *theory of types*, which, however, is awkward to apply and was not accepted by a majority of mathematicians.[6] At approximately the same time, Zermelo proposed an alternative solution, which in time and with the contributions of many evolved into the contemporary theory of sets.

In his first publication on the subject in 1908, Zermelo took a pragmatic view of the problem. No doubt the General Comprehension Principle was not generally valid, Russell's paradox had made that clear. On the other hand, the specific applications of this principle in the proofs of basic facts about sets (like those in Chapter **2**) are few, simple and seemingly non-contradictory.

> Under such circumstances there is at this point nothing left for us to do but to proceed in the opposite direction [from that of the General Comprehension Principle] and, starting from set theory as it is historically given, to seek out the principles required for establishing the foundations of this mathematical discipline. In solving the problem we must, on the one hand, restrict these principles sufficiently to exclude all contradictions and, on the other, take them sufficiently wide to retain all that is valuable in this theory.

In other words, Zermelo proposed to replace the direct *intuitions* of Cantor about sets which led us to the faulty General Comprehension Principle with some *axioms*, hypotheses about sets which we accept with little a priori justification, simply because they are necessary for the proofs of the fundamental results of the existing theory and seemingly free of contradiction.

---

[6]The theory of types has had a strong influence in the development of analytic philosophy and logic, and some of its basic ideas eventually have also found their place in set theory.

Such were the philosophically dubious beginnings of **axiomatic set theory**, surely one of the most significant achievements of 20th century science. From its inception, however, the new theory had a substantial advantage in the genius of Zermelo, who selected an extraordinarily natural and pliable axiomatic system. None of Zermelo's axioms has yet been discarded or seriously revised and (until very recently) only one basic new axiom was added to his seven in the decade 1920-1930. In addition, despite the opportunistic tone of the cited quotation, each of Zermelo's axioms expresses a property of sets which is intuitively obvious and was already well understood from its uses in classical mathematics. With the experience gained from working out the consequences of these axioms over the years, a new intuitive notion of "grounded set" has been created which does not lead to contradictions and for which the axioms of set theory are clearly true. We will reconsider the problem of foundation of set theory after we gain experience by the study of its basic mathematical results.

The basic model for the axiomatization of set theory was Euclidean geometry, which for 2000 years had been considered the "perfect" example of a rigorous, mathematical theory. If nothing else, the axiomatic method clears the waters and makes it possible to separate what might be confusing and self-contradictory in our intuitions about the objects we are studying, from simple errors in logic we might be making in our proofs. As we proceed in our study of axiomatic set theory, it will be useful to remind ourselves occasionally of the example of Euclidean geometry.

3.6. The axiomatic setup. We assume at the outset that there is a **domain** or **universe** $\mathcal{W}$ **of objects**, some of which are **sets**, and certain **definite conditions and operations** on $\mathcal{W}$, among them the basic conditions of *identity*, *sethood* and *membership*:

$$x = y \iff x \text{ is the same object as } y,$$
$$\text{Set}(x) \iff x \text{ is a set},$$
$$x \in y \iff \text{Set}(y) \text{ and } x \text{ is a member of } y.$$

We call the objects in $\mathcal{W}$ which are not sets **atoms**, but we do not require that any atoms exist, i.e., we allow the possibility that all objects are sets. Definite conditions and operations are neither sets nor atoms.

This is the way every axiomatic theory begins. In Euclidean geometry for example, we start with the assumption that there are *points*, *lines* and several other geometrical objects and that some basic, definite conditions and operations are specified on them, e.g., it makes sense to ask if a "point P lies on the line L", or "to construct a line joining two given points". We then proceed to formulate the classical axioms of Euclid about these objects and to derive theorems from them. Actually Euclidean geometry is quite complex: there are several types of basic objects and a long list of intricate axioms about them. By contrast, Zermelo's set theory is quite austere: we just have sets and atoms and only seven fairly simple axioms relating them. In the remainder of

this chapter we will introduce six of these axioms with a few comments and examples. It is a bit easier to put off stating his last, seventh axiom until we first gain some understanding of the consequences of the first six in the next few chapters.

3.7. (**I**) **Axiom of Extensionality.** *For any two sets A, B,*

$$A = B \iff (\forall x)[x \in A \iff x \in B].$$

3.8. (**II**) **Emptyset and Pairset Axiom**. *(a) There is a special object $\emptyset$ which we will call a set, but which has no members. (b) For any two objects x, y, there is a set A whose only members are x and y, so that it satisfies the equivalence*

$$t \in A \iff t = x \lor t = y. \tag{3-2}$$

The Axiom of Extensionality implies that only one empty set exists, and that for any two objects x, y, only one set A can satisfy (3-2). We denote this **doubleton** of x and y by

$$\{x, y\} =_{\mathrm{df}} \text{ the unique set } A \text{ with sole members } x, y.$$

If $x = y$, then $\{x, x\} = \{x\}$ is the **singleton** of the object x.

Using this axiom we can construct many simple sets, e.g.,

$$\emptyset, \ \{\emptyset\}, \ \{\{\emptyset\}\}, \ \{\emptyset, \{\emptyset\}\}, \ \{\{\emptyset\}, \{\{\emptyset\}\}\}, \dots ,$$

but each of them has at most two members!

3.9. Exercise. *Prove that $\emptyset \neq \{\emptyset\}$.*

3.10. (**III**) **Separation Axiom** *or* **Axiom of Subsets**. *For each set A and each unary, definite condition P, there exists a set B which satisfies the equivalence*

$$x \in B \iff x \in A \,\&\, P(x). \tag{3-3}$$

From the Extensionality Axiom again, it follows that only one B can satisfy (3-3) and we will denote it by

$$B = \{x \in A \mid P(x)\}.$$

A characteristic contribution of Zermelo, this axiom is obviously a restriction of the General Comprehension Principle which implies many of its trouble-free consequences. For example, we can use it to define the operations of intersection and difference on sets,

$$A \cap B =_{\mathrm{df}} \{x \in A \mid x \in B\},$$
$$A \setminus B =_{\mathrm{df}} \{x \in A \mid x \notin B\}.$$

The proof of Russell's paradox yields a theorem:

3.11. Theorem. *For each set A, the set*

$$r(A) =_{\mathrm{df}} \{x \in A \mid x \notin x\} \tag{3-4}$$

is not a member of A. It follows that the collection of all sets is not a set, i.e., there is no set V such that

$$x \in V \iff \mathrm{Set}(x).$$

PROOF. Notice first that $r(A)$ is a set by the Separation Axiom. Assuming that $r(A) \in A$, we have (as before) the equivalence

$$r(A) \in r(A) \iff r(A) \notin r(A),$$

which is absurd.   ⊣

3.12. (**IV**) **Powerset Axiom.** *For each set A, there exists a set B whose members are the subsets of A, i.e.,*

$$X \in B \iff \mathrm{Set}(X) \,\&\, X \subseteq A. \tag{3-5}$$

Here $X \subseteq A$ is an abbreviation of $(\forall t)[t \in X \implies t \in A]$. The Axiom of Extensionality implies that for each A, only one set B can satisfy (3-5); we call it the **powerset** of A and we denote it by

$$\mathcal{P}(A) =_{\mathrm{df}} \{X \mid \mathrm{Set}(X) \,\&\, X \subseteq A\}.$$

3.13. Exercise. $\mathcal{P}(\emptyset) = \{\emptyset\}$ *and* $\mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$.

3.14. Exercise. *For each set A, there exists a set B whose members are exactly all singletons of members of A, i.e.,*

$$x \in B \iff (\exists t \in A)[x = \{t\}].$$

3.15. (**V**) **Unionset Axiom.** *For each set $\mathcal{E}$, there exists a set B whose members are the members of the members of $\mathcal{E}$, i.e., so that it satisfies the equivalence*

$$t \in B \iff (\exists X \in \mathcal{E})[t \in X]. \tag{3-6}$$

The Axiom of Extensionality implies again that for each $\mathcal{E}$, only one set can satisfy (3-6); we call it the **unionset** of $\mathcal{E}$ and we denote it by

$$\bigcup \mathcal{E} =_{\mathrm{df}} \{t \mid (\exists X \in \mathcal{E})[t \in X]\}.$$

The unionset operation is obviously most useful when $\mathcal{E}$ is a **family of sets**, i.e., a set all of whose members are also sets. This is the case for the simplest application, which (finally) gives us the binary, union operation on sets: we set

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

using axioms (**II**) and (**V**), and we compute

$$t \in A \cup B \iff (\exists X \in \{A, B\})[t \in X]$$
$$\iff t \in A \vee t \in B.$$

3.16. Exercise. $\bigcup \emptyset = \bigcup \{\emptyset\} = \emptyset$.

3.17. (**VI**) **Axiom of Infinity.** *There exists a set I which contains the empty set $\emptyset$ and the singleton of each of its members, i.e.,*

$$\emptyset \in I \,\&\, (\forall x)[x \in I \implies \{x\} \in I].$$

We have not given yet a rigorous definition of "infinite", but it is quite obvious that any I with the properties in the axiom must be infinite, since (**VI**) implies

$$\emptyset \in I, \; \{\emptyset\} \in I, \; \{\{\emptyset\}\} \in I, \dots$$

and the objects $\emptyset$, $\{\emptyset\}$, ... are all distinct sets by the Extensionality Axiom. The intuitive understanding of the axiom is that it demands precisely the existence of the set

$$I = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \dots\},$$

but it is simpler (and sufficient) to assume of I only the stated properties, which imply that it contains all these complex singletons.

It was a commonplace belief among philosophers and mathematicians of the 19th century that the existence of infinite sets could be proved, and in particular the set of natural numbers could be "constructed" out of thin air, "by logic alone". All the proposed "proofs" involved the faulty General Comprehension Principle in some form or another. We know better now: logic can codify the valid forms of reasoning but it cannot prove the existence of anything, let alone infinite sets. By taking account of this fact cleanly and explicitly in the formulation of his axioms, Zermelo made a substantial contribution to the process of purging logic of ontological concerns and separating the mathematical development of the theory of sets from logic, to the benefit of both disciplines.

3.18. Axioms for definite conditions and operations. Zermelo understood *definite conditions* intuitively, he described them much as we did in **3.4** and he applied the Separation Axiom using various quite complex conditions without any special argument that they are, indeed, "definite". We will do the same, because the business of proving definiteness is boring and not particularly illuminating. For the sake of completeness, however, we list here the only properties of definiteness that we will actually use.

1. The following basic conditions are definite:

$$x = y \iff_{df} x \text{ and } y \text{ are the same object,}$$
$$\mathrm{Set}(x) \iff_{df} x \text{ is a set,}$$
$$x \in y \iff_{df} \mathrm{Set}(y) \text{ and } x \text{ is a member of } y,$$

2. For each object c and each n, the constant n-ary operation

$$F(x_1, \dots, x_n) = c$$

is definite.
3. Each *projection operation*

$$F_i(x_1, \dots, x_n) = x_i \quad (1 \le i \le n)$$

is definite.
4. If P is a definite condition of $n + 1$ arguments and for each tuple of objects $\vec{x} = x_1, \dots, x_n$ there exists exactly one w such that $P(\vec{x}, w)$ is true, then the operation

$$F(\vec{x}) = \text{the unique } w \text{ such that } P(\vec{x}, w)$$

is definite.

5. If Q is an m-ary definite condition, each F_i is an n-ary definite operation for $i = 1, \ldots, m$ and

$$P(\vec{x}) \iff_{\text{df}} Q(F_1(\vec{x}), \ldots, F_m(\vec{x})),$$

then the condition P is also definite.

6. If Q, R and S are definite conditions of the appropriate number of arguments, then so are the following conditions which are obtained from them by applying the elementary operations of logic:

$$P_1(\vec{x}) \iff_{\text{df}} \neg P(\vec{x}) \iff P(\vec{x}) \text{ is false,}$$
$$P_2(\vec{x}) \iff_{\text{df}} Q(\vec{x}) \,\&\, R(\vec{x}) \iff \text{both } Q(\vec{x}) \text{ and } R(\vec{x}) \text{ are true,}$$
$$P_3(\vec{x}) \iff_{\text{df}} Q(\vec{x}) \vee R(\vec{x}) \iff \text{either } Q(\vec{x}) \text{ or } R(\vec{x}) \text{ is true,}$$
$$P_4(\vec{x}) \iff_{\text{df}} (\exists y)S(\vec{x}, y) \iff \text{for some } y, S(\vec{x}, y) \text{ is true,}$$
$$P_5(\vec{x}) \iff_{\text{df}} (\forall y)S(\vec{x}, y) \iff \text{for every } y, S(\vec{x}, y) \text{ is true.}$$

All the conditions and operations we will use can be proved definite by appealing to these basic properties. Aside from one problem at the end of this chapter, however, for the logically minded, we will omit these technical proofs of definiteness and it is best for the reader to forget about them too: they detract from the business at hand, which is the study of sets, not definite conditions and operations.

3.19. Classes. Having gone to all the trouble to discredit the General Principle of Comprehension, we will now profess that *for every unary, definite condition P there exists a* **class**

$$A = \{x \mid P(x)\}, \tag{3-7}$$

such that for every object x,

$$x \in A \iff P(x). \tag{3-8}$$

To give meaning to this principle and prove it, we need to define classes. Every set will be a class, but because of the Russell Paradox **3.5**, there must be classes which are not sets, else (3-8) leads immediately to the Russell Paradox in the case $P(x) \iff \text{Set}(x) \,\&\, x \notin x$.

First let us agree that for every unary, definite condition P we will write synonymously

$$x \in P \iff P(x).$$

For example, if Set is the basic condition of sethood, we write interchangeably

$$x \in \text{Set} \iff \text{Set}(x) \iff x \text{ is a set.}$$

This is just a convenient notational convention.

A unary definite condition P is **coextensive** with a set A if the objects which satisfy it are precisely the members of A, in symbols

$$P =_{\text{e}} A \iff_{\text{df}} (\forall x)[P(x) \iff x \in A]. \tag{3-9}$$

For example, if

$$P(x) \iff x \neq x,$$

then $P =_e \emptyset$. By the Russell Paradox **3.5**, not every P is coextensive with a set. On the other hand, *a unary, definite condition P is coextensive with at most one set*; because if $P =_e A$ and also $P =_e B$, then for every x,

$$x \in A \iff P(x) \iff x \in B,$$

and $A = B$ by the Axiom of Extensionality.

By definition, a **class** *is either a set or a unary definite condition which is not coextensive with a set*. With each unary condition P, we associate the class

$$\{x \mid P(x)\} =_{df} \begin{cases} \text{the unique set } A \text{ such that } P =_e A, \\ \qquad\qquad\qquad\qquad\qquad \text{if } P =_e A \text{ for some set } A, \\ P, \quad \text{otherwise.} \end{cases} \qquad (3\text{-}10)$$

Now if $A =_{df} \{x \mid P(x)\}$, then either P is coextensive with a set, in which case $P =_e A$ and by the definition $x \in A \iff P(x)$; or P is not coextensive with any set, in which case $A = P$ and

$$x \in A \iff x \in P \iff P(x) \quad \text{(by the notational convention).}$$

This is exactly the *General Comprehension Principle for Classes* enunciated in (3-7) and (3-8).

3.20. Exercise. *For every set A,*

$$\{x \mid x \in A\} = A,$$

and, in particular, every set is a class. Show also that

$$\{X \mid \text{Set}(X) \,\&\, X \subseteq A\} = \mathcal{P}(A).$$

3.21. Exercise. *The class $\mathcal{W}$ of all objects is not a set, and neither is the class* Set *of all sets.*

If P is an n-ary definite condition and F an n-ary definite operation, we set

$$\{F(\vec{x}) \mid P(\vec{x})\} =_{df} \{w \mid (\exists \vec{x})[P(\vec{x}) \,\&\, w = F(\vec{x})]\}. \qquad (3\text{-}11)$$

For example, with $F(x) = \{x\}$,

$$\{\{x\} \mid x = x\} = \{w \mid (\exists x)[w = \{x\}]\} = \text{the class of all singletons.}$$

3.22. Exercise. *The class $\{\{x\} \mid x = x\}$ of all singletons is not a set.*

3.23. Exercise. *For every class A,*

$$A \text{ is a set} \iff \text{for some class } B, \ A \in B$$
$$\iff \text{for some set } X, \ A \subseteq X,$$

where inclusion among classes is defined as if they were sets,

$$A \subseteq B \iff_{df} (\forall x)[x \in A \implies x \in B].$$

3.24. The Axioms of Choice and Replacement: a warning. Our axiomatization of set theory will not be complete until we introduce Zermelo's last Axiom of Choice in Chapter **8** and the later Axiom of Replacement in Chapter **11**. While there are good reasons for these postponements which we will explain in due course, there are also good reasons for adding the axioms of Choice

and Replacement: many basic set theoretic arguments need them, and among these are some of the simplest claims of Chapter **2**. Thus, until Chapter **8**, we will need to be extra careful and make sure that our constructions indeed can be justified by axioms (**I**) – (**VI**) and that we have not sneaked in some "obvious" assertion about sets not yet proved or assumed. In a few places we will formulate and prove something weaker than the whole truth, when the proof of the whole truth needs one of the missing axioms. Now this is good: it will keep us on our toes and make us understand better the art of reasoning from axioms.

3.25. About atoms. Most recent developments of axiomatic set theory assume at the outset the so-called **Principle of Purity**, that *there are no atoms*: all objects of the basic domain are *sets*. There is a certain appealing simplicity to this conception of a mathematical world in which everything is a set. We have followed Zermelo in allowing atoms (without demanding them), primarily because this makes the theory more naturally relevant to the natural sciences: we want our results to apply to sets of *planets*, *molecules* or *frogs*, and frogs are not sets. In any case, it comes at little cost, we simply have to say "object" in some situations where the atom banners would say "set". It is important to notice, however, that none of the axioms requires the existence of atoms, so none of the consequences we will derive from them depends on the existence of atoms: everything we will prove remains true in the domain of pure sets, provided only that it satisfies the Zermelo axioms, as we stated them.

3.26. Axioms as closure properties of the universe $\mathcal{W}$. Whatever the domain $\mathcal{W}$ of our axiomatic set theory may be, it is clear that it does not contain all "objects of our intuition or thought" in Cantor's expression; $\mathcal{W}$ is not a set, and it is certainly a perfectly legitimate mathematical object of our intuition about which we intend to have many thoughts. Granting that $\mathcal{W}$ is not all there is, we can fruitfully conceive of the axioms as imposing *closure conditions* on it. We have assumed (so far) that $\mathcal{W}$ contains $\emptyset$, that it is closed under the operations of pairing $\{x, y\}$ (**II**), powerset $\mathcal{P}(X)$ (**IV**), and unionset $\bigcup \mathscr{E}$ (**V**), that it includes every definite subcollection of every set (**III**), and that it contains some set I with the stipulated property of the Axiom of Infinity (**VI**). It is also possible to understand the Axiom of Extensionality (**I**) as a closure property of $\mathcal{W}$: in its non-trivial direction, it says that for any two sets A, B,

$$A \neq B \Longrightarrow (\exists t)[t \in (A \setminus B) \cup (B \setminus A)], \tag{3-12}$$

i.e., every inequality $A \neq B$ between two sets is witnessed by some legitimate object $t \in \mathcal{W}$ which belongs to one and not to the other.

This understanding of the meaning of the axioms is compatible with two different conceptions of the universe $\mathcal{W}$. One is that it is huge, amorphous, difficult to understand and impossible to define; but every object in it is concrete, definite, whole, and this is enough to justify the closure properties of $\mathcal{W}$ embodied by the axioms. Let us call this *the large view*. The *small view* is that $\mathcal{W}$ consists precisely of those objects whose existence is "guaranteed"

by the axioms, those which can be "constructed" by applying the axioms repeatedly: the axioms are satisfied because we deliberately put in $\mathcal{W}$ all the objects required by the closure properties they express. Neither conception is precise, to be sure, but they are different. On the small view, for example, there are no atoms since none of the axioms demands their existence, while the large view allows frogs among the objects in $\mathcal{W}$.

Both of these views can be defended and they have played significant roles in the philosophy of set theory, and even in its mathematical practice, by suggesting the kind of questions one should ask. We will come back to discuss the issue in Chapter **11** and Appendix **B**, when we will be in a position to be less flippant about it. In the meantime, we will often speak of the axioms as closure conditions on $\mathcal{W}$, a useful heuristic device which is compatible with every philosophical approach to the subject.

Problems for Chapter 3

x3.1. For each non-empty set $\mathcal{E}$ and each $X \in \mathcal{E}$, we define the *intersection of $\mathcal{E}$ via X* by

$$\bigcap_X \mathcal{E} =_{\mathrm{df}} \{x \in X \mid (\forall U \in \mathcal{E})[x \in U]\}.$$

Show that for any two members X, Y of $\mathcal{E}$,

$$\bigcap_X \mathcal{E} = \bigcap_Y \mathcal{E},$$

i.e., the intersection $\bigcap_X \mathcal{E}$ is independent of the specific X we used in its definition, and hence we can use for it the notation $\bigcap \mathcal{E}$ which does not exhibit X. Show also that $A \cap B = \bigcap\{A, B\}$.

x3.2. For any two sets A, B, determine whether each of the following classes is or is not a set.

1. $\{\{\emptyset, x\} \mid x \in A\}$.
2. $\{x \mid \mathrm{Set}(x) \,\&\, x \neq \emptyset\}$.
3. $\{\{x, y\} \mid x \in A \,\&\, y \in B\}$.
4. $\{\mathcal{P}(X) \mid X \subseteq A\}$.

x3.3. Prove rigorously that the following conditions and operations are definite, using only (**I**) – (**VI**) and the axioms in **3.18**. (Here c is some arbitrary object.)

$$
\begin{aligned}
P_1(x) &\Longleftrightarrow_{\mathrm{df}} x \in c, & F_1(x, y) &=_{\mathrm{df}} \{x, y\}, \\
P_2(x, y, z) &\Longleftrightarrow_{\mathrm{df}} z \in x, & F_2(X) &=_{\mathrm{df}} \bigcup X, \\
P_3(X, Y) &\Longleftrightarrow_{\mathrm{df}} X \subseteq Y, & F_3(X) &=_{\mathrm{df}} \mathcal{P}(X), \\
P_4(X, Y) &\Longleftrightarrow_{\mathrm{df}} X \cap Y = \emptyset, & F_4(x) &=_{\mathrm{df}} \{x\}, \\
P_5(X, Y) &\Longleftrightarrow_{\mathrm{df}} \mathcal{P}(X) \subseteq Y. & F_5(X, Y) &=_{\mathrm{df}} X \cup Y.
\end{aligned}
$$

In the remaining problems of this section we consider Zermelo's notion of *equivalence*, which intuitively holds between two sets when they are equinumerous. These problems will be trivial after we introduce *functions* within
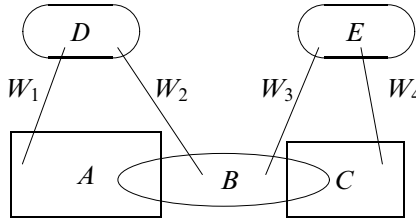
FIGURE **3.1**. The hypothesis for Part 3 of Problem **x3.5**.

Zermelo's axiomatic theory in the next Chapter, but right now they are challenging.

3.27. Definition. Recall that two sets A, B are **disjoint** if $A \cap B = \emptyset$. A set W is a **connection** of the two disjoint sets A and B (according to Zermelo) if the following three conditions hold:

1. $Z \in W \implies (\exists x \in A, y \in B)[Z = \{x, y\}]$.
2. For each $x \in A$, there is exactly one $y \in B$ such that $\{x, y\} \in W$.
3. For each $y \in B$, there is exactly one $x \in A$ such that $\{x, y\} \in W$.

x3.4. For any two disjoint sets A, B, the class

$$\Sigma(A, B) = \{W \mid W \text{ is a connection of } A \text{ with } B\}$$

is a set.

3.28. Definition. Two sets A, B are **equivalent according to Zermelo** if there exists a third set C disjoint from both of them and connections of A with C and of B with C, in symbols

$$A \sim_Z B \iff (\exists C, W, W')[A \cap C = \emptyset \,\&\, B \cap C = \emptyset$$
$$\&\, W \in \Sigma(A, C) \,\&\, W' \in \Sigma(B, C)].$$

****x3.5.** The condition of equivalence according to Zermelo has the following properties, for any three sets A, B, C:

$$A \sim_Z A,$$
$$\text{if } A \sim_Z B, \text{ then } B \sim_Z A,$$
$$\text{if } (A \sim_Z B \,\&\, B \sim_Z C), \text{ then } A \sim_Z C.$$

HINT. To prove $A \sim_Z A$, you need to find some set C such that $A \cap C = \emptyset$ and there is a connection W of A with C. The hypothesis for the (last) transitivity property is illustrated in Figure **3.1**.

CHAPTER 4

ARE SETS ALL THERE IS?

Our next goal is to determine whether the basic results of naive set theory in Chapter **2** can be proved on the basis of the axioms of Zermelo. Right at the start we hit a snag: to define the crucial notion of *equinumerosity* we need functions; to define *countable sets* we need the specific set $\mathbb{N}$ of natural numbers; the fundamental theorem **2.21** of Cantor is about the set $\mathbb{R}$ of real numbers, etc. Put another way, the results of Chapter **2** are not only about sets, but about points, numbers, functions, Cartesian products and many other mathematical objects which are plainly not sets. Where will we find these objects in the axioms of Zermelo which speak only about sets?

An obvious solution is to assume that these non-sets are among the *atoms* which are allowed by Zermelo's theory and to add axioms which express our basic intuitions about points, numbers, functions, etc. This is possible but awkward and there is a much better solution.

A typical example of the method we will adopt is the "identification" of the (directed) geometric line Π with the set $\mathbb{R}$ of real numbers, via the correspondence which "identifies" each point P on the line with its coordinate $x(P)$ with respect to a fixed choice of an origin O. What is the precise meaning of this "identification"? *Certainly not that points are real numbers.* Men have always had direct geometric intuitions about points which have nothing to do with their coordinates and which existed before Descartes discovered analytic geometry. Every Athenian of the classical period understood the meaning of the sentence

> Phaliron is between Piraeus and Sounion along the Saronic coast[7]

even though he was (by necessity) ignorant of analytic geometry. In fact, many educated ancient Athenians had an excellent understanding of the Pythagorean Theorem, without knowing how to coordinetize the plane. What we mean by the "identification" of Π with $\mathbb{R}$ is that the correspondence $P \mapsto x(P)$ gives a **faithful representation** of Π in $\mathbb{R}$ which allows us to give arithmetic definitions for all the useful geometric notions and to study the mathematical properties of Π **as if points were real numbers**. For example, the

---

[7]These are seaside suburbs of Athens.

quoted sentence above is expressed by the inequalities

$$x(\text{Piraeus}) < x(\text{Phaliron}) < x(\text{Sounion}),$$

assuming that the coordinates increase in the easterly direction. In the same way, we will discover within the universe of sets *faithful representations* of all the mathematical objects we need, and we will study set theory on the basis of the lean axiomatic system of Zermelo **as if all mathematical objects were sets**. The delicate problem in specific cases is to formulate precisely the correct definition of a "faithful representation" and to prove that one such exists.

4.1. Ordered pair. We consider first the basic (**ordered**) **pair** operation. Intuitively, the pair (x, y) of two objects x and y is the "thing" which has a "first member" x and a "second member" y, and it is different from the unordered pair $\{x, y\}$ since (for example) $\{0, 1\} = \{1, 0\}$ while $(0, 1) \neq (1, 0)$. Thus, the first characteristic property of the ordered pair is the following:

(OP1) $(x, y) = (x', y') \iff x = x' \,\&\, y = y'.$

There is a second, perhaps less obvious characteristic property of pairs which makes it possible to define Cartesian products: for any two sets A, B,

(OP2)   the class $A \times B =_{\text{df}} \{(x, y) \mid x \in A \,\&\, y \in B\}$ is a set.

Thus, the problem of representing the notion of "pair" in set theory takes the following precise form: we must define a definite operation (x, y) such that (OP1) and (OP2) follow from the axioms of Zermelo.

4.2. Lemma. *The **Kuratowski pair** operation*

$$(x, y) =_{\text{df}} \{\{x\}, \{x, y\}\} \tag{4-1}$$

has properties (OP1) *and* (OP2).

Proof. (OP1). The direction $\Longleftarrow$ is obvious. For the non-trivial direction $\Longrightarrow$, we distinguish two cases.

If $x = y$, then $\{x, y\} = \{x, x\} = \{x\}$, the set $(x, y) = \{\{x\}, \{x\}\} = \{\{x\}\}$ is a singleton, hence the set (x', y') which is assumed equal to it is also a singleton, so that $x' = y'$ and $(x', y') = \{\{x'\}\}$; and since this last singleton is equal to $\{\{x\}\}$, we have $x = x'$ and, hence, also $y = x = x' = y'$.

If $x \neq y$, then the members of (x, y) are the singleton $\{x\}$ and the doubleton $\{x, y\}$, and these must correspond with the members $\{x'\}$ and $\{x', y'\}$ of the equal set (x', y'), so that we must have $\{x\} = \{x'\}$, $\{x, y\} = \{x', y'\}$, and then, immediately, $x = x'$ and $y = y'$.

(OP2). It is enough to prove that for any two sets A, B, there is a set C such that

$$x \in A \,\&\, y \in B \Longrightarrow \{\{x\}, \{x, y\}\} \in C; \tag{4-2}$$

because then

$$A \times B = \{z \in C \mid (\exists x \in A)(\exists y \in B)[z = (x, y)]\},$$

and $A \times B$ is a set by the Subset Axiom (**III**). To prove (4-2), we compute:

$$
\begin{aligned}
x \in A, \ y \in B &\implies \{x\}, \ \{x, y\} \subseteq (A \cup B) \\
&\implies \{x\}, \ \{x, y\} \in \mathcal{P}(A \cup B) \\
&\implies \{\{x\}, \ \{x, y\}\} \subseteq \mathcal{P}(A \cup B) \\
&\implies (x, y) \in \mathcal{P}(\mathcal{P}(A \cup B)),
\end{aligned}
$$

so that we can take $C = \mathcal{P}(\mathcal{P}(A \cup B))$. ⊣

We now fix a specific definite operation (x, y) which satisfies (OP1) and (OP2), perhaps the Kuratowski pair defined in the proof of **4.2**, perhaps some other: from now on we may forget the specific definition chosen, the only thing that counts is that the pair operation satisfies (OP1) and (OP2).

4.3. Exercise. *Let*

$$
\mathrm{Pair}(z) \iff_{\mathrm{df}} (\exists x)(\exists y)[z = (x, y)],
$$

$$
\mathrm{First}(z) =_{\mathrm{df}} \begin{cases} \text{the unique } x \text{ such that } (\exists y)[z = (x, y)], \text{ if } \mathrm{Pair}(z), \\ z, \text{ otherwise}, \end{cases}
$$

$$
\mathrm{Second}(z) =_{\mathrm{df}} \begin{cases} \text{the unique } y \text{ such that } (\exists x)[z = (x, y)], \text{ if } \mathrm{Pair}(z), \\ z, \text{ otherwise}. \end{cases}
$$

It follows that

$$
\mathrm{Pair}(z) \iff z = (\mathrm{First}(z), \mathrm{Second}(z)).
$$

Using the ordered pair we can easily define triples, quadruples, etc., as well as the corresponding products, e.g.,

$$
(x, y, z) =_{\mathrm{df}} (x, (y, z)), \tag{4-3}
$$
$$
(x, y, z, w) =_{\mathrm{df}} (x, (y, z, w)) = (x, (y, (z, w))), \tag{4-4}
$$
$$
A \times B \times C =_{\mathrm{df}} A \times (B \times C), \tag{4-5}
$$

etc. By this definition, a tuple of length $n + 1$ is a pair with second member a tuple of length n.

4.4. Exercise. *For all x, y, z, x', y', z',*

$$
(x, y, z) = (x', y', z') \iff x = x' \ \& \ y = y' \ \& \ z = z'.
$$

4.5. Disjoint union. For each set A and a fixed object <u>blue</u>, we can think of the set of pairs $\{\underline{\mathrm{blue}}\} \times A$ as a "blue copy" of A, the act of replacing each $a \in A$ by the pair $(\underline{\mathrm{blue}}, a)$ being the set theoretic equivalent of painting a blue. We fix two such distinct "colors",

$$
\underline{\mathrm{blue}} =_{\mathrm{df}} \emptyset, \quad \underline{\mathrm{white}} =_{\mathrm{df}} \{\emptyset\}, \tag{4-6}
$$

and we define the **disjoint union** of two sets by

$$
A \uplus B =_{\mathrm{df}} (\{\underline{\mathrm{blue}}\} \times A) \cup (\{\underline{\mathrm{white}}\} \times B).
$$

The construction is pictured in Figure **4.1**, and the notion is useful, as we will see. It should be clear that the specific identity of <u>blue</u> and <u>white</u> must be deliberately and instantly forgotten: all that matters is that <u>blue</u> $\neq$ <u>white</u>.
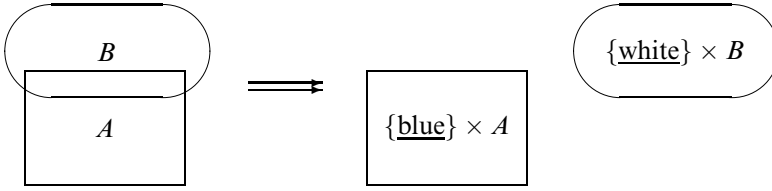
FIGURE **4.1**. Constructing the disjoint union.

4.6. Exercise. *Show that for all sets A, B, $A \uplus \emptyset \subseteq A \uplus B$. Assuming that we use the Kuratowski pair to define products, give an example where the plausible inclusion $A \subseteq A \uplus B$ is not true.*

Next we consider the notion of **relation** which permeates mathematics. Intuitively, a *binary relation R* between objects $x \in A$ and $y \in B$ is a condition which is satisfied by some $x \in A$, $y \in B$ and fails for others. For example, the relation

$$x \, R \, y \iff_{\text{df}} x \text{ is a son of } y$$

is defined on $A = \{\text{men}\}$, $B = \{\text{women}\}$ and holds for x, y precisely if y has given birth to x. The obvious way to represent a binary relation in set theory is to identify it with its *extension*, the set of pairs which satisfy it.

4.7. Relations. A **binary relation** on the sets A, B is any subset R of the Cartesian product $A \times B$. We will use synonymously the notations

$$R(x, y) \iff x \, R \, y \iff (x, y) \in R.$$

Obvious examples of binary relations are the **identity** and the relations of **membership** and **subsethood** restricted to some set A,

$$x =_A y \iff_{\text{df}} x \in A \,\&\, y \in A \,\&\, x = y,$$
$$x \in_A Y \iff_{\text{df}} x \in A \,\&\, Y \subseteq A \,\&\, x \in Y,$$
$$X \subseteq_A Y \iff_{\text{df}} X \subseteq Y \subseteq A,$$

which by the definition are represented respectively by the sets

$$=_A \,=_{\text{df}} \{(x, y) \in A \times A \mid x = y\},$$
$$\in_A \,=_{\text{df}} \{(x, Y) \in A \times \mathcal{P}(A) \mid x \in Y\},$$
$$\subseteq_A \,=_{\text{df}} \{(X, Y) \in \mathcal{P}(A) \times \mathcal{P}(A) \mid X \subseteq Y\}.$$

4.8. Relations and definite conditions. The definite conditions of identity $=$, membership $\in$ and subsethood $\subseteq$ on the domain $\mathcal{W}$ of all objects are not relations according to **4.7**, in fact they are not even "coextensive" with sets of pairs, because their extensions are "too large". This is important, the distinction between relations and definite conditions: briefly, *every relation determines a definite condition but* (in general) *the converse does not hold*. The precise situation is detailed in the next Exercise. In practice we will often refer to *the relation $=$ on the set A*, meaning (without possibility of confusion) the restriction $=_A$ as we just defined it.

4.9. Exercise. (1) *For every binary relation $R \subseteq (A \times B)$, the condition*

$$R^*(x, y) \iff_{df} (x, y) \in R$$

is definite. (Nothing to prove here, unless you want to practice applying **3.18**.)
   (2) *For every binary definite condition P and any two sets A, B, the* **restriction**

$$P_{A,B} =_{df} \{(x, y) \in A \times B \mid P(x, y)\}$$

of P to $A \times B$ is a binary relation on A, B.

4.10. Properties of relations on a set A. Binary relations with both arguments ranging over the same set are especially important, and they are classified and studied according to the structural properties they may enjoy. Here are three such properties which come up often, in various combinations, where $P \subseteq A \times A$:

$$P \text{ is } \textbf{reflexive} \iff_{df} (\forall x \in A)[x \, P \, x],$$
$$P \text{ is } \textbf{symmetric} \iff_{df} (\forall x, y \in A)[x \, P \, y \implies y \, P \, x],$$
$$P \text{ is } \textbf{transitive} \iff_{df} (\forall x, y, z \in A)[[x \, P \, y \, \& \, y \, P \, z] \implies x \, P \, z].$$

We call P an **equivalence relation** on A if it has all three of these properties.

Equivalence relations are very useful and we will meet examples of them in practically every chapter of these Notes. They are often denoted by symbols like $\sim, \approx, \simeq$, so that their three characteristic properties take the form

$$x \sim x,$$
$$x \sim y \implies y \sim x,$$
$$[x \sim y \, \& \, y \sim z] \implies x \sim z.$$

4.11. Exercise. *On each set A, the identity relation $\{(x, y) \mid x = y \in A\}$, the identically true relation $\{(x, y) \mid x, y \in A\}$ and for each $B \subseteq A$ the relation*

$$x \sim_{A/B} y \iff_{df} x = y \in A \vee [x \in B \, \& \, y \in B]$$

are all equivalence relations.

4.12. Proposition. *Suppose $\sim$ is an equivalence relation on the set A, let for each $x \in A$,*

$$[x/\sim] = \{y \in A \mid x \sim y\} \tag{4-7}$$

be the **equivalence class**[8] *of x, and let*

$$\llbracket A/\sim \rrbracket = \{[x/\sim] \in \mathcal{P}(A) \mid x \in A\} \tag{4-8}$$

be the set of all equivalence classes. It follows that $[x/\sim] \neq \emptyset$ for each $x \in A$, and for all $x, y \in A$,

$$x \sim y \iff [x/\sim] = [y/\sim], \tag{4-9}$$
$$x \nsim y \iff [x/\sim] \cap [y/\sim] = \emptyset, \tag{4-10}$$

---

[8] Each $[x/\sim]$ is obviously a set, a subset of A, and it would be more appropriately named the "equivalence set" of x, but the classical terminology goes way back and has been frozen.

so that $[\![A/\!\sim]\!]$ is a family of non-empty and pairwise disjoint subsets of A such that $\bigcup [\![A/\!\sim]\!] = A$.

Conversely, for each family $\mathscr{E}$ of non-empty and pairwise disjoint subsets of A such that $A = \bigcup \mathscr{E}$, the relation

$$x \sim y \iff_{\mathrm{df}} (\exists X \in \mathscr{E})[x \in X \,\&\, y \in X]$$

is an equivalence relation on A and $[\![A/\!\sim]\!] = \mathscr{E}$.

PROOF. Each $[x/\!\sim] \neq \emptyset$, since $x \in [x/\!\sim]$. By the transitivity and symmetry of $\sim$,

$$t \sim x \,\&\, x \sim y \Longrightarrow t \sim y, \quad t \sim y \,\&\, x \sim y \Longrightarrow t \sim x$$

so that

$$x \sim y \implies (\forall t \in A)[t \sim x \iff t \sim y]$$
$$\implies [x/\!\sim] = [y/\!\sim].$$

This implies immediately both (4-9) and (4-10). For the converse, the reflexivity and symmetry of $\sim$ are trivial. If $x \sim y$ and $y \sim z$, then there exist sets X, Y in $\mathscr{E}$ such that $x, y \in X$, $y, z \in Y$, so in particular $y \in X \cap Y$ and since the sets in $\mathscr{E}$ are pairwise disjoint, we have $X = Y$, so $x \sim z$. $\dashv$

4.13. Exercise. *What are the equivalence classes of the equivalence relations in Exercise* **4.11**?

Following up the same idea, we identify each **ternary relation** R on the sets A, B, C with the set of triples which satisfy it, so that a ternary relation on A, B, C is simply an arbitrary subset of $A \times B \times C$. We will use synonymously the notations

$$R(x, y, z) \iff_{\mathrm{df}} (x, y, z) \in R.$$

As with relations, we represent functions in set theory by identifying them with their "graphs".

4.14. Functions. A **function** (or **mapping** or **transformation**) $f : A \to B$ with domain the set A and range the set B is any subset $f \subseteq (A \times B)$ which satisfies the condition

$$(\forall x \in A)(\exists ! y \in B)[(x, y) \in f],$$

in more detail,

$$(\forall x \in A)(\exists y \in B)[(x, y) \in f],$$
$$\text{and } (x, y) \in f \,\&\, (x, y') \in f \Longrightarrow y = y'.$$

If we picture the product $A \times B$ in the plane as in Figure **4.2**, this means that every "vertical line" intersects the set f in exactly one point: for $x \in A$ and $f : A \to B$, we will write, as usual,

$$f(x) =_{\mathrm{df}} \text{ the unique } y \in B \text{ such that } (x, y) \in f \qquad (x \in A).$$
$$= \text{ the value of } f \text{ on the element } x$$

The function space
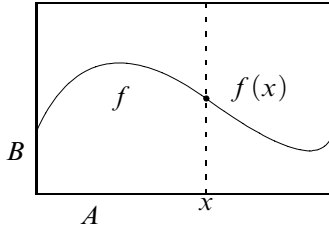
FIGURE **4.2**. A function as a set of ordered pairs.

$$(A \to B) =_{\text{df}} \{f \subseteq A \times B \mid f : A \to B\}$$
$$= \{f \in \mathcal{P}(A \times B) \mid (\forall x \in A)(\exists! y \in B)(x, y) \in f\}$$

is a set by the Subset Axiom (**III**).

We will use all the familiar notations and abbreviations in connection with functions, e.g., sometimes writing the argument without the parentheses or as an index,

$$f(x) = f x = f_x.$$

The $\mapsto$ notation is also useful; for example, an **indexed family of sets** is a function

$$A = (i \mapsto A_i)_{i \in I} : I \to E$$

for some $I \neq \emptyset$ and some E, where each A_i is a set. We refer to I as the **index set** and we define the union and intersection of the family in the usual way,

$$\bigcup_{i \in I} A_i =_{\text{df}} \{x \in \bigcup E \mid (\exists i \in I)[x \in A_i]\},$$
$$\bigcap_{i \in I} A_i =_{\text{df}} \{x \in \bigcup E \mid (\forall i \in I)[x \in A_i]\}. \tag{4-11}$$

We can also define the **product** of an indexed family, the set of functions which select for each $i \in I$ one element from the value A_i,

$$\prod_{i \in I} A_i =_{\text{df}} \{f : I \to \bigcup_{i \in I} A_i \mid (\forall i \in I)[f(i) \in A_i]\}. \tag{4-12}$$

Injections, surjections, bijections (correspondences), images and pre-images of functions are defined exactly as in the Introduction. We will be using the following notations for the relevant function spaces:

$$(A \rightarrowtail B) =_{\text{df}} \{f \in (A \to B) \mid f \text{ is an injection, one-to-one}\},$$
$$(A \twoheadrightarrow B) =_{\text{df}} \{f \in (A \to B) \mid f \text{ is a surjection, onto } B\},$$
$$(A \rightarrowtail\!\!\!\rightarrow B) =_{\text{df}} (A \rightarrowtail B) \cap (A \twoheadrightarrow B) \text{ (bijection, correspondence)}.$$

For each $X \subseteq A$, the **restriction** $f \upharpoonright X$ of a function $f : A \to B$ is obtained by cutting f down so it is defined only on X,

$$f \upharpoonright X =_{\text{df}} \{(x, y) \in f \mid x \in X\}. \tag{4-13}$$

It is also useful to notice that the basic condition of "functionhood"

$$\text{Function}(f) \iff_{\text{df}} (\exists A)(\exists B)[f \in (A \to B)] \tag{4-14}$$

is evidently definite. When we refer to a *function f* without identifying specific sets A, B such that $f : A \to B$, we will mean any set f which satisfies the condition Function(f).

4.15. Exercise. *Prove from the axioms, that for each function f , the* **domain of definition** *of f*

$$\mathrm{Domain}(f) =_{\mathrm{df}} \{x \mid (\exists y)[(x, y) \in f]\}$$

and the **image** *of f*

$$\mathrm{Image}(f) =_{\mathrm{df}} \{y \mid (\exists x)[(x, y) \in f]\}$$

are sets, and for each set B,

$$\text{if } \mathrm{Image}(f) \subseteq B, \text{ then } f : \mathrm{Domain}(f) \to B.$$

As a consequence,

$$\text{if } \mathrm{Function}(f), \text{ then } f : \mathrm{Domain}(f) \to \mathrm{Image}(f).$$

4.16. About functions as sets of ordered pairs. This "identification" of a function $f : A \to B$ with the set of pairs $\{(x, y) \in A \times B \mid f(x) = y\}$ has generated some controversy, because we have natural "operational" intuitions about functions and by "function" we often mean a formula or a rule of computation. For example, the two functions on the reals

$$f(x, y) =_{\mathrm{df}} (x + y)^2,$$
$$g(x, y) =_{\mathrm{df}} x^2 + 2xy + y^2$$

are identified in set theory, although they are obviously different as computation rules. There is no problem with this if we keep clear in our minds that the "definition" **4.14** does not replace the intuitive notion of function but only represents it within *set theory*, faithfully for the uses to which we put this notion *within set theory*:[9] first (and foremost), to define equinumerosity and the size comparison condition for sets with no reference to objects outside axiomatic set theory,

$$A =_c B \iff_{\mathrm{df}} (\exists f)[f : A \rightarrowtail\!\!\!\rightarrow B] \iff (A \rightarrowtail\!\!\!\rightarrow B) \neq \emptyset,$$
$$A \leq_c B \iff_{\mathrm{df}} (\exists f)[f : A \rightarrowtail B] \iff (A \rightarrowtail B) \neq \emptyset.$$
$$A <_c B \iff_{\mathrm{df}} A \leq_c B \,\&\, A \neq_c B.$$

4.17. Exercise. *Prove from the axioms that $A =_c B \implies \mathcal{P}(A) =_c \mathcal{P}(B)$, and identify which of the axioms you are using.*

4.18. Exercise. *Prove from the axioms that if $A =_c A'$ and $B =_c B'$, then*

$$A \uplus B =_c A' \uplus B', \quad A \times B =_c A' \times B', \quad (A \to B) =_c (A' \to B').$$

---

[9]Whether the intuitive notion of function-as-computation-rule can also be represented faithfully in set theory is an interesting problem, for which there does not exist yet a generally accepted solution.
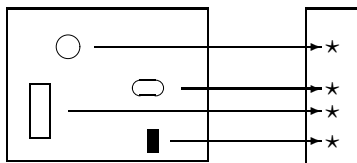
FIGURE **4.3**. Cantor's construction of $|A|$ for a four-element set.

4.19. Cantor's notion of cardinal numbers. Ironically, one of the most difficult, intuitive mathematical notions to represent faithfully in set theory is that of *cardinal number*, a most basic concept of the subject. Here is how Cantor introduced it in the same 1895 paper from which we quoted the "definition" of sets in the Introduction:

> Every set A has a definite 'power', which we will also call its 'cardinal number'.
>
> We will call by the name 'power' or 'cardinal number' of A the general concept, which by means of our active faculty of thought, arises from the set A when we make abstraction of its various elements x and of the order in which they are given.
>
> We denote the result of this double act of abstraction, the cardinal number or power of A by $\overline{\overline{A}}$. Since every element x, if we abstract from its nature becomes a 'unit', the cardinal number $\overline{\overline{A}}$ is a definite set composed of units, and this number has existence in our minds as an intellectual image or projection of the given set A.

After some discussion, Cantor infers from this "definition" that cardinal numbers have the following two fundamental properties:

$$A =_c \overline{\overline{A}},$$
$$\text{if } A =_c B, \text{ then } \overline{\overline{A}} = \overline{\overline{B}}.$$

The first of these flows quite naturally from Cantor's conception: the process of abstraction which associates with each $x \in A$ a corresponding "unit" u_x evidently defines a correspondence $x \mapsto u_x$ between A and $\overline{\overline{A}}$. Cantor gives a brief argument for the second whose key phrase is that

> $\overline{\overline{A}}$ grows, so to speak, out of A in such a way that from every element x of A a special unit of A arises.

To get the second property out of this we must assume that the "special units of A" depend only on "how many" members A has and not the nature of these members, which begs the question of cardinality, but there it is.

There is a third, more technical property of cardinal numbers, which Cantor uses routinely with no special mention to define and study operations which act on infinite families of sets: *for every family of sets $\mathscr{E}$, $\{\overline{\overline{X}} \mid X \in \mathscr{E}\}$ is a set*. Thus, however we understand Cantor's construction, it is quite clear what we

must do to represent it faithfully in set theory. We substitute modern notation for Cantor's awkward double bar symbolism.

4.20. Problem of Cardinal Assignment: *to define an operation* $|A|$ *on the class of sets which satisfies*

(C1) $$A =_c |A|,$$

(C2) $$\text{if } A =_c B, \text{ then } |A| = |B|,$$

(C3) $$\text{for each set of sets } \mathscr{E}, \{|X| \mid X \in \mathscr{E}\} \text{ is a set.}$$

The problem is quite difficult and it was not solved until the twenties, by von Neumann, whose elegant construction uses both the Axioms of Choice and Replacement. We will present it in Chapter **12**, as the culmination of a lot of work. In the meantime, notice that there are plenty of definite operations which satisfy (C1) and (C3), including the obvious $|A| = A$! And as it turns out, these two properties suffice for the development of a very satisfactory theory of cardinality.

4.21. Cardinal numbers. A (weak) **cardinal assignment** is any definite operation on sets $A \mapsto |A|$ which satisfies (C1), and (C3), and it is a **strong cardinal assignment** if it also satisfies (C2). The **cardinal numbers** (relative to a given cardinal assignment) are its values,

(C4) $$\text{Card}(\kappa) \iff \kappa \in \text{Card} \iff_{\text{df}} (\exists A)[\kappa = |A|].$$

4.22. Exercise. *Prove that for any cardinal assignment and any two sets* A, B,

$$|A| = |B| \Longrightarrow A =_c B,$$

so that the converse of (C2) is true of all cardinal assignments, including the weak ones.

We fix one, specific (possibly weak) cardinal assignment and we define the arithmetical operations on the cardinals as follows:

$$\kappa + \lambda =_{\text{df}} |\kappa \uplus \lambda| \quad =_c \kappa \uplus \lambda,$$
$$\kappa \cdot \lambda =_{\text{df}} |\kappa \times \lambda| \quad =_c \kappa \times \lambda,$$
$$\kappa^\lambda =_{\text{df}} |(\lambda \to \kappa)| =_c (\lambda \to \kappa).$$

The infinitary operations are defined similarly:[10]

$$\sum_{i \in I} \kappa_i =_{\text{df}} |\{(i, x) \in I \times \bigcup_{i \in I} \kappa_i \mid x \in \kappa_i\}|,$$
$$\prod_{i \in I} \kappa_i =_{\text{df}} |\prod_{i \in I} \kappa_i|.$$

The motivation is clear, e.g., the sum $\kappa + \lambda$ is the "number of elements" in the set we get by putting together disjoint copies of κ and λ.

---

[10]It is traditional to use the cap Greek Π to denote both the Cartesian product of sets and the cardinal operation of infinite product, and it does not really cause any confusion.

Notice that there is only one choice for $|\emptyset|$,

$$0 =_{df} |\emptyset| = \emptyset, \qquad (4\text{-}15)$$

since only $|\emptyset| = \emptyset$ satisfies $\emptyset =_c |\emptyset|$. It is also convenient to set

$$1 =_{df} |\{0\}|, \quad 2 =_{df} |\{0, 1\}| \qquad (4\text{-}16)$$

so we have handy names for the cardinal numbers of singletons and double-tons.

4.23. Exercise. *For all sets A, B, $|A \cup B| \leq_c |A| + |B|$, and*

$$\textit{if } A \cap B = \emptyset, \textit{ then } |A \cup B| =_c |A| + |B|.$$

4.24. Exercise. *For all cardinal numbers, if $\kappa_1 =_c \kappa_2$ and $\lambda_1 =_c \lambda_2$, then*

$$\kappa_1 + \lambda_1 =_c \kappa_2 + \lambda_2, \quad \kappa_1 \cdot \lambda_1 =_c \kappa_2 \cdot \lambda_2, \quad \kappa_1^{\lambda_1} =_c \kappa_2^{\lambda_2}.$$

4.25. Cardinal arithmetic. It looks quite silly to develop the theory of a weak cardinal assignment which could be just the identity $|X| = X$, but the notation of cardinal numbers and the arithmetical operations on them is useful for expressing simply complex "equinumerosities". Consider the formula

$$\kappa^{(\lambda+\mu)} =_c \kappa^\lambda \cdot \kappa^\mu. \qquad (4\text{-}17)$$

It looks obvious, it is true by Exercise **4.28**, and it expresses exactly the same fact as

$$((\lambda \uplus \mu) \to \kappa) =_c (\lambda \to \kappa) \times (\mu \to \kappa), \qquad (4\text{-}18)$$

more simply, or so some would say. More significantly:

(1) the systematic development of formulas like (4-17) leads to a *cardinal arithmetic* which in the end suggests new (and useful) facts about equinumerosities by analogy with ordinary arithmetic; and

(2) when we do construct von Neumann's strong cardinal assignment, we will have already proved all the interesting facts about cardinals with $=_c$ in place of $=$: all we will need to do is remove in our minds the subscript $_c$ from facts we already understand, because of the following, simple fact.

4.26. Exercise. *If cardinal numbers are defined using a strong cardinal assignment, then for all cardinals κ, λ,*

$$|\kappa| = \kappa \textit{ and } \kappa =_c \lambda \iff \kappa = \lambda. \qquad (4\text{-}19)$$

The basic technique for proving identities of cardinal arithmetic is to use systematically (C1) and the replacement properties of Exercise **4.18**. To prove the associativity of cardinal addition, for example, we compute:

$$\begin{aligned}
\kappa + (\lambda + \mu) &=_c \kappa \uplus (\lambda + \mu) && \text{by def.,} \\
&=_c \kappa \uplus (\lambda \uplus \mu) && \text{by def., (C1) and } \mathbf{4.18}, \\
&=_c (\kappa \uplus \lambda) \uplus \mu && \text{by a direct argument,} \\
&=_c (\kappa + \lambda) + \mu && \text{reversing the steps.}
\end{aligned}$$

The mathematical essence of the proof is the alleged "direct argument", which in this case is quite easy:

4.27. Exercise. *Prove that for any three sets K, L, M,*

$$K \uplus (L \uplus M) =_c (K \uplus L) \uplus M.$$

4.28. Exercise. *Prove* (4-17), *by showing first that for any three sets K, L, M,*

$$\text{if } L \cap M = \emptyset, \text{ then } ((L \cup M) \to K) =_c (L \to K) \times (M \to K).$$

To see how the more technical condition (C3) comes into play, consider the equation

$$\left| \bigcup_{i \in I} A_i \right| =_c \sum_{i \in I} |A_i|, \tag{4-20}$$

which should certainly be true when the sets in the family $(i \mapsto A_i)_{i \in I}$ are pairwise disjoint. To make sense of it, before trying to prove it, we must know that *there is a function $(i \mapsto |A_i|)$*, and the proof of this requires (C3), as follows:

4.29. Lemma. *For every indexed family of sets $A = (i \mapsto A_i)_{i \in I}$, there exists a function $f : I \to f[I]$ such that*

$$f(i) = |A_i| \quad (i \in I).$$

PROOF. By (C3) with

$$\mathscr{E} = \{A_i \mid i \in I\} = A[I],$$

there exists a set W which contains every $|A_i|$ for $i \in I$, and we can set

$$f =_{\mathrm{df}} \{(i, w) \in I \times W \mid w = |A_i|\}. \hspace{2em} \dashv$$

As it happens, equations like (4-20) cannot be proved without the Axiom of Choice, so we will have little need of (C3) before Chapter **8**.

4.30. Structured sets. A *topological space* is a set X of *points* endowed with a *topological structure*, which is determined by a collection $\mathscr{T}$ of subsets of X satisfying the following three properties:

1. $\emptyset, X \in \mathscr{T}$.
2. $A, B \in \mathscr{T} \Longrightarrow A \cap B \in \mathscr{T}$.
3. For every family $\mathscr{E} \subseteq \mathscr{T}$ of sets in $\mathscr{T}$, the unionset $\bigcup \mathscr{E}$ is also in $\mathscr{T}$.

A family of sets $\mathscr{T}$ with these properties is called a **topology** on X, with **open sets** its members and **closed sets** the complements of open sets relative to X, i.e., all $X \setminus G$ with G open.

Notions like this of sets "endowed" with structure abound in mathematics: there are graphs, groups, vector spaces, sheaves, manifolds, partially ordered sets, etc. etc. In each of these cases we have a set X, typically called "the space", and a complex of related objects which impose a structure on the space—functions, families of sets, other spaces with their own structure, etc. The pairing operation provides a simple and flexible way to model such notions faithfully in set theory.

A **structured set** is a pair

$$U = (A, \mathscr{S}), \tag{4-21}$$

where $A = \text{Field}(U)$ is a set, the *field* or *space* of U, and $\mathscr{S}$ is an arbitrary object, the *frame*[11] of U.

For example, a **topological space** is a structured set $(X, \mathscr{T})$, where the frame $\mathscr{T}$ is a topology on X, as above. A *group* is a structured set

$$U = (G, (e, \cdot)) \qquad (4\text{-}22)$$

where $e \in G$ and $\cdot : G \times G \to G$ is a binary function, satisfying the *group axioms*, which do not concern us here. Notice that by the definition of triple (4-3), definition (4-22) is equivalent to

$$U = (G, e, \cdot). \qquad (4\text{-}23)$$

It is quite common that the frame of a structured set is a tuple of objects, and then the structured set is also a tuple, with its field as the first element. We will meet numerous examples of this in the sequel.

Following usual mathematical practice, we will systematically confuse a structured set with its field when the frame is understood from the context or is not relevant. For example, we will refer to "the topological space X" rather than "$(X, \mathscr{T})$", with "points" the members of X, "subsets" the subsets of X, etc. In the general case, the members of a structured set U are the members of $\text{Field}(U)$,

$$x \in U \iff_{\text{df}} x \in \text{Field}(U), \qquad (4\text{-}24)$$

the subsets of U are the subsets of $\text{Field}(U)$, etc. Notice that the terminological convention (4-24) cannot possibly cause a misunderstanding: since we have (deliberately) not settled on a specific pairing operation—and have even left open the possibility that $(A, \mathscr{S})$ may be an atom (!)—the statement

$$x \in (A, \mathscr{S})$$

cannot possibly mean anything until we define it, and we just did this by (4-24).

Problems for Chapter 4

The definition of ordered pair in the proof of **4.2** is due to the Polish set theorist and topologist Kuratowski. A few years before Kuratowski's construction, the American analyst Wiener had discovered the following, somewhat more complex but interesting solution of this problem.

x4.1. (Wiener) The properties (OP1) and (OP2) in **4.1** hold with the following definition of pair:

$$(x, y) =_{\text{df}} \{\{\emptyset, \{x\}\}, \{\{y\}\}\}.$$

---

[11]It would be more suggestive to call $\mathscr{S}$ *the structure* of the structured set $(A, \mathscr{S})$, but the word is heavily overloaded in logic and set theory and it is best to avoid attaching it to one more precise notion. Some people call "structures" what we have called "structured sets" here, at least when they are simple enough.

x4.2. The properties (OP1) and (OP2) in **4.1** hold with the following definition of pair:
$$(x, y) =_{df} \{\{0, x\}, \{1, y\}\},$$
where 0, 1 are two distinct objects, e.g., those defined by (4-16). HINT. Compute $(0, 0), (0, 1), (1, 1)$ and $(1, 0)$ (with this pair) to see what goes on, and then take cases on whether $x = y$ or not.

x4.3. Prove from the axioms that for all sets A, B, C,
$$((A \times B) \to C) =_c (A \to (B \to C)).$$

x4.4. Prove from the axioms the theorem of Cantor **2.21**, that for every set A, $A <_c \mathcal{P}(A)$. Which axioms do you need?

x4.5. A binary relation $\sim \subseteq (A \times A)$ is an equivalence relation on A if and only if there exists some set Q and a surjection
$$\pi : A \twoheadrightarrow Q \qquad\qquad (4\text{-}25)$$
such that
$$x \sim y \iff \pi(x) = \pi(y). \qquad\qquad (4\text{-}26)$$

When (4-25) and (4-26) hold, we call Q a **quotient** of A by $\sim$ and π a **determining surjection** of $\sim$. The proof of **4.12** yields the quotient $[\![A/\sim]\!]$ and the determining surjection $(x \mapsto [x/\sim])$, but in specific cases there exist other quotients which help us understand better the structure of the equivalence relation at hand.

x4.6. Suppose $\sim$ is an equivalence relation on A and $\pi : A \to A$ satisfies
$$x \sim y \Longrightarrow \pi(x) = \pi(y) \in [x/\sim].$$
Prove that π is a determining surjection witnessing that its image $\pi[A] \subseteq A$ is a quotient of A by $\sim$.

x4.7. Fix an element $x_0 \in A$ in some set and define on the function space $(A \to B)$ the relation
$$f \sim g \iff_{df} f(x_0) = g(x_0).$$
Prove that $\sim$ is an equivalence relation and find a determining surjection $\pi : (A \to B) \twoheadrightarrow B$ which witnesses that B is a quotient of $(A \to B)$ by $\sim$.

x4.8. Let $x_0 \neq x_1$ be two distinct elements in A and find a determining surjection which witnesses that $(B \times B)$ is a quotient of $(A \to B)$ by the equivalence relation
$$f \sim g \iff_{df} f(x_0) = g(x_0) \,\&\, f(x_1) = g(x_1).$$

x4.9. Suppose $\sim$ is an equivalence relation on A and $f : A \to A$ is a function which **respects** $\sim$, i.e.,
$$x \sim y \Longrightarrow f(x) \sim f(y).$$
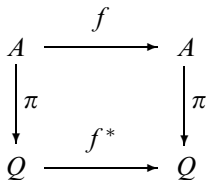
FIGURE **4.4**.

Let Q be any quotient of A by $\sim$. Prove that there exists a unique function $f^* : Q \to Q$ such that the diagram in Figure **4.4** *commutes*, i.e., $\pi f = f^*\pi$,

$$f^*(\pi x) = \pi(f(x)), \quad (x \in A),$$

where $\pi : A \twoheadrightarrow Q$ is a determining surjection.

x4.10. For all cardinal numbers, κ, λ, μ,

$$\kappa + 0 =_c \kappa, \quad \kappa \cdot 0 =_c 0, \quad \kappa \cdot 1 =_c \kappa.$$

x4.11. For all cardinal numbers κ, λ, μ,

$$\kappa + (\lambda + \mu) =_c (\kappa + \lambda) + \mu,$$
$$\kappa + \lambda =_c \lambda + \kappa.$$

x4.12. For all cardinal numbers κ, λ, μ,

$$\kappa \cdot (\lambda \cdot \mu) =_c (\kappa \cdot \lambda) \cdot \mu,$$
$$\kappa \cdot \lambda =_c \lambda \cdot \kappa,$$
$$\kappa \cdot (\lambda + \mu) =_c \kappa \cdot \lambda + \kappa \cdot \mu.$$

x4.13. For all cardinals κ, $|\mathcal{P}(\kappa)| =_c 2^\kappa$.

x4.14. For all cardinal numbers κ, λ, μ,

$$\kappa^0 =_c 1, \quad \kappa^1 =_c \kappa, \quad \kappa^2 =_c \kappa \cdot \kappa.$$

x4.15. For all cardinal numbers κ, λ, μ,

$$(\kappa \cdot \lambda)^\mu =_c \kappa^\mu \cdot \lambda^\mu,$$
$$\kappa^{(\lambda+\mu)} =_c \kappa^\lambda \cdot \kappa^\mu,$$
$$(\kappa^\lambda)^\mu =_c \kappa^{\lambda \cdot \mu}.$$

x4.16. For all cardinal numbers κ, λ, μ

$$\kappa \leq_c \mu \implies \kappa + \lambda \leq_c \mu + \lambda,$$
$$\kappa \leq_c \mu \implies \kappa \cdot \lambda \leq_c \mu \cdot \lambda,$$
$$\lambda \leq_c \mu \implies \kappa^\lambda \leq_c \kappa^\mu \quad (\kappa \neq 0),$$
$$\kappa \leq_c \lambda \implies \kappa^\mu \leq_c \lambda^\mu.$$

For what values of λ, μ does the third implication fail when $\kappa = 0$?

Caution! As we will see later, strict inequalities between infinite cardinal numbers are not always respected by the algebraic operations, for example, we may have

$$\kappa <_c \mu, \text{ but } \kappa + \lambda =_c \mu + \lambda.$$

x4.17. For all A, B and all cardinals κ, λ,

$$\prod_{i \in A} B = (A \to B), \quad \prod_{i \in \lambda} \kappa = \kappa^\lambda.$$

x4.18. Suppose $a \neq b$ are two distinct objects and κ_a, κ_b are cardinals, and prove that

$$\kappa_a + \kappa_b =_c \sum_{i \in \{a,b\}} \kappa_i,$$
$$\kappa_a \cdot \kappa_b =_c \prod_{i \in \{a,b\}} \kappa_i.$$

x4.19. Prove that for all indexed families of cardinals,

$$\kappa \cdot \sum_{i \in I} \lambda_i =_c \sum_{i \in I} \kappa \cdot \lambda_i.$$

x4.20. Show that $\kappa \cdot \lambda = 0 \iff \kappa = 0 \vee \lambda = 0$. Show also one of the directions of the equivalence

$$\prod_{i \in I} \kappa_i = 0 \iff (\exists i \in I)[\kappa_i = 0]. \tag{4-27}$$

(If you think you can show both directions of (4-27), think again and find your error, since one direction requires the Axiom of Choice.)

x4.21. The notion of equivalence according to Zermelo **3.28** coincides with equinumerosity, i.e.,

$$A =_c B \iff A \sim_Z B.$$

The definition **2.6** of infinite and finite sets refers to the set $\mathbb{N}$ of natural numbers and we cannot study these concepts axiomatically before we give a definition of $\mathbb{N}$ directly from the axioms. There is, however, another, simpler definition of these notions which we can give now and which we will later prove (with the Axiom of Choice) equivalent to **2.6**.

4.31. Definition. A set A is **infinite according to Dedekind** if there exists an injection

$$f : A \rightarrowtail B \subsetneq A$$

from A into a proper subset of itself. If A is not Dedekind-infinite, then it is **Dedekind-finite**.

x4.22. If A is Dedekind-infinite and $A =_c B$, then B is also Dedekind-infinite.

x4.23. If A is Dedekind-finite, then every subset of A is also Dedekind-finite.

x4.24. Every set I which satisfies the conditions

$$\emptyset \in I, \ (\forall x)[x \in I \implies \{x\} \in I]$$

is Dedekind-infinite.
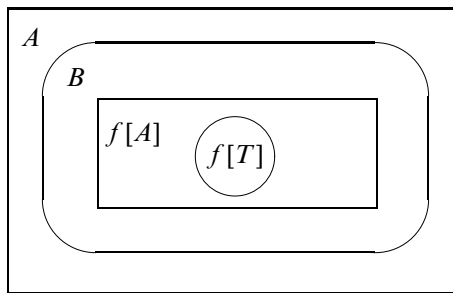
FIGURE **4.5**. Zermelo's proof of the Schröder-Bernstein Theorem.

Most of the properties of Dedekind-finite sets require the Axiom of Choice for their proof. Here is one which does not, but it requires some thinking.

****x4.25.** If A is Dedekind-finite and $t \notin A$, then $A \cup \{t\}$ is also Dedekind-finite. HINT. Assume $\pi : A \cup \{t\} \rightarrowtail A \cup \{t\}$ misses some point in $A \cup \{t\}$, and consider cases on whether it misses t; or $\pi(t) = t$; or "otherwise", which is where the thinking is required.

The classical proof of the Schröder-Bernstein Theorem **2.26** uses induction on the natural numbers and we cannot justify it now. In the next two problems we outline a very different proof (due to Zermelo), somewhat opaque in its motivation but elegant, short and in no way dependent on the natural numbers.

****x4.26.** If $A' \subseteq B \subseteq A$ and $A =_c A'$, then also $A =_c B$. HINT. Suppose $f : A \rightarrowtail A'$ is a correspondence which witnesses that $A =_c f[A] = A'$, and

$$Q = B \setminus f[A]$$

is the set of objects in B which are not in the image of A by f. We define the family of subsets of A

$$\mathscr{T} = \{X \mid Q \cup f[X] \subseteq X\}$$

and we first verify that its intersection is a member of it,

$$T =_{\mathrm{df}} \bigcap \mathscr{T} \in \mathscr{T},$$

so that $Q \cup f[T] \subseteq T$. With a bit more work we can show that, in fact, $T = Q \cup f[T]$; this identity then implies that

$$B = T \cup (f[A] \setminus f[T]),$$

which completes the proof, since T and $(f[A] \setminus f[T])$ are disjoint sets and their union is (easily now) equinumerous with A.

****x4.27.** Use Problem **x4.26** to give a proof of the Schröder-Bernstein Theorem from the axioms. HINT. If $f : A \rightarrowtail C$ and $g : C \rightarrowtail A$, then

$$A =_c gf[A] \subseteq g[C] \subseteq A, \quad g[C] =_c C.$$

CHAPTER 5

THE NATURAL NUMBERS

Our fundamental intuitive understanding of the natural numbers is that there is a (least) number 0, that every number n has a successor Sn, and that if we start with 0 and construct in sequence the successor of every number

$$0, \ S0 = 1, \ S1 = 2, \ S2 = 3, \ \ldots$$

forever, then in time we will reach every natural number. In set theoretic terms we can capture this intuition by the following axiomatic characterization.

5.1. Definition. A **Peano system** or **system of natural numbers** is any structured set

$$(\mathbb{N}, 0, S) = (\mathbb{N}, (0, S))$$

which satisfies the following conditions.

1. $\mathbb{N}$ is a set which contains the element 0, $0 \in \mathbb{N}$.
2. S is a function on $\mathbb{N}$, $S : \mathbb{N} \to \mathbb{N}$.
3. S is an injection, $Sn = Sm \Longrightarrow n = m$.
4. For each $n \in \mathbb{N}$, $Sn \neq 0$.
5. **Induction Principle.** For each $X \subseteq \mathbb{N}$,

$$[0 \in X \ \& \ (\forall n \in \mathbb{N})[n \in X \Longrightarrow Sn \in X]] \Longrightarrow X = \mathbb{N}.$$

These obvious properties of the natural numbers are called the **axioms of Peano** in honor of the Italian logician and mathematician who first proposed them as an axiomatic foundation of number theory, following their earlier formulation by Dedekind. Most significant among them is the Induction Principle, whose typical application is illustrated in the proof of the next lemma.

5.2. Lemma. *In a Peano system* $(\mathbb{N}, 0, S)$, *every element* $n \neq 0$ *is a successor,*

$$\textit{if } n \neq 0, \ \textit{then } (\exists m \in \mathbb{N})[n = Sm],$$

and for each n, $Sn \neq n$.

PROOF. To prove the first assertion by the Induction Principle, it is enough to show that the set

$$X = \{n \in \mathbb{N} \mid n = 0 \vee (\exists m \in \mathbb{N})[n = Sm]\}$$

satisfies the conditions

$$0 \in X, \quad (\forall n \in \mathbb{N})[n \in X \Longrightarrow Sn \in X],$$

and both of these are obvious from the definition of X. In the same way, for the second assertion it is enough to verify that $S0 \neq 0$ (which holds because, in general, $Sn \neq 0$) and that $Sn \neq n \Longrightarrow SSn \neq Sn$: this holds because S is one-to-one, so that $SSn = Sn \Longrightarrow Sn = n$. ⊣

Number theory is one of the richest and most sophisticated fields of mathematics and it is by no means obvious that it can be developed on the basis of these five, simple properties; in fact they do not suffice, one also needs to use set theory which (in its naive form) Peano took for granted, as part of "logic". Here we will only show that the axioms imply the first, most basic properties of addition, multiplication and the ordering on the natural numbers, which is all we need. The proofs we will give, however, are characteristic examples of the use of the Peano axioms in the more advanced parts of the theory of numbers.

If number theory can be developed from the Peano axioms, then to give a faithful representation of the natural numbers in set theory, it is enough to prove from the axioms the following two theorems.

5.3. Theorem (Existence of the natural numbers). *There exists at least one Peano system* $(\mathbb{N}, 0, S)$.

5.4. Theorem (Uniqueness of the natural numbers). *For any two Peano systems* $(\mathbb{N}_1, 0_1, S_1)$ *and* $(\mathbb{N}_2, 0_2, S_2)$, *there exists one* (*and only one*) *bijection*

$$\pi : \mathbb{N}_1 \rightarrowtail\!\!\!\rightarrow \mathbb{N}_2$$

such that

$$\pi(0_1) = 0_2,$$
$$\pi(S_1 n) = S_2 \pi(n) \quad (n \in \mathbb{N}_1).$$

A bijection π which satisfies these identities is an **isomorphism** of the two systems $(\mathbb{N}_1, 0_1, S_1)$ and $(\mathbb{N}_2, 0_2, S_2)$, so that the theorem asserts that *any two Peano systems are* (*uniquely*) *isomorphic*.

The Existence Theorem is very simple and we can prove it immediately.

5.5. Proof of the existence of natural numbers, **5.3**. The Axiom of Infinity (**VI**) guarantees the existence of a set I such that

$$\emptyset \in I,$$
$$(\forall n)[n \in I \Longrightarrow \{n\} \in I].$$

Using this I, first we define the family of sets

$$\mathscr{I} = \{X \subseteq I \mid \emptyset \in X \,\&\, (\forall n)[n \in X \Longrightarrow \{n\} \in X]\}$$

so that obviously $I \in \mathscr{I}$, and then we set

$$\mathbb{N} = \bigcap \mathscr{I}, \quad 0 = \emptyset, \quad S = \{(n, m) \in \mathbb{N} \times \mathbb{N} \mid m = \{n\}\}.$$
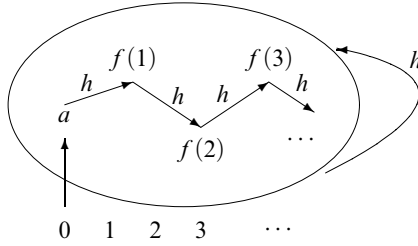
FIGURE **5.1**. The Recursion Theorem.

To finish off the proof, it suffices to verify that this triple $(\mathbb{N}, 0, S)$ is a Peano system. To begin with, $\mathbb{N} \in \mathscr{I}$, because $X \in \mathscr{I} \Longrightarrow \emptyset \in X$ and hence $\emptyset \in \bigcap \mathscr{I} = \mathbb{N}$, and by the same thinking,

$$n \in \mathbb{N} \Longrightarrow (\forall X \in \mathscr{I})[n \in X] \Longrightarrow (\forall X \in \mathscr{I})[\{n\} \in X] \Longrightarrow \{n\} \in \mathbb{N}.$$

This implies immediately the first two of the Peano axioms, the next two hold because (in general, for all n, m) $\{n\} = \{m\} \Longrightarrow n = m$ and $\{n\} \neq \emptyset$, and the Induction Principle follows directly from the definition of $\mathbb{N}$ as an intersection. ⊣

To prove the Uniqueness Theorem **5.4**, we need the next fundamental result of axiomatic number theory.

5.6. Recursion Theorem. *Assume that $(\mathbb{N}, 0, S)$ is a Peano system, E is some set, $a \in E$, and $h : E \to E$ is some function: it follows that there exists exactly one function $f : \mathbb{N} \to E$ which satisfies the identities*

$$f(0) = a,$$
$$f(Sn) = h(f(n)) \quad (n \in \mathbb{N}).$$

The Recursion Theorem justifies the usual way by which we define functions on the natural numbers, **by recursion**[12] (or induction): to define $f : \mathbb{N} \to E$, first we specify the value $f(0) = a$ and then we supply a function $h : E \to E$ which determines the value $f(Sn)$ of f at every successor Sn from the value $f(n)$ at its predecessor n, $f(Sn) = h(f(n))$. Our basic intuition about the natural numbers with which we started this chapter clearly justifies such definitions, so we should also be able to justify them on the basis of the axioms.

Before we establish the Recursion Theorem, let us use it in the next proof which is a typical example of the way it is applied.

5.7. Proof of the uniqueness of the natural numbers, **5.4**. We assume that $(\mathbb{N}_1, 0_1, S_1)$ and $(\mathbb{N}_2, 0_2, S_2)$ are Peano systems. *By the Recursion Theorem on*

---

[12]The terms "recursion" and "induction" are often used synonymously in mathematics. We will follow the more recent convention which distinguishes **recursive definitions** from **inductive proofs**.

$(\mathbb{N}_1, 0_1, S_1)$ with $E = \mathbb{N}_2$, $a = 0_2$, $h = S_2$, there exists exactly one function $\pi : \mathbb{N}_1 \to \mathbb{N}_2$ which satisfies the identities

$$\pi(0_1) = 0_2,$$
$$\pi(S_1 n) = S_2 \pi(n) \quad (n \in \mathbb{N}_1),$$

and it suffices to verify that this π is a (one-to-one) correspondence.

(1) π **is a surjection**, $\pi : \mathbb{N}_1 \twoheadrightarrow \mathbb{N}_2$. Obviously $0_2 \in \pi[\mathbb{N}_1]$ since $0_2 = \pi(0_1)$, and

$$
\begin{aligned}
m \in \pi[\mathbb{N}_1] &\implies (\exists n \in \mathbb{N}_1)[m = \pi(n)] \\
&\implies (\exists n \in \mathbb{N}_1)[S_2 m = S_2 \pi(n) = \pi(S_1 n)] \\
&\implies S_2 m \in \pi[\mathbb{N}_1],
\end{aligned}
$$

so that by *the Induction Principle on* $(\mathbb{N}_2, 0_2, S_2)$, $\pi[\mathbb{N}_1] = \mathbb{N}_2$.

(2) π **is an injection**, $\pi : \mathbb{N}_1 \rightarrowtail \mathbb{N}_2$. It suffices to verify that if we set

$$X = \{n \in \mathbb{N}_1 \mid (\forall m \in \mathbb{N}_1)[\pi(m) = \pi(n) \implies m = n]\},$$

then

$$0_1 \in X, \quad n \in X \implies S_1 n \in X,$$

since together with *the Induction Principle on* $(\mathbb{N}_1, 0_1, S_1)$, this implies that $X = \mathbb{N}_1$, which means that π is an injection. For the first condition,

$$
\begin{aligned}
m \neq 0_1 &\implies m = S_1 m' \text{ for some } m', \text{ by Lemma } \mathbf{5.2} \\
&\implies \pi(m) = \pi(S_1 m') = S_2 \pi(m') \neq 0_2,
\end{aligned}
$$

so that if $\pi(m) = \pi(0_1) = 0_2$, then $m = 0_1$ and $0_1 \in X$. For the second condition, it is enough to show that

$$n \in X \,\&\, \pi(m) = \pi(S_1 n) \implies m = S_1 n.$$

By the hypothesis

$$\pi(m) = \pi(S_1 n) = S_2 \pi(n) \neq 0_2,$$

which implies that $m \neq 0_1$, since $\pi(0_1) = 0_2$ and $0_1 \in X$. By Lemma $\mathbf{5.2}$ again, $m = S_1 m'$ for some $m' \in \mathbb{N}_1$,

$$\pi(m) = \pi(S_1 m') = S_2 \pi(m')$$

and the hypothesis $\pi(m) = \pi(S_1 n)$ yields

$$S_2 \pi(m') = S_2 \pi(n),$$

which implies $\pi(m') = \pi(n)$. This, in turn, implies $m' = n$ because $n \in X$, so that $m = S_1 m' = S_1 n$, the required conclusion. $\dashv$

5.8. Proof of the Recursion Theorem. Assume the hypotheses and define first the set $\mathscr{A}$ of all *approximations* of the function which we want to construct:

$$p \in \mathscr{A} \Longleftrightarrow_{\text{df}} \text{Function}(p) \tag{5-1}$$
$$\& \text{Domain}(p) \subseteq \mathbb{N} \,\&\, \text{Image}(p) \subseteq E$$
$$\& \, 0 \in \text{Domain}(p) \,\&\, p(0) = a$$
$$\& \, (\forall n \in \mathbb{N})[Sn \in \text{Domain}(p)$$
$$\Longrightarrow n \in \text{Domain}(p) \,\&\, p(Sn) = h(p(n))].$$

In words, each $p \in \mathscr{A}$ is a function with domain a subset of $\mathbb{N}$ and values in E; by the third clause, $\text{Domain}(p)$ contains 0; and by the last one, $\text{Domain}(p)$ is "closed downward", i.e., if $Sn \in \text{Domain}(p)$, then $n \in \text{Domain}(p)$, and the value $p(Sn)$ is determined by $p(n)$. Some examples of approximations are

$$\{(0, a)\}, \{(0, a), (S0, f(a))\}, \{(0, a), (S0, f(a)), (SS0, f(f(a)))\}, \dots$$

and they suggest how the required function f is "built up" stage-by-stage by the recursive definition. To prove the theorem, we need to show (from the axioms only, without " ... " or "etc.") that there is exactly one approximation with domain of definition all of $\mathbb{N}$.

Lemma. *For all $p, q \in \mathscr{A}$ and $n \in \mathbb{N}$,*

$$n \in \text{Domain}(p) \cap \text{Domain}(q) \Longrightarrow p(n) = q(n).$$

Proof. The set

$$X = \{n \in \mathbb{N} \mid (\forall p, q \in \mathscr{A})\Big[n \in \text{Domain}(p) \cap \text{Domain}(q) \Longrightarrow p(n) = q(n)\Big]\}$$

clearly contains 0, since every $p \in \mathscr{A}$ satisfies $p(0) = a$. If

$$n \in X \,\&\, p \in \mathscr{A} \,\&\, q \in \mathscr{A} \,\&\, Sn \in \text{Domain}(p) \cap \text{Domain}(q),$$

then

$$p(Sn) = h(p(n)) \qquad \text{because } p \in \mathscr{A},$$
$$= h(q(n)) \qquad \text{because } p(n) = q(n),$$
$$= q(Sn) \qquad \text{because } q \in \mathscr{A},$$

so that if $n \in X$, then $Sn \in X$. It follows by the Induction Principle that $X = \mathbb{N}$, which completes the proof. $\dashv$ (Lemma)

The Lemma implies immediately that *at most one* function $f : \mathbb{N} \to E$ belongs to $\mathscr{A}$, so to complete the proof of the theorem we need only show that *at least one* such f exists. This is the union

$$f = \bigcup \mathscr{A} = \{(n, w) \mid (\exists p \in \mathscr{A})[n \in \text{Domain}(p) \,\&\, p(n) = w]\},$$

which is a function, because

$$(n, w) \in f \,\&\, (n, w') \in f \implies (\exists p, q \in \mathscr{A})[(n, w) \in p \,\&\, (n, w') \in q]$$
$$\implies w = w' \text{ from the Lemma,}$$

and then the definition of $\mathscr{A}$ and a similar calculation shows that $f \in \mathscr{A}$. The only thing left is to verify that $\mathrm{Domain}(f) = \mathbb{N}$, and for that we will use once more the Induction Principle. To begin with, $0 \in \mathrm{Domain}(f)$, since $0 \in \mathrm{Domain}(p)$ for every $p \in \mathscr{A}$ and $\mathscr{A} \neq \emptyset$. If $n \in \mathrm{Domain}(f)$, then there exists some function $p \in \mathscr{A}$ with $n \in \mathrm{Domain}(p)$, and hence (easily)

$$q = p \cup \{(Sn, h(p(n)))\} \in \mathscr{A},$$

so that $Sn \in \mathrm{Domain}(q) \subseteq \mathrm{Domain}(f)$. $\dashv$

5.9. The natural numbers. We now fix a specific Peano system $(\mathbb{N}, 0, S)$ whose members we will henceforth call **natural numbers**, or just **numbers**, when there cannot be any confusion. Following Cantor, we denote the cardinal number of $\mathbb{N}$ by the first Hebrew letter with the subscript 0, pronounced "aleph-zero",

$$\aleph_0 =_{\mathrm{df}} |\mathbb{N}|. \tag{5-2}$$

Later we will meet its followers $\aleph_1$, $\aleph_2$, etc. Functions $a : \mathbb{N} \to A$ with domain $\mathbb{N}$ are called (infinite) **sequences** and we often write their argument as a subscript,

$$a_n = a(n) \quad (n \in \mathbb{N}, a : \mathbb{N} \to A).$$

An obvious choice for $\mathbb{N}$ would be the system which we constructed in the proof of the Existence Theorem **5.3**, where $0 = \emptyset$,

$$\mathbb{N} = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \dots \}$$

and $Sn = \{n\}$. Another choice which some would prefer on philosophical grounds is to assert that there exists, in fact, a set

$$\mathbb{N} = \{0, 1, 2, \dots \}$$

of the "true natural numbers", which are not sets, and the successor function S is nothing like the artificial $(n \mapsto \{n\})$, but it is the "natural successor function" which associates with each number n "the next number" Sn. Zermelo's theory allows such non-sets (like the "true numbers") as atoms, and requires only one thing: that the system of natural numbers satisfies the Peano axioms, something which every serious person will surely grant. As far as the mathematical theory of numbers and sets is concerned, these two (and all other) choices of the objects we will call *numbers* are equivalent, since we will base all our proofs on the Peano axioms alone.

5.10. The Schröder-Bernstein Theorem 2.26. At this point, we should review the proof of this important result and verify that *we can now derive it from the axioms*; this is because the recursive definitions of the sequences of sets $\{A_n\}_{n \in \mathbb{N}}$ and $\{B_n\}_{n \in \mathbb{N}}$ are justified by the Recursion Theorem, and their basic properties are established by induction and simple manipulations of functions, which can all be based on the results in Chapter **4**.

There are many reformulations of the Recursion Theorem which are useful in applications. We state here just two of them, whose proofs illustrate how the theorem is used. Two more such results are included in the problems, **x5.20** and **x5.21**.
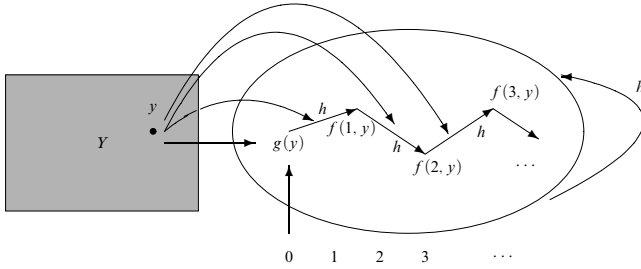
FIGURE **5.2**. The Recursion Theorem with parameters.

5.11. Corollary (Recursion with parameters). *For any two sets Y, E and functions*

$$g : Y \to E, \quad h : E \times Y \to E,$$

there exists exactly one function $f : \mathbb{N} \times Y \to E$ which satisfies the identities

$$f(0, y) = g(y) \qquad\qquad (y \in Y),$$
$$f(Sn, y) = h(f(n, y), y) \quad (y \in Y, n \in \mathbb{N}).$$

PROOF. For each $y \in Y$, we define the function $h_y : E \to E$ by the formula

$$h_y(w) = h(w, y),$$

and by the Recursion Theorem we know that there exists *exactly one* function

$$f_y : \mathbb{N} \to E$$

which satisfies the identity

$$f_y(0) = g(y),$$
$$f_y(Sn) = h_y(f_y(n)) = h(f_y(n), y).$$

It follows immediately that the function $f : \mathbb{N} \times Y \to E$ defined by the formula

$$f(n, y) =_{\text{df}} f_y(n) \quad (y \in Y, n \in \mathbb{N})$$

satisfies the conclusion of the Corollary.   ⊣

5.12. Corollary (Recursion with the argument as parameter). *For every set E, each $a \in E$, and every function $h : E \times \mathbb{N} \to E$, there is exactly one function $f : \mathbb{N} \to E$ which satisfies the identities*

$$f(0) = a,$$
$$f(Sn) = h(f(n), n).$$

Similarly, with parameters, for every

$$g : Y \to E, \quad h : E \times \mathbb{N} \times Y \to E,$$

there exists exactly one function $f : \mathbb{N} \times Y \to E$ which satisfies the identities

$$f(0, y) = g(y) \qquad\qquad (y \in Y),$$
$$f(Sn, y) = h(f(n, y), n, y) \qquad (y \in Y,\ n \in \mathbb{N}).$$

Proof. For the version with parameters, we define first a function

$$\phi : \mathbb{N} \times Y \to \mathbb{N} \times E$$

by recursion with parameters **5.11**, where the component functions First and Second are those of Exercise **4.3**:

$$\phi(0, y) = (0, g(y))$$
$$\phi(Sn, y) = (S\,\mathrm{First}(\phi(n, y)), h(\mathrm{Second}(\phi(n, y)), \mathrm{First}(\phi(n, y)), y)).$$

By induction on n, immediately,

$$\mathrm{First}(\phi(n, y)) = n,$$

and so the function ϕ satisfies the equations

$$\phi(0, y) = (0, g(y)), \quad \phi(Sn, y) = (Sn, h(\mathrm{Second}(\phi(n, y)), n, y)).$$

This implies immediately that the function

$$f(n, y) = \mathrm{Second}(\phi(n, y))$$

satisfies the required identities.

The uniqueness is easy to prove directly by induction on n. $\qquad\qquad \dashv$

We now use these results to define and establish the basic properties of addition, multiplication and the ordering on $\mathbb{N}$.

5.13. Addition and multiplication. The addition function on the natural numbers is defined by the recursion

$$\begin{aligned} n + 0 &= n, \\ n + (Sm) &= S(n + m), \end{aligned} \qquad (5\text{-}3)$$

and multiplication is defined next, using addition, by the recursion

$$\begin{aligned} n \cdot 0 &= 0, \\ n \cdot Sm &= (n \cdot m) + n. \end{aligned} \qquad (5\text{-}4)$$

In more detail, we know from **5.11** that there exists exactly one function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ which satisfies the identities

$$f(0, n) = g(n),$$
$$f(Sm, n) = h(f(m, n), n),$$

where the functions g and h have been given as sets of pairs,

$$g = \{(n, n) \in \mathbb{N} \times \mathbb{N} \mid n \in \mathbb{N}\},$$
$$h = \{((z, n), w) \in (\mathbb{N} \times \mathbb{N}) \times \mathbb{N} \mid w = Sz\},$$

and we define addition by

$$n + m =_{\mathrm{df}} f(m, n),$$

i.e., $+ = \{((n, m), w) \mid ((m, n), w) \in f\}$. Such scholastic details do not enhance understanding (rather the opposite) and we will avoid them in the future.

5.14. Theorem. *Addition is* **associative***, i.e., it satisfies the identity*

$$(n + m) + k = n + (m + k) \tag{5-5}$$

PROOF. First for $k = 0$,

$$(n + m) + 0 = n + m = n + (m + 0),$$

using twice the identity $w + 0 = w$ directly from the definition of addition. Inductively, assuming that for some k

$$(n + m) + k = n + (m + k), \tag{5-6}$$

we compute:

$$\begin{aligned}
(n + m) + Sk &= S((n + m) + k) \\
&= S(n + (m + k)) \text{ by (5-6)} \\
&= n + S(m + k) \\
&= n + (m + Sk)
\end{aligned}$$

where the steps we did not justify follow from the definition of addition. ⊣

The commutativity of addition is not quite so simple and requires two lemmas.

5.15. Lemma. *For every natural number n, $0 + n = n$.*

PROOF. By induction, $0+0 = 0$ follows from the definition, and if $0+n = n$, then $0 + Sn = S(0 + n) = Sn$. ⊣

5.16. Lemma. *For all n, m, $n + Sm = Sn + m$.*

PROOF. By induction on m, first for $m = 0$, immediately from the definition:

$$n + S0 = S(n + 0) = Sn = Sn + 0.$$

At the induction step, we assume that for some m

$$n + Sm = Sn + m \tag{5-7}$$

and we must show that

$$n + SSm = Sn + Sm.$$

Compute:

$$\begin{aligned}
n + SSm &= S(n + Sm) &&\text{by the definition} \\
&= S(Sn + m) &&\text{from (5-7)} \\
&= Sn + Sm &&\text{by the definition.} \qquad \dashv
\end{aligned}$$

5.17. Theorem. *Addition is a* **commutative** *function, i.e., it satisfies the identity*

$$n + m = m + n.$$

PROOF is by induction on m, the basis being immediate from Lemma **5.15**. At the induction step, we assume that for some specific m

$$n + m = m + n \tag{5-8}$$

and we compute:

$$
\begin{aligned}
n + Sm &= S(n + m) &&\text{by the definition} \\
&= S(m + n) &&\text{from (5-8)} \\
&= m + Sn &&\text{by the definition} \\
&= Sm + n &&\text{by Lemma } \textbf{5.16}. \qquad\qquad \dashv
\end{aligned}
$$

5.18. Exercise. *For every natural number n, the function $(s \mapsto n + s)$ is one-to-one, so that $n + s = n + t \Longrightarrow s = t$, and in particular*

$$\text{if } n + s = n, \text{ then } s = 0.$$

5.19. Definition. A binary relation $\leq$ on a set P is a **partial ordering** if it is reflexive, transitive and antisymmetric, i.e., for all $x, y, z \in P$,

$$
\begin{aligned}
&x \leq x &&\text{(reflexivity),} \\
&x \leq y \,\&\, y \leq z \Longrightarrow x \leq z &&\text{(transitivity),} \\
&x \leq y \,\&\, y \leq x \Longrightarrow x = y, &&\text{(antisymmetry).}
\end{aligned}
$$

In connection with partial orderings we will also use the notation

$$x < y \iff_{\mathrm{df}} x \leq y \,\&\, x \neq y.$$

The partial ordering $\leq$ is **total**, or **linear**, or simply an **ordering**, if, in addition, any two elements of P are **comparable** in $\leq$, i.e.,

$$(\forall x, y \in P)[x \leq y \vee y \leq x],$$

or equivalently

$$(\forall x, y \in P)[x < y \vee x = y \vee y < x].$$

5.20. Definition. The binary relation $\leq$ on P is a **wellordering** of P, if it is a total ordering of P and, in addition, every non-empty subset of P has a least element,

$$(\forall X \subseteq P)[X \neq \emptyset \Longrightarrow (\exists x \in X)(\forall y \in X)[x \leq y]].$$

Correct English would have us call these "good orderings", and in fact this is what they are called in every other language, but the awkward "wellordering" has been established so firmly that it is hopeless to try and change it.

5.21. Definition. The order relation $\leq$ on the natural numbers is defined by the equivalence

$$n \leq m \iff_{\mathrm{df}} (\exists s)[n + s = m].$$

The most basic property of $\leq$ is:

5.22. Lemma. *For all natural numbers n, m,*

$$n \leq Sm \iff n \leq m \vee n = Sm.$$

PROOF. If $n \leq Sm$, then there is some t such that $n + t = Sm$ by the definition, and we consider two cases. CASE (1), $t = 0$. Now $n + 0 = Sm$, hence $n = Sm$. CASE (2), $n + t = Sm$ for some $t \neq 0$. Now, by Lemma **5.2**, $t = Ss$ for some s, so that $n + Ss = Sm$, hence $S(n + s) = Sm$, hence $n + s = m$ because S is an injection and hence $n \leq m$. The converse direction of the Lemma is easier.   ⊣

5.23. Theorem. *The relation $\leq$ on the natural numbers is a wellordering.*

PROOF. Reflexivity is immediate from $n + 0 = n$ and transitivity holds because $n + s = m \,\&\, m + t = k \Longrightarrow n + (s + t) = k$. To prove antisymmetry, notice that if $n + s = m$ and $m + t = n$, then $n + (s + t) = n$ and by Exercise **5.18** we have $s + t = 0$; this implies $t = 0$ (otherwise $s + t$ is a successor) and hence $m = n$.

Proof of linearity. We show that $(\forall n)[n \leq m \vee m \leq n]$, by induction on m. Notice first that for every n, $n \leq Sn$, because $n + S0 = Sn$.

BASIS. For every n, $0 + n = n$ and hence $0 \leq n$.

INDUCTION STEP. We assume the induction hypothesis

$$(\forall n)[n \leq m \vee m \leq n]$$

and show that for each n, $n \leq Sm \vee Sm \leq n$. The induction hypotheses naturally splits the proof up in two cases. If $n \leq m$, then $n \leq Sm$ because $m \leq Sm$ and $\leq$ is transitive. If $m \leq n$, then for some t, $m + t = n$, and again we have two cases: if $t = 0$, then $n = m \leq Sm$, and if $t \neq 0$, then $t = Ss$ for some s, so $m + Ss = n$ and from Lemma **5.16** $Sm + s = n$, hence $Sm \leq n$.

Proof of the wellordering property. Towards a contradiction, suppose that X is non-empty but has no least element and let

$$Y = \{n \in \mathbb{N} \mid (\forall m \leq n)[m \notin X]\},$$

so that obviously

$$Y \cap X = \emptyset. \tag{5-9}$$

It is enough to show that $0 \in Y$ and $n \in Y \Longrightarrow Sn \in Y$, because then $Y = \mathbb{N}$ by the Induction Principle and hence $X = \emptyset$ by (5-9), which is a contradiction.

BASIS. $0 \in Y$. Since 0 is the least number, we must have $0 \notin X$ (otherwise X would have a least member) and also $m \leq 0 \Longrightarrow m = 0 \Longrightarrow m \notin X$, so $0 \in Y$.

INDUCTION STEP. The induction hypothesis $n \in Y$ and the definition of Y imply that $(\forall m \leq n)m \notin X$, and then we know from Lemma **5.22** that $m \leq Sn \Longrightarrow m \leq n \vee m = Sn$. Hence to verify that $Sn \in Y$, it is enough to show that $Sn \notin X$. But if Sn were a member of X, then it would be the least member of X since

$$m < Sn \iff m \leq n \quad \text{by Lemma } \mathbf{5.22}$$
$$\Longrightarrow m \notin X \quad \text{by the ind. hyp.}$$

This shows that $\leq$ has the wellordering property and completes the proof of the theorem.   ⊣

About wellorderings, in general, we will say a lot in Chapter **7**. In the special case of the natural numbers, the fact that $\mathbb{N}$ is well ordered by $\leq$ is another manifestation of the Induction Principle.

Before we begin studying the applications of recursion to the theory of finite and countable sets, we should recall the warning issued in **3.24**: some of them require the Axiom of Choice and we will not be able to justify them axiomatically until we add that axiom to our system in Chapter **8**. Most, however, can be established by judicious applications of the general method of proof which can be symbolized by the coupling

$$\text{recursive definition} \; - \; \text{inductive proof}.$$

First we repeat some of the definitions of Chapter **2**, with the axiomatic notions now at our disposal.

5.24. Definition. For any two natural numbers $n \leq m$, the (half-open) **interval** from n to m is the set

$$[n, m) =_{\mathrm{df}} \{k \in \mathbb{N} \mid n \leq k \,\&\, k < m\}.$$

5.25. Exercise. *For each n, $[n, n) = \emptyset$, and for all $n \leq m$,*

$$[n, Sm) = [n, m) \cup \{m\}.$$

5.26. Definition. A set A is **finite** if there exists some natural number n such that $A =_c [0, n)$; **infinite** if it is not finite; and **countable** if it is finite or equinumerous with $\mathbb{N}$. By Proposition **2.7** (which follows easily from the axioms),

$$A \text{ is countable} \iff A \leq_c \mathbb{N}.$$

The **finite cardinals** are the cardinal numbers of finite sets.

The next crucial property of finite sets is the first, basic result in the field of *combinatorics*.

5.27. Pigeonhole Principle. *Every injection $f : A \rightarrowtail A$ of a finite set into itself is also a surjection, i.e., $f[A] = A$.*

PROOF. It is enough to prove that for each natural number m and each g,

$$g : [0, m) \rightarrowtail [0, m) \Longrightarrow g[[0, m)] = [0, m), \qquad (5\text{-}10)$$

for the following reason: if $f : A \rightarrowtail A$ is an injection and $\pi : A \rightarrowtail\!\!\!\rightarrow [0, m)$ witnesses that A is finite, we define $g : [0, m) \to [0, m)$ by the equation

$$g(i) = \pi(f(\pi^{-1}(i))) \quad (i < m),$$

so that (easily)

$$f(x) = \pi^{-1}(g(\pi(x))) \quad (x \in A). \qquad (5\text{-}11)$$

Now g is an injection, as a composition of injections, so that from (5-10) it is a bijection; but then f is also a bijection, because it is a composition of bijections, by (5-11).

$$0 \quad 1 \quad 2 \quad \cdots \qquad\qquad u \quad \cdots \quad m-1 \;\; m$$

$$0 \quad 1 \quad 2 \quad \cdots \qquad v \qquad \cdots \qquad m-1 \;\; m$$

g
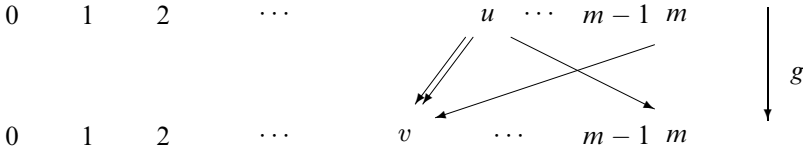
FIGURE **5.3**. CASE (3) in the Pigeonhole Principle proof.

The proof of (5-10) is (naturally) by induction on m. It is important to note at the outset that what we will show is the general assertion

$$(\forall g) \left[g : [0, m) \rightarrowtail [0, m) \Longrightarrow g[[0, m)] = [0, m) \right], \qquad (5\text{-}12)$$

because in the verification of the induction step for some g we will need the induction hypothesis for various other g's.

BASIS. (5-10) is trivial when $m = 0, 1$, because there is only one function $g : [0, m) \to [0, m)$ in these cases and it is a bijection.

INDUCTION STEP. We assume (5-12) for some $m > 1$ and we proceed to prove that every injection

$$g : [0, Sm) \rightarrowtail [0, Sm)$$

is a surjection. From Exercise **5.25** we know that

$$[0, Sm) = [0, m) \cup \{m\},$$

and the proof naturally splits into three cases.

CASE (1). $m \notin \text{Image}(g)$. Consider the *restriction* h of g to the interval $[0, m)$, which is defined by

$$h(i) = g(i) \quad (i < m),$$

i.e., as a set of pairs, $h = g \setminus \{(m, g(m))\}$. This takes all its values in $[0, m)$ and it is certainly an injection, so the induction hypothesis holds for it and hence $h[[0, m)] = [0, m)$. This means that $g[[0, m)] = [0, m)$, which is absurd, because the Case Hypothesis implies $g(m) < m$ so that the value $g(m)$ is taken on twice and g is not an injection.

CASE (2). $g(m) = m$. By the same reasoning, the restriction h is a bijection $h : [0, m) \rightarrowtail [0, m)$, and hence (trivially now) g is also a bijection.

CASE (3). There exist numbers $u, v < m$ such that

$$g(u) = m, \quad g(m) = v.$$

In this most interesting case, we define the function $h' : [0, m) \to [0, m)$ by the formula

$$h'(i) = \begin{cases} g(i), & \text{if } i < m \,\&\, i \neq u, \\ v, & \text{if } i = u. \end{cases}$$

Now h' is an injection, because it agrees with g at all arguments except u, where it takes the value v; and $v \neq g(j)$, for every $j < m$, because $g(m) = v$ and g is an injection. The induction hypothesis applied to h' implies that $h'[[0, m)] = [0, m)$, and using this (easily), $g[[0, Sm)] = [0, Sm)$. $\dashv$

As a first application we can give a rigorous proof of the following "obvious" result.

5.28. Corollary. *The set $\mathbb{N}$ of natural numbers is infinite.*

Proof. The function $(n \mapsto Sn)$ is an injection of $\mathbb{N}$ into $\mathbb{N} \setminus \{0\}$.   ⊣

It follows that "infinite, countable" means precisely "equinumerous with $\mathbb{N}$", in accordance with our basic intuitions: a set A is countable, infinite just when $|A| =_c \aleph_0$.

5.29. Corollary. *For each finite set A, there exists exactly one natural number n such that $A =_c [0, n)$. We let*

$$\#(A) =_{\text{df}} \text{ the unique } n \in \mathbb{N} \text{ such that } [A =_c |A| =_c [0, n)] \qquad (5\text{-}13)$$

and we naturally call $\#(A)$ the number of elements of A.

Proof. If $A =_c [0, n)$ and $A =_c [0, m)$ with $n < m$, then $[0, n) =_c [0, m)$ and the correspondence $\pi : [0, m) \rightarrowtail [0, n)$ contradicts the Pigeonhole Principle, since $[0, n)$ is a proper subset of $[0, m)$.   ⊣

From this point on we can proceed to prove all the basic properties of finite sets by induction on the number of elements in them, which is essentially their cardinal number. The method is illustrated in the problems.

5.30. Strings. In Chapter **2** we used the n-fold Cartesian product A^n to represent sequences of length n from a given set A. This is not convenient when we wish to study the set of all finite sequences from A, and it is better to represent these using functions on initial segments of $\mathbb{N}$. For each set A, we define the set of **finite sequences**, **words** or **strings** from A by

$$\begin{aligned} A^{(n)} &=_{\text{df}} \{u \subseteq \mathbb{N} \times A \mid \text{Function}(u) \,\&\, \text{Domain}(u) = [0, n)\}, \\ A^* &=_{\text{df}} \bigcup_{n=0}^{\infty} A^{(n)}, \end{aligned} \qquad (5\text{-}14)$$

and we let

$$\text{lh}(u) =_{\text{df}} \max\{i \mid i = 0 \lor i - 1 \in \text{Domain}(u)\} \quad (u \in A^*); \qquad (5\text{-}15)$$

this is the **length** of the string u, so that $\text{lh}(u) = 0$ exactly when u is the empty string $\emptyset$. We also let

$$u \sqsubseteq v \iff_{\text{df}} u \subseteq v \quad (u, v \in A^*), \qquad (5\text{-}16)$$

and we call u an **initial segment** of v if $u \sqsubseteq v$. If $a_0, \dots, a_{n-1}$ are elements of A, we let

$$\langle a_0, \dots, a_{n-1} \rangle =_{\text{df}} \{(0, a_0), \dots, (n - 1, a_{n-1})\} \qquad (5\text{-}17)$$

be the sequence of these objects and, in particular (with $n = 0, 1$),

$$\langle \, \rangle = \emptyset, \quad \langle a \rangle =_{\text{df}} \{(0, a)\}. \qquad (5\text{-}18)$$

For any two strings u, v, the string

$$u \star v = \langle u(0), \dots, u(\text{lh}(u) - 1), v(0), \dots, v(\text{lh}(v) - 1) \rangle \qquad (5\text{-}19)$$

is the **concatenation** of the strings u and v.

For each $f : \mathbb{N} \to A$ and each natural number n,

$$\overline{f}(n) =_{\mathrm{df}} f \upharpoonright [0, n) = \{(i, f(i)) \mid i < n\} \qquad (5\text{-}20)$$

is the **restriction** of f to the initial segment $[0, n)$ of $\mathbb{N}$. For example,

$$\overline{f}(0) = \emptyset, \quad \overline{f}(1) = \{(0, f(0))\}, \ldots,$$

and we can recover f from $\overline{f}$, since

$$i < n \Longrightarrow f(i) = \overline{f}(n)(i).$$

5.31. Definition. For each cardinal number κ and each $n \in \mathbb{N}$, we set

$$\kappa^n =_{\mathrm{df}} |\kappa^{(n)}|.$$

5.32. Proposition. *For each countably infinite set A and each $n > 0$,*

$$A =_c A \times A =_c A^{(n)} =_c A^*.$$

As equations of cardinal arithmetic, these read:

$$\aleph_0 =_c \aleph_0 \cdot \aleph_0 =_c \aleph_0{}^n =_c |\aleph_0{}^*|. \qquad (5\text{-}21)$$

PROOF. The inequalities from left to right are trivial, so by the Schröder-Bernstein Theorem it is enough to show $\mathbb{N}^* \leq_c \mathbb{N}$. We need to start with some injection

$$\rho : \mathbb{N} \times \mathbb{N} \rightarrowtail \mathbb{N}, \qquad (5\text{-}22)$$

and let us first suppose that we have one. Using it, we define by recursion an injection $\pi_n : \mathbb{N}^{(n+1)} \rightarrowtail \mathbb{N}$, for each n, so that

$$\pi_0(u) = u(0),$$
$$\pi_{n+1}(u) = \rho(\pi_n(u \upharpoonright [0, n + 1)), u(n + 1));$$

in full detail (for the last time), this comes from the Recursion Theorem, by setting

$$\pi_0 = \{(u, w) \mid u \in \mathbb{N}^{(1)}, (0, w) \in u\},$$
$$\pi_{n+1} = \{(u, w) \mid u \in \mathbb{N}^{(n+2)}, w = \rho(\pi_n(u \upharpoonright [0, n + 1), u(n + 1))\}.$$

Finally, the function

$$\pi(u) = (\mathrm{lh}(u) - 1, \pi_{\mathrm{lh}(u)-1}(u))$$

proves that $\bigcup_{n=0}^{\infty} \mathbb{N}^{(n+1)} \leq_c \mathbb{N} \times \mathbb{N}$, from which the full result follows immediately by using ρ once more. As far as choosing a ρ to start with, everyone has their favorite way of coding pairs and Cantor's illustrated in Figure **2.2** will certainly do. Here is another one, due to Gödel and pictured in Figure **5.4**:

$$\rho(m, n) = \begin{cases} (m + 1)^2 - 1, & \text{if } m = n, \\ n^2 + m, & \text{if } m < n, \\ m^2 + m + n, & \text{if } n < m. \end{cases}$$

The proof that it actually works is fun (Problem **x5.24**). $\dashv$

3	9	10	11	
2	4	5	8	
1	1	3	7	
0	0	2	6	12
	0	1	2	3

FIGURE **5.4**. A pairing of $\mathbb{N} \times \mathbb{N}$ with $\mathbb{N}$.

It is sometimes useful to think of A^* as a generalization of $\mathbb{N}$, generated from the empty string $\langle\,\rangle$ (instead of 0) by iterating the **appending operations**

$$S_a(u) = u \star \langle a \rangle \quad (a \in A),$$

one for each $a \in A$. This picture motivates the following, simple but useful method of definition by recursion on A^*.

5.33. String Recursion Theorem. *For any two sets A, E, each $a \in E$ and each function $h : E \times A \to E$, there is exactly one function $f : A^* \to E$ which satisfies the identities*

$$f(\langle\,\rangle) = a,$$
$$f(u \star \langle x \rangle) = h(f(u), x) \quad (u \in A^*, x \in A).$$

Similarly, with parameters, for any

$$g : Y \to E, \quad h : E \times A \times Y \to E,$$

there exists exactly one function $f : A^* \times Y \to E$ which satisfies the identities

$$f(\langle\,\rangle, y) = g(y) \qquad\qquad (y \in Y),$$
$$f(u \star \langle x \rangle, y) = h(f(u, y), x, y) \qquad (u \in A^*, y \in Y, x \in A).$$

PROOF. For the version with parameters, we appeal to Corollary **5.12** to obtain a function $\phi : \mathbb{N} \times A^* \times Y \to E$ which satisfies the identities

$$\phi(0, u, y) = g(y),$$
$$\phi(n + 1, u, y) = h(\phi(n, u, y), u(n), y),$$

and we set

$$f(u, y) = \phi(\mathrm{lh}(u), u, y).$$

Clearly $f(\langle\,\rangle, y) = \phi(0, u, y) = g(y)$, so $f(u, y)$ satisfies the first of the two required identities. To see that it also satisfies the second, we show by an easy

induction on n that for any two strings u, v,

$$(\forall i < n)[u(i) = v(i)] \Longrightarrow \phi(n, u, y) = \phi(n, v, y); \qquad (*)$$

this is immediate at the basis since $\phi(0, u, y) = g(y)$ does not depend on u, and in the induction step, assuming that $(\forall i < n + 1)[u(i) = v(i)]$,

$$\begin{aligned}
\phi(n + 1, u, y) &= h(\phi(n, u, y), u(n), y) \\
&= h(\phi(n, v, y), v(n), y) \quad \text{(ind. hyp. and hyp.)} \\
&= \phi(n + 1, v, y).
\end{aligned}$$

Now, using $(*)$, we compute, for any u with $\mathrm{lh}(u) = n$ and any $x \in A$,

$$\begin{aligned}
f(u \star \langle x \rangle, y) &= \phi(n + 1, u \star \langle x \rangle, y) \\
&= h(\phi(n, u \star \langle x \rangle, y), x, y) \\
&= h(\phi(n, u, y), x, y) \qquad \text{(by } (*)) \\
&= h(f(u, y), x, y) \qquad\quad \text{(by the def. of } f)
\end{aligned}$$

The uniqueness of $f(u, y)$ is easy, by induction on $\mathrm{lh}(u)$. $\dashv$

5.34. The continuum. The classical notation for the cardinal of $\mathcal{P}(\mathbb{N})$ uses the German (Fraktur) letter 'c',

$$\mathfrak{c} =_{\mathrm{df}} |\mathcal{P}(\mathbb{N})| =_c 2^{\aleph_0}. \qquad (5\text{-}23)$$

It is quite easy to establish the basic facts about $\mathfrak{c}$ using the properties of $\aleph_0$ in (5-21) and elementary cardinal arithmetic. For example:

$$\mathfrak{c} \cdot \mathfrak{c} =_c 2^{\aleph_0} \cdot 2^{\aleph_0} =_c 2^{\aleph_0 + \aleph_0} =_c 2^{\aleph_0} =_c \mathfrak{c}.$$

The Schröder-Bernstein Theorem is also very useful: for example

$$\mathfrak{c} =_c 2^{\aleph_0} \leq_c \aleph_0^{\aleph_0} \leq_c \mathfrak{c}^{\aleph_0} =_c (2^{\aleph_0})^{\aleph_0} =_c 2^{\aleph_0 \cdot \aleph_0} =_c 2^{\aleph_0} =_c \mathfrak{c},$$

which by Schröder-Bernstein implies that

$$\mathfrak{c} =_c \aleph_0^{\aleph_0} =_c \mathfrak{c}^{\aleph_0}.$$

Some of the problems ask for computations of this type. On the other hand, the equinumerosity $\mathbb{R} =_c \mathcal{P}(\mathbb{N})$ will follow easily from the axioms once we have defined the real numbers $\mathbb{R}$ in Appendix **A**, so that the Continuum Hypothesis is equivalent to the proposition

(**CH**) $(\forall \kappa \leq_c \mathfrak{c})[\kappa \leq_c \aleph_0 \vee \kappa =_c \mathfrak{c}].$

We will discuss **CH** extensively in Chapter **10**, it is not that easy to resolve.

Problems for Chapter 5

x5.1. Multiplication on the natural numbers is associative.

x5.2. Multiplication on the natural numbers is commutative.

x5.3. Exponentiation on the natural numbers is defined by the following recursion on m:

$$n^0 = 1,$$
$$n^{Sm} = n^m \cdot n.$$

Show that it satisfies the following identities (for $n \neq 0$):

$$n^{(m+k)} = n^m \cdot n^k,$$
$$n^{(m \cdot k)} = (n^m)^k.$$

x5.4. Suppose $(\mathbb{N}_1, 0_1, S_1)$ and $(\mathbb{N}_2, 0_2, S_2)$ are two Peano systems, $+_1$, $\cdot_1$, $+_2$, $\cdot_2$ are the functions of addition and multiplication in these systems, and $\pi : \mathbb{N}_1 \rightarrowtail\!\!\!\rightarrow \mathbb{N}_2$ is the "canonical" isomorphism between them defined in the proof of Theorem **5.4**. Show that π is an isomorphism with respect to addition and multiplication also, i.e., for all $n, m \in \mathbb{N}_1$,

$$\pi(n +_1 m) = \pi(n) +_2 \pi(m), \quad \pi(n \cdot_1 m) = \pi(n) \cdot_2 \pi(m).$$

x5.5. Suppose $(\mathbb{N}_1, 0_1, S_1)$ and $(\mathbb{N}_2, 0_2, S_2)$ are two Peano systems, $\leq_1, \leq_2$ are the respective wellorderings and $\pi : \mathbb{N}_1 \rightarrowtail\!\!\!\rightarrow \mathbb{N}_2$ is the canonical isomorphism. Show that π is order preserving, i.e., for all $n, m \in \mathbb{N}_1$,

$$n \leq_1 m \iff \pi(n) \leq_2 \pi(m).$$

x5.6. Every subset B of an interval $[0, n)$ is equinumerous with some $[0, m)$, where $m \leq n$. It follows that if A is finite and $B \subseteq A$, then B is finite and $\#(B) \leq \#(A)$.

Every cardinal number is a set; a **finite cardinal** κ is a cardinal number which is a finite set.

x5.7. Prove that for every finite cardinal number κ,

$$\kappa =_c [0, \#(\kappa)).$$

x5.8. Show that for all n, m, $[0, m) =_c [n, n + m)$ and infer that the union of two finite sets A, B is finite and such that

$$\text{if } A \cap B = \emptyset, \text{ then } \#(A \cup B) = \#(A) + \#(B).$$

It follows that for any two finite cardinals κ, λ,

$$\#(\kappa + \lambda) = \#(\kappa) + \#(\lambda).$$

x5.9. If $\mathscr{E}$ is a finite set and every member of it is a finite set, then the unionset $\bigcup \mathscr{E}$ is also finite.

x5.10. The product of two finite sets A, B is finite and such that

$$\#(A \times B) = \#(A) \cdot \#(B).$$

It follows that for any two finite cardinals κ, λ,

$$\#(\kappa \cdot \lambda) = \#(\kappa) \cdot \#(\lambda).$$

x5.11. The powerset of every finite set A is finite and

$$\#(\mathcal{P}(A)) = 2^{\#(A)}.$$

It follows that for every finite cardinal κ,

$$\#(2^\kappa) = 2^{\#(\kappa)}.$$

x5.12. For all finite cardinals κ, λ,

$$\#(\kappa^\lambda) = \#(\kappa)^{\#(\lambda)}.$$

x5.13. Show that if A is finite, then every surjection $f : A \twoheadrightarrow A$ is an injection. (This is an alternative version of the Pigeonhole Principle.)

x5.14. For all cardinals κ, $2^\kappa \neq_c \aleph_0$. (*Careful*: we have not proved the Comparability Hypothesis **3.1**, and so you cannot appeal to it.)

x5.15. $\mathfrak{c} + \mathfrak{c} =_c \aleph_0 \cdot \mathfrak{c} =_c \mathfrak{c} \cdot \mathfrak{c} =_c \mathfrak{c}$.

x5.16. $\mathfrak{c}^{\mathfrak{c}} =_c 2^{\mathfrak{c}}$.

x5.17. For every cardinal number $\kappa >_c 1$, if $\kappa \cdot \kappa =_c \kappa$, then $2^\kappa =_c \kappa^\kappa$.

x5.18. For each cardinal number κ and each $n \in \mathbb{N}$,

$$\kappa^n =_c \kappa^{|[0,n)|},$$

where the left side is defined by **5.31** and the right side is cardinal exponentiation.

x5.19. For each $n \neq 0$, $\mathfrak{c}^n =_c |\mathfrak{c}^*| =_c \mathfrak{c}$.

x5.20 (Simultaneous Recursion Theorem). For any two sets E_1, E_2, elements $a_1 \in E_1$, $a_2 \in E_2$ and functions $h_1 : E_1 \times E_2 \to E_1$, $h_2 : E_1 \times E_2 \to E_2$, there exist unique functions

$$f_1 : \mathbb{N} \to E_1, \quad f_2 : \mathbb{N} \to E_2$$

which satisfy the identities

$$\begin{aligned} f_1(0) &= a_1, & f_2(0) &= a_2, \\ f_1(n+1) &= h_1(f_1(n), f_2(n)), & f_2(n+1) &= h_2(f_1(n), f_2(n)). \end{aligned}$$

****x5.21** (Nested Recursion Theorem). Prove that for any three functions

$$g : Y \to E, \ h : E \times \mathbb{N} \times Y \to E, \text{ and } \pi : \mathbb{N} \times Y \to Y.$$

there is exactly one function $f : \mathbb{N} \times Y \to E$ which satisfies the identities

$$f(0, y) = g(y), \quad f(Sn, y) = h(f(n, \pi(n, y)), n, y).$$

HINT. Define recursively a function $\phi : \mathbb{N} \to (\mathbb{N} \times (\mathbb{N} \to Y))$ such that the required function is $f(n, y) = \text{Second}(\phi(n))(y)$.

x5.22. Give a recursive definition of the *factorial*

$$f(n) = 1 \cdot 2 \cdots (n-1) \cdot n \quad \text{(with } f(0) = 1, \text{ conventionally)}.$$

x5.23. Give an explicit definition of the function $f : A^* \to A^*$ on the strings from some set A which is defined by the string recursion

$$f(\langle\,\rangle) = \langle\,\rangle, \quad f(u \star \langle x\rangle) = \langle x\rangle \star f(u).$$

x5.24. The function ρ in the proof of **5.32** is a bijection.

***x5.25.** Every partial ordering $\leq$ on a finite set P has a **linearization**, i.e., some linear ordering $\leq'$ of P exists such that $x \leq y \Longrightarrow x \leq' y$.

***x5.26. The marriage problem.** Suppose B is a finite set and $h : B \to \mathcal{P}(G)$ is a function, such that for each $x \in B$, $h(x)$ is a finite subset of G and

$$X \subseteq B \Longrightarrow \#(X) \leq \#(\bigcup\{h(x) \mid x \in X\}), \tag{5-24}$$

so in particular each $h(x) \neq \emptyset$. Prove that there exists an injection $f : B \rightarrowtail G$ such that

$$(\forall x \in B)[f(x) \in h(x)]. \tag{5-25}$$

Show also that the hypothesis (5-24) is necessary for the existence of some injection f which satisfies (5-25). Hint. Take cases, on whether or not there exists some $\emptyset \neq C \subsetneq B$ such that $\#(C) = \#(\bigcup\{h(x) \mid x \in C\})$. The name of the problem comes from the traditional interpretation, in which B is a set of boys, G is a set of available girls and h assigns to each boy x the (finite) set $h(x)$ of girls that he would be willing to marry. There are many other applications of the problem, more useful and less sexist, for example when B is a set of courses, G is a set of professors and h assigns to each course the set of professors who can teach it ("the scheduling problem").

The next problem justifies one more form of recursive definition which is often useful in applications.

x5.27. Complete Recursion. For each $h : E^* \to E$, there exists exactly one function $f : \mathbb{N} \to E$ which satisfies the identity

$$f(n) = h(\overline{f}(n));$$

similarly, for each $h : E^* \times \mathbb{N} \to E$, there exists exactly one function $f : \mathbb{N} \to E$ which satisfies the identity

$$f(n) = h(\overline{f}(n), n).$$

The next problem gives a characterization of countable, infinite sets directly in terms of the membership relation, with no appeal to the defined notions of $\mathbb{N}$ and "function".

x5.28. Prove the equivalence:

$$A =_c \mathbb{N} \iff (\exists \mathscr{E})[A = \bigcup\mathscr{E} \,\&\, \emptyset \in \mathscr{E} \,\&\, (\forall u \in \mathscr{E})(\exists! y \notin u)[u \cup \{y\} \in \mathscr{E}]$$
$$\&\, (\forall Z)[[\emptyset \in Z \,\&\, (\forall u \in Z)(\exists! y \notin u)[u \cup \{y\} \in Z \cap \mathscr{E}]] \Longrightarrow \mathscr{E} \subseteq Z]].$$

FIXED POINTS

The primary significance of the Recursion Theorem **5.6** is foundational, since the result justifies on the basis of the axioms a method of definition of functions which is intuitively obvious. From a purely mathematical point of view, however, we can also view **5.6** as a theorem of *existence and uniqueness* of solutions for systems of identities of the form

$$\begin{aligned} f(0) &= a, \\ f(Sx) &= h(f(x)) \quad (x \in \mathbb{N}), \end{aligned} \tag{6-1}$$

where $a \in E$ and $h : E \to E$ are given and the function $f : \mathbb{N} \to E$ is *the unknown*. In this chapter we will prove an elegant generalization of the Recursion Theorem in the context of the theory of partial orderings, which implies the existence and uniqueness of solutions for systems of functional identities much more general than (6-1). The Continuous Least Fixed Point Theorem **6.21** is fundamental for the *theory of computation*, it is the basic mathematical tool of the so-called *fixpoint theory of programs*. In the next chapter we will show that it is a special case of a much deeper Fixed Point Theorem of Zermelo, which is intimately related to the theory of wellorderings and rich in set theoretic consequences: for example it implies directly the Hypothesis of Cardinal Comparability, **3.1**. Thus, in addition to its purely mathematical significance, the Continuous Least Fixed Point Theorem yields also an interesting point of contact between classical set theory and today's theoretical computer science.

In their simplest and most natural expressions, the Fixed Point Theorems are somewhat abstract and apparently unrelated to the solution of systems of functional identities to which we intend to apply them. To understand what they say and how to use them, we will need to introduce first some basic notions from the theory of partial orderings.

6.1. A **partially ordered set** or simply **poset** is a structured set

$$P = (\text{Field}(P), \leq_P),$$

where $\text{Field}(P)$ is an arbitrary set and $\leq_P$ is a partial ordering on $\text{Field}(P)$, i.e., a reflexive, transitive and antisymmetric binary relation. Notice that $\leq_P$ determines P because it is reflexive,

$$x \in \text{Field}(P) \iff x \leq_P x,$$

Figure **6.1**. A discrete and a flat poset.

so we can specify a poset P completely by defining its partial ordering $\leq_P$. In practice, however, the partial ordering $\leq_P$ is often clear from the context and we will tend to identify a poset P with its field Field(P), following the general convention about structured sets discussed in **4.30**. For example, by *the poset* $\mathbb{N}$ we obviously mean the pair $(\mathbb{N}, \leq)$, where $\leq$ is the usual ordering on the set of natural numbers. By this convention, *the points of P* are the members of Field(P), *a subset $I \subseteq P$* is a subset $I \subseteq$ Field(P), etc. Each $I \subseteq P$ is a poset in its own right, partially ordered by the restriction of $\leq_P$ to I,

$$x \leq_I y \iff_{\text{df}} x \leq_P y \mathbin{\&} x \in I \mathbin{\&} y \in I, \tag{6-2}$$

which is (easily) a partial ordering. We will often deal with posets which have a least element, and it will be convenient to use the same, standard symbol $\perp$ (read "bottom" or "least") for it, just as we use the same symbol 0 for the additive unit of every number system:

$$\perp = \perp_P =_{\text{df}} \text{ the least element of } P \text{ (if it exists).} \tag{6-3}$$

Any set A can be viewed as a **discrete poset** in which no two elements are comparable, i.e., partially ordered by the identity relation

$$x \leq y \iff x = y \quad (x, y \in A).$$

Just above these in complexity are the **flat posets** which have a least element, the only element involved in any comparisons: i.e.,

$$x \leq_P y \iff x = \perp \vee x = y.$$

The simplest non-empty poset is a singleton $\{\perp\}$, which is both discrete and flat. Additional examples of posets are the sets of natural numbers $\mathbb{N}$, rationals $\mathbb{Q}$ and reals $\mathbb{R}$, with their usual orderings. These are all linear (totally ordered) posets. There is a large variety of finite posets and their study constitutes an important research area of mathematics, but we will not be much concerned with it here. Mostly we will use them as counterexamples. In drawing posets we indicate $x < y$ by placing y *above or to the right of x and drawing a polygonal line* (sometimes directed, to avoid ambiguity) *from x to y*, which may pass through other points, e.g., $c < e$ in Figure **6.2**.[13]

6.2. Definition. Let P be a poset, $S \subseteq P$ and $M \in P$ a member of P.

---

[13]There are those who draw posets growing *to the right*, those who draw them growing *up* and even those who draw them growing *down*; to the best of my knowledge, nobody pictures posets growing to the left and it does not appear that any of the three common choices is dominant.
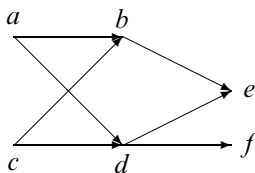
FIGURE **6.2**. A finite poset

1. M is an **upper bound** of S if it is greater than or equal to every element of S, i.e., if $(\forall x \in S)[x \leq M]$.
2. M is **maximum** in S if it is a member and an upper bound of S, i.e., if $M \in S \,\&\, (\forall x \in S)[x \leq M]$.
3. M is a **least upper bound** of S if it is an upper bound and also less than or equal to every other upper bound of S, i.e., if

$$(\forall x \in S)[x \leq M] \,\&\, (\forall M')[(\forall x \in S)[x \leq M'] \Longrightarrow M \leq M'].$$

If M_1, M_2 are both least upper bounds of S, then $M_1 \leq M_2$ (because M_2 is an upper bound and M_1 is a least upper bound) and symmetrically $M_2 \leq M_1$, so that $M_1 = M_2$, i.e., S has at most one least upper bound. When it exists, the least upper bound of a set S is denoted by

$$\sup S = \text{ the least upper bound of } S. \tag{6-4}$$

The term "sup" from the Latin *supremum* (maximum) is justified by the following observation.

6.3. Exercise. *If M is maximum of a set S in a poset P, then M is also the least upper bound of S.*

6.4. Exercise. *In the poset of Figure **6.2**, find subsets S with the following properties*: (1) *S has no upper bound.* (2) *S has upper bounds but no least upper bound.* (3) *S has a least upper bound but no maximum element.*

6.5. Exercise. *In any poset P, an element M is the least upper bound of the empty set $\emptyset$ if and only if M is the least element of P, i.e.,*

$$\bot = \sup \emptyset \tag{6-5}$$

if $\bot$ or $\sup \emptyset$ exists.

6.6. Exercise. *The powerset $\mathcal{P}(A)$ of every set A is partially ordered by the relation*

$$X \subseteq_A Y \iff_{\mathrm{df}} X \subseteq Y \subseteq A,$$

so that $\bot = \emptyset$ and for every $S \subseteq \mathcal{P}(A)$, the union $\bigcup S$ is the least upper bound of S.[14]

---

[14]The partial ordering of $\mathcal{P}(A)$ is the restriction $\subseteq_A$ of the definite condition $X \subseteq Y$ to $\mathcal{P}(A)$, and we often skip the subscript in referring to it, cf. **4.8**.
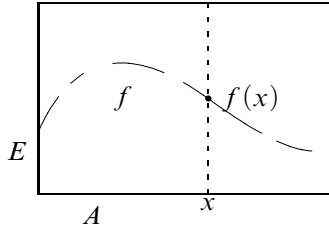
Figure **6.3**. A partial function $f : A \rightharpoonup E$.

Less trivial and more interesting for our purposes is the next example of a poset.

6.7. Definition. A **partial function** on a set A to a set E is any function with domain of definition some subset of A and values in E, in symbols

$$f : A \rightharpoonup E \iff_{\mathrm{df}} \mathrm{Function}(f) \,\&\, \mathrm{Domain}(f) \subseteq A \,\&\, \mathrm{Image}(f) \subseteq E. \quad (6\text{-}6)$$

For example, $(n \mapsto (n-1))$ is a partial function on the natural numbers defined only when $n \neq 0$, $(x \mapsto \sqrt{x})$ is a partial function on the reals with domain of definition $\{x \mid x \geq 0\}$, etc. A finite sequence $u \in A^*$ is a partial function $u : \mathbb{N} \rightharpoonup A$, as we defined it in **5.30**. The empty set $\emptyset$ is (trivially) a partial function (with empty domain of definition!) and every (total) function on A to E is also a partial function, since (6-6) does not exclude $\mathrm{Domain}(f) = A$,

$$\emptyset : A \rightharpoonup E, \quad f : A \to E \Longrightarrow f : A \rightharpoonup E.$$

We will use systematically the convenient half-arrow notation for partial functions (recently established in computer science), as well as the common notations

$$f(x){\downarrow} \iff_{\mathrm{df}} x \in \mathrm{Domain}(f), \quad f(x){\uparrow} \iff_{\mathrm{df}} x \notin \mathrm{Domain}(f) \quad (6\text{-}7)$$

to indicate that a partial function is defined or undefined at some point; we sometimes read $f(x){\downarrow}$ as $f(x)$ *converges* or f *converges at* x, and $f(x){\uparrow}$ as $f(x)$ *diverges* or f *diverges at* x.

6.8. Definition. For each A and E,

$$(A \rightharpoonup E) =_{\mathrm{df}} \{f \subseteq A \times E \mid f : A \rightharpoonup E\} \quad (6\text{-}8)$$

is the set of all partial functions from A to E, in analogy with the notation $(A \to E)$ for the set of all (total) functions from A to E, cf. **4.14**. The set $(A \rightharpoonup E)$ is partially ordered by the relation $\subseteq$,

$$f \subseteq g \iff (\forall x \in A)[f(x){\downarrow} \Longrightarrow [g(x){\downarrow} \,\&\, f(x) = g(x)]],$$

with least element $\bot = \emptyset$.

6.9. Exercise. *For each A, E,*

$$(A \to E) = \{f \restriction X \mid f : A \to E \,\&\, X \subseteq A\}.$$

Function restrictions are defined in (4-13).

It is harder to find least upper bounds for sets in these partial function posets than in powersets: if, for example, $S \subseteq (\mathbb{N} \to \mathbb{N})$ contains the two (total) constant functions $x \mapsto 0$ and $x \mapsto 1$, then any upper bound $f : \mathbb{N} \to \mathbb{N}$ of S would satisfy both $f(0) = 0$ and $f(0) = 1$, which is absurd. On the other hand, linearly ordered subsets of $(A \to E)$ have least upper bounds, and this is a fruitful property of these posets, worth a name.

6.10. Definition. A **chain** in a poset P is any linearly ordered subset S of P, i.e., a set satisfying

$$(\forall x, y \in S)[x \le y \vee y \le x].$$

A poset P is *chain-complete* or **inductive** if every chain in P has a least upper bound.

6.11. Exercise. *The empty set is (trivially) a chain, hence every inductive poset P has a least element $\bot = \sup \emptyset$.*

6.12. Exercise. *Every flat poset is inductive; a discrete poset is inductive only when it has exactly one element, in which case it is also flat.*

6.13. Exercise. *The image $\{x_n \mid n \in \mathbb{N}\}$ of a non-decreasing sequence*

$$x_0 \le x_1 \le x_2 \le \cdots$$

is a chain; thus, every non-decreasing sequence has a **limit** *in an inductive poset,*

$$\lim_n x_n =_{\mathrm{df}} \sup \{x_n \mid n \in \mathbb{N}\}. \tag{6-9}$$

6.14. Proposition. (1) *For each set A, the powerset $\mathcal{P}(A)$ is inductive.*

(2) *For any two sets A, E, the poset $(A \to E)$ of all partial functions from A to E is inductive.*

(3) *For every poset P, the set*

$$\mathrm{Chains}(P) = \{S \subseteq P \mid S \text{ is a chain}\}$$

of all chains in P (partially ordered under $\subseteq$) is inductive.

PROOF. (1) is immediate by Exercise **6.6**. (2) If $S \subseteq (A \to E)$ is a chain, then the union $\bigcup S$ is a partial function and obviously, $\bigcup S = \sup S$. (3) This is also proved by observing that the union of a chain of chains in a poset is also a chain. $\dashv$

6.15. Exercise. *Neither $\mathbb{N}$ (with the usual ordering) nor the finite poset of Figure **6.2** are inductive.*

6.16. Exercise. *For each set E, the set $P = E^* \cup (\mathbb{N} \to E)$ of finite and infinite sequences from E is an inductive poset, under $\subseteq$.*

6.17. Exercise. *For any two sets A, E, the poset*

$$(A \rightarrowtail E) =_{\mathrm{df}} \{f \in (A \to E) \mid f \text{ is one-to-one}\}$$

of **partial injections** *from A to E (partially ordered by $\subseteq$) is inductive.*

We will find the most significant applications of the fixed point theorems in partial function posets; but the proofs will use only the fact that these posets are inductive, and there are lots of other interesting examples. Some of them are described in the problems at the end of the chapter.

Finally, we need to delineate the type of functions on inductive posets which must, necessarily, have fixed points.

6.18. Definition. A mapping[15] $\pi : P \to Q$ on a poset P to another is **monotone** if for all $x, y \in P$,

$$x \leq_P y \Longrightarrow \pi(x) \leq_Q \pi(y).$$

A monotone mapping need not be strictly increasing in the sense of

$$x <_P y \Longrightarrow \pi(x) <_Q \pi(y),$$

e.g., every constant mapping is monotone.

Notice that if $\pi : P \to Q$ is monotone and $S \subseteq P$ is a chain, then the image $\pi[S]$ is also a chain; because given $x = \pi(u)$, $y = \pi(v)$ with $u, v \in S$, either $u \leq v$, which implies $x = \pi(u) \leq \pi(v) = y$, or $v \leq u$, which similarly implies $y \leq x$. This makes the next definition meaningful.

6.19. Definition. A monotone mapping $\pi : P \to Q$ on an inductive poset to another is **countably continuous** if for every non-empty, countable chain $S \subseteq P$,

$$\pi(\sup S) = \sup \pi[S].$$

6.20. Exercise. *A monotone mapping $\pi : P \to Q$ on one inductive poset to another is countably continuous if and only if for every non-decreasing sequence $x_0 \leq_P x_1 \leq_P \ldots$ of elements in P,*

$$\pi(\lim_n x_n) = \lim_n \pi(x_n).$$

Here the limit on the left is taken in P and the limit on the right is taken in Q.

6.21. Continuous Least Fixed Point Theorem. *Every countably continuous, monotone mapping $\pi : P \to P$ on an inductive poset into itself has exactly one* **strongly least fixed point** *x^*, which is characterized by the two properties*

$$\pi(x^*) = x^*, \tag{6-10}$$

$$(\forall y \in P)[\pi(y) \leq y \Longrightarrow x^* \leq y]. \tag{6-11}$$

PROOF. The **orbit** of the least element $\bot$ under π is defined by the simple recursion on the natural numbers,

$$x_0 = \bot,$$

$$x_{n+1} = \pi(x_n).$$

---

[15]It is convenient to refer to $\pi : P \to Q$ as a "mapping" rather than a "function" (which means the same thing), because in the interesting applications P is some partial function space $(A \rightharpoonup E)$, Q may be another partial function space, π takes partial functions as arguments and (possibly) values and there are altogether too many functions around. Notice also that, pedantically, $\pi : \text{Field}(P) \to \text{Field}(Q)$ is a mapping from the field of P to that of Q.
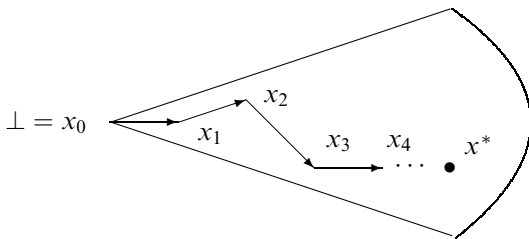
FIGURE **6.4**. The Continuous Least Fixed Point Theorem.

Clearly $x_0 = \bot \leq x_1$, and by a trivial induction (using the monotonicity of π), for every n, $x_n \leq x_{n+1}$. Thus, the limit

$$x^* =_{\mathrm{df}} \lim_n x_n = \sup \{x_n \mid n \in \mathbb{N}\} \qquad (6\text{-}12)$$

exists by **6.13**, and by the countable continuity of π,

$$\pi(x^*) = \pi(\lim_n x_n) = \lim_n \pi(x_n) = \lim_n x_{n+1} = x^*.$$

For the second claimed property of x^*, we assume $\pi(y) \leq y$ and show by induction that for every n, $x_n \leq y$. BASIS. $x_0 = \bot \leq y$, trivially. INDUCTION STEP. The Induction Hypothesis gives us $x_n \leq y$, and we compute:

$$\begin{aligned} x_n \leq y &\implies \pi(x_n) \leq \pi(y), &\text{(because } \pi \text{ is monotone)}, \\ &\implies x_{n+1} \leq \pi(y) \leq y, &\text{(by the assumption on } y). \end{aligned}$$

Thus, y is an upper bound of the set of values $\{x_n \mid n \in \mathbb{N}\}$ of the sequence, and hence, $x^* = \sup \{x_n \mid n \in \mathbb{N}\} \leq y$. ⊣

To apply the Continuous Least Fixed Point Theorem, we must formulate the problem at hand as a question of existence and (sometimes) uniqueness of solutions for an equation of the form $\pi(x) = x$, where $\pi : P \to P$ is monotone and countably continuous on some inductive poset P. This is typically the hardest part: to bring the problem in a form in which **6.21** can be applied. *Verification of the countable continuity of π is not necessary*: because we will show in the next chapter that **6.21** *remains true if we simply remove the hypothesis of countable continuity of π*. In any case, most applications involve simple monotone mappings on partial function posets for which it is often trivial to recognize a much stronger, natural continuity property.

6.22. Definition. A partial function $g : A \rightharpoonup E$ is **finite** if it has finite domain, i.e., if it is a finite set of ordered pairs.

A mapping $\pi : (A \rightharpoonup E) \to (B \rightharpoonup M)$ from one partial function space into another is **continuous**, if it is **monotone** and **compact**, i.e., for each $f : A \rightharpoonup E$, and all $y \in B$ and $v \in M$,

$$\pi(f)(y) = v \implies (\exists f_0 \subseteq f)[f_0 \text{ is finite } \& \pi(f_0)(y) = v]. \qquad (6\text{-}13)$$

The notation is a bit convoluted, but what it means is quite simple: to compute $\pi(f)(y)$, we first compute the partial function $f' = \pi(f)$ and then we evaluate it at y, $\pi(f)(y) = f'(y)$: a monotone mapping π is continuous if each value $\pi(f)(y)$ of $\pi(f)$, whenever defined, depends only on finitely many values of f. In fact, we can combine the conditions of monotonicity and compactness in the following, simple characterization of continuity for partial function space mappings which often allows the immediate—"by inspection"—recognition that they are continuous.

6.23. Proposition. *A mapping $\pi : (A \rightharpoonup E) \to (B \rightharpoonup M)$ is continuous if and only if it satisfies both* (6-13) *and its converse, i.e., if for every $f : A \rightharpoonup E$ and all $y \in B$ and $v \in M$:*

$$\pi(f)(y) = v \iff (\exists f_0 \subseteq f)[f_0 \text{ is finite } \& \ \pi(f_0)(y) = v]. \qquad (6\text{-}13^*)$$

For example, the mapping $\pi : (\mathbb{N} \rightharpoonup \mathbb{N}) \to (\mathbb{N} \rightharpoonup \mathbb{N})$ defined by

$$\pi(f) = (n \mapsto f(n) + f(n^2))$$

is continuous, because (obviously) for every f and n,

$$\pi(f)(n) = \pi(f_0)(n),$$

where f_0 is the restriction of f to the two-element set $\{n, n^2\}$.

Proof of **6.23**. Suppose first that π is continuous, $f : A \rightharpoonup E$, $y \in B$, and $v \in M$. If $\pi(f)(y) = b$, then by the compactness of π, there exists some finite $f_0 \subseteq f$ such that $\pi(f_0)(y) = v$; and if such a finite $f_0 \subseteq f$ exists, then the monotonicity of π implies that $\pi(f)(y) = v$.

For the converse, assume that $(6\text{-}13^*)$ holds for every $f : A \rightharpoonup E$ and all $y \in B, v \in M$. This gives immediately the compactness of π. To verify that π is monotone, suppose $f \subseteq g$ and $\pi(f)(y) = v$; by the $\implies$ direction of $(6\text{-}13^*)$, there is some finite $f_0 \subseteq f$ such that $\pi(f_0)(y) = v$; but now $f_0 \subseteq g$, and so by the $\impliedby$ direction of $(6\text{-}13^*)$, $\pi(g)(y) = v$. $\dashv$

6.24. Exercise. *Show that the mapping $\pi : (\mathbb{N} \rightharpoonup \mathbb{N}) \to (\mathbb{N} \rightharpoonup \mathbb{N})$ defined by*

$$\pi(f) = (n \mapsto \Sigma_{i=0}^{n} f(i))$$

is continuous. Compute $\pi(n \mapsto 2n)(2)$ for this π.

6.25. Definition. A function $f : X \to Y$ from one topological space to another is (topologically) **continuous**, if the inverse image $f^{-1}[G]$ of every open subset of Y is an open subset of X. The definition of a **topological space** was sneaked in **4.30**, as the first example of a structured set.

6.26. Exercise. *A function $f : X \to Y$ from one topological space to another is continuous if and only if the inverse image $f^{-1}[F]$ of every closed subset of Y is closed in X.*

One might guess that our using the term "continuous" in Definition **6.22** is not entirely accidental and that the notion of **6.22** has something to do with topological continuity. Indeed it does: the notions are equivalent when the

proper topology is put on partial function posets, but we will have no need of this fact and will leave it for Problem **x6.21**.

6.27. About topology. General (pointset) topology is to set theory like parsley to Greek food: some of it gets in almost every dish, but there are no great "parsley recipes" that the good Greek cook needs to know. Many notions and results of set theory are connected to topological ideas, but it is quite rare that you can prove an interesting theorem about sets by quoting some deep topological result. To avoid getting distracted with side issues, we will follow the general policy of giving the most direct, set theoretically natural definitions and proofs of the notions and results we need and leave the connections with topology for the problems. Occasionally the most natural approach is topological.

6.28. Lemma. *If $S \subseteq (A \rightharpoonup E)$ is a non-empty chain in a partial function poset and $f_0 \subseteq \sup S$ is a finite function, then there exists some $g \in S$ such that $f_0 \subseteq g$.*

PROOF is by induction on the number of elements in the domain of f_0.

BASIS. $f_0 = \emptyset$ is the partial function which is nowhere defined. There is some $g \in S$ since S is non-empty, and $\emptyset \subseteq g$.

INDUCTION STEP. The domain of f_0 has $n + 1$ elements, so

$$f_0 = f_1 \cup \{(x, w)\} \subseteq \sup S,$$

where f_1 is a finite, partial function with just n elements in its domain, and by induction hypothesis, there exists some $g_1 \in S$ such that $f_1 \subseteq g_1$. Since $(x, w) \in \sup S$, there must also exist some $h' \in S$ such that $(x, w) \in h'$, and since S is a chain, either $g_1 \subseteq h'$ or $h' \subseteq g_1$; the g we need is the larger of these two partial functions. ⊣

6.29. Lemma. *Every continuous mapping $\pi : (A \rightharpoonup E) \to (B \rightharpoonup M)$ is countably continuous, in fact, for every* (not necessarily countable) *non-empty chain $S \subseteq (A \rightharpoonup E)$,*

$$\pi(\sup S) = \sup \pi[S].$$

PROOF. Suppose $S \subseteq (A \rightharpoonup E)$ is a non-empty chain with union $f = \sup S$. If $g \in S$, then $g \leq f$, and since π is monotone we have $\pi(g) \leq \pi(f)$, so that

$$\sup \pi[S] = \sup \{\pi(g) \mid g \in S\} \leq \pi(f).$$

For the converse inequality, we need to show that if $\pi(f)(y) = v$, then there exists some $g \in S$ such that $\pi(g)(y) = v$. By the continuity of π, there exists a finite $f_0 \subseteq f$, such that already $\pi(f_0)(y) = v$; by the preceding Lemma, there exists some $g \in S$ so that $f_0 \subseteq g$; and by the monotonicity of π, this implies $\pi(f_0) \subseteq \pi(g)$. In particular, since $\pi(f_0)(y) = v$, we have $\pi(g)(y) = v$, so this is the g we need. ⊣

The Continuous Least Fixed Point Theorem is evidently a simple corollary of the Recursion Theorem on the natural numbers **5.6**. In fact, it implies **5.6**

by a fairly direct argument, which is worth looking at, as it illustrates how we intend to apply **6.21**.

6.30. Proof of the Recursion Theorem from 6.21. For each given $a \in E$ and function $h : E \to E$, we define the mapping

$$\pi : (\mathbb{N} \rightharpoonup E) \to (\mathbb{N} \rightharpoonup E)$$

by the formula

$$\pi(f) = f', \text{ where } f'(x) = \begin{cases} a, & \text{if } x = 0, \\ h(f(x-1)), & \text{if } x > 0, \end{cases}$$

where f is any partial function from $\mathbb{N}$ to E and we understand the definition naturally, so that

$$x > 0 \Longrightarrow [f'(x)\downarrow \iff h(f(x-1))\downarrow \iff f(x-1)\downarrow].$$

Written out in detail, the mapping π associates a set of pairs $f' \subseteq (\mathbb{N} \times E)$ with every $f \in (\mathbb{N} \rightharpoonup E)$ and it is defined by the equation

$$\begin{aligned} \pi(f) = &\{(0, a)\} \\ &\cup \{(x, h(w)) \mid x > 0 \,\&\, (x-1, w) \in f\} \quad (f : \mathbb{N} \rightharpoonup E). \quad (6\text{-}14) \end{aligned}$$

From this we get that for every f and x,

$$\pi(f)(x) = \pi(f_0)(x),$$

where $f_0 = \{(0, a)\}$ if $x = 0$ and $f_0 = \{(x-1, f(x-1))\}$ if $x > 0$, so that π is continuous by Proposition **6.23**, and hence countably continuous. Thus by **6.21**, it has a fixed point: that is, some partial function $f^* : \mathbb{N} \rightharpoonup E$ exists which satisfies $f^* = \pi(f^*)$, so that, immediately,

$$f^*(0) = a, \qquad\qquad\qquad\qquad\qquad (6\text{-}15)$$
$$f^*(x+1) = h(f^*(x)) \quad (f^*(x)\downarrow). \qquad\qquad (6\text{-}16)$$

Theorem **6.21** does not guarantee that this f^* is a *total function*, with domain of definition the entire $\mathbb{N}$, but this can be verified by an easy induction on x using the identities (6-15) and (6-16). $\dashv$

Consider next a case where it is not quite so obvious how to define the function we want directly by the Recursion Theorem.

6.31. Proposition. *For each function $h : \mathbb{N} \to \mathbb{N}$ and each infinite set $A \subseteq \mathbb{N}$ of natural numbers, there exists a (total) function $f : \mathbb{N} \to \mathbb{N}$ which satisfies the identity*

$$f(n) = \begin{cases} 0, & \text{if } n \in A, \\ h(f(n+1)), & \text{if } n \notin A. \end{cases} \qquad\qquad (6\text{-}17)$$

PROOF. We define the mapping

$$\pi : (\mathbb{N} \rightharpoonup \mathbb{N}) \to (\mathbb{N} \rightharpoonup \mathbb{N})$$

on the inductive poset of all unary, partial functions on $\mathbb{N}$ by the formula

$$\pi(f) = f', \text{ where } f'(n) = \begin{cases} 0, & \text{if } n \in A, \\ h(f(n+1)), & \text{if } n \notin A. \end{cases} \tag{6-18}$$

In full detail, this means we set

$$\pi(f) = \{(n, 0) \mid n \in A\} \cup \{(n, h(w)) \mid n \notin A \,\&\, (n+1, w) \in f\},$$

which implies by inspection that π is continuous, hence countably continuous. Thus we must have a fixed point f which satisfies (6-17), and it is enough to prove that this f is total. Assume towards a contradiction that $f(n)\uparrow$ for some n. Notice that by (6-17), this means that $n \notin A$, else $f(n)\downarrow$, in fact, $f(n) = 0$. We will prove by induction on i that $f(n+i)\uparrow$, which implies again that for all i, $n+i \notin A$, so that $A \subseteq [0, n)$ is finite, contradicting the hypothesis. BASIS. If $i = 0$, then $f(n+0) = f(n)\uparrow$, by assumption. INDUCTION STEP. Assume that $f(n+i)\uparrow$, so that by (6-17), once more, $n+i \notin A$. Now this implies that $f(n+i) = h(f(n+i+1))$ so that $f(n+i+1)\downarrow \implies f(n+1)\downarrow$ (since h is total), which violates the induction hypothesis. $\dashv$

6.32. Exercise. *Prove in detail that the mapping π in this proof is continuous.*

As a third, typical application of the Continuous Least Fixed Point Theorem we consider the *Euclidean algorithm*.

6.33. Proposition. (1) *There exists exactly one partial function $f : \mathbb{N} \times \mathbb{N} \rightharpoonup \mathbb{N}$ with domain of definition $\{(n, m) \mid n, m \neq 0\}$ which satisfies the following identities for all $0 < n < m$:*

$$\begin{aligned} f(m, n) &= f(n, m), \\ f(n, n) &= n, \\ f(n, m) &= f(n, m - n). \end{aligned} \tag{6-19}$$

(2) *The unique f^* which satisfies the system (6-19) computes the greatest common divisor of any two natural numbers different from 0,*

$$f^*(n, m) = \gcd(n, m) \tag{6-20}$$

$$= \text{the largest } k \text{ which divides evenly}$$

$$\text{both natural numbers } n, m.$$

PROOF. With each partial function $f : \mathbb{N} \times \mathbb{N} \rightharpoonup \mathbb{N}$ we associate the partial function $f' : \mathbb{N} \times \mathbb{N} \rightharpoonup \mathbb{N}$ which is defined by the formula

$$f'(n, m) = \begin{cases} f(m, n), & \text{if } n > m > 0, \\ n, & \text{if } n = m > 0, \\ f(n, m - n) & \text{if } 0 < n < m, \end{cases}$$

and we set

$$\pi(f) = f'.$$

The mapping $\pi : ((\mathbb{N} \times \mathbb{N}) \rightharpoonup \mathbb{N}) \to ((\mathbb{N} \times \mathbb{N}) \rightharpoonup \mathbb{N})$ is continuous, by inspection. It follows that there exists a least partial function $f^* : (\mathbb{N} \times \mathbb{N}) \rightharpoonup \mathbb{N}$ which satisfies
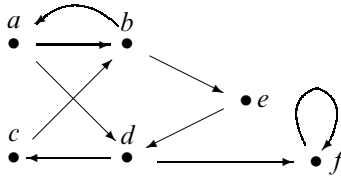
$$\pi(f^*) = f^*,$$

FIGURE **6.5**

and this is (easily) equivalent with the system (6-19). Proof that for all $n, m \neq 0$

$$f^*(n, m){\downarrow} \ \& \ f^*(n, m) = \gcd(n, m)$$

is by induction on the sum $n + m$. (Take cases whether $n > m > 0$, $n = m > 0$ or $0 < n < m$, and use the simple property of the natural numbers, that *for $0 < n < m$, the common divisors of n, m are precisely the same as the common divisors of $n, m - n$.*)   ⊣

In this example we do not need the Least Fixed Point Theorem to prove the existence of a solution for the system (6-19), since we can verify directly that the function gcd is a solution. Despite this, the proposition is important because it yields a characterization of the function gcd which suggests a specific—and simple—method for computing it. For example, using only the identities of the system, we compute:

$$
\begin{aligned}
\gcd(231, 165) &= \gcd(165, 231) = \gcd(165, 66) = \gcd(66, 165) \\
&= \gcd(66, 99) \quad = \gcd(66, 33) \quad = \gcd(33, 66) \\
&= \gcd(33, 33) \quad = 33.
\end{aligned}
$$

This computation of the value $\gcd(231, 165)$ is much simpler than the trivial one, where we would search for the greatest common divisor by testing in sequence all the numbers from 165 moving down, until we would find some common divisor of 165 and 231. The example is quite general: the characterization of a partial function f as the least solution of a system of simple identities typically yields an **algorithm**, a "recipe" for the "mechanical" computation of the values of f, and this is the underlying reason for the significance of the Continuous Least Fixed Point Theorem in theoretical computer science.

We end with a simple result about graphs which is related to the ideas of this chapter, see Problems **x6.16** and **x6.17**.

6.34. Definition. A **graph** is a structured set $(G, \rightarrow_G)$, where the set of **edges** $\rightarrow_G \subseteq G \times G$ is an arbitrary binary relation on the set of **nodes** G. The **transitive closure** of a graph G is the graph $\overline{G} = (G, \Rightarrow_G)$, where

$$x \Rightarrow_G y \iff_{\mathrm{df}} \text{there is a path from } x \text{ to } y \text{ in } G$$

$$\iff (\exists z_0, \ldots, z_n)[x = z_0 \rightarrow_G z_1 \ \& \ \cdots \ \& \ z_{n-1} \rightarrow_G z_n = y].$$

We draw graphs much like posets, except we forget about the convention of "growing up or towards the right": $x \to_G y$ holds if there is an arrow from x to y, and $x \Rightarrow_G y$ holds if you can move from x to y along the arrows of the diagram. In Figure **6.5** we have $f \to f$, $a \Rightarrow a$ and $a \Rightarrow c$, but $f \not\Rightarrow d$.

6.35. Proposition. *For each graph G, the transitive closure relation $\Rightarrow_G$ satisfies the equivalence*

$$x \Rightarrow_G y \iff x \to_G y \vee (\exists z \in G)[x \to_G z \, \& \, z \Rightarrow_G y]. \qquad (6\text{-}21)$$

PROOF. Suppose first that (skipping the conjunction signs)

$$x = z_0 \to_G z_1 \to_G z_2 \to_G \cdots \to_G z_n = y;$$

if $n = 1$, we have $x \to_G y$, and if $n > 1$, then $x \to_G z_1$ and $z_1 \Rightarrow_G y$ (by the definition of $\Rightarrow_G$), so we have the right-hand side of (6-21), taking $z = z_1$. The converse is equally simple, taking cases on the two disjuncts of the right-hand side. $\dashv$

Problems for Chapter 6

x6.1. For every partial ordering $\leq$ on a set A, the **converse relation**

$$x \leq' y \iff_{df} y \leq x$$

is also a partial ordering. Of the inductive posets $(A \to E)$ and $\mathcal{P}(A)$, which one has an inductive, converse poset?

x6.2. Suppose $\leq_E$ is an inductive partial ordering on the set E, A is a set and $\leq$ is the "pointwise" partial ordering on the function space $(A \to E)$,

$$f \leq g \iff_{df} (\forall x \in A)[f(x) \leq_E g(x)] \quad (f, g : A \to E).$$

Prove that $\leq$ is an inductive partial ordering on $(A \to E)$.

x6.3. If the partial orderings $\leq_1, \leq_2$ on the respective sets P_1, P_2 are inductive, then the following relation $\leq$ on the Cartesian product $P_1 \times P_2$ is also inductive:

$$(x_1, x_2) \leq (y_1, y_2) \iff_{df} x_1 \leq_1 y_1 \, \& \, x_2 \leq_2 y_2.$$

With this partial ordering, the poset $P_1 \times P_2$ is called **the product** of the two posets P_1 and P_2.

x6.4. Suppose P_1, P_2, Q are inductive posets. A mapping $\pi : P_1 \times P_2 \to Q$ is **separately monotone** if for each $x_1 \in P_1$, the mapping $(x_2 \mapsto \pi(x_1, x_2))$ on P_2 is monotone, and symmetrically for each $x_2 \in P_2$. Prove that π is monotone (on the product poset) if and only if it is separately monotone,

x6.5. Suppose P_1, P_2, Q are inductive posets. A mapping $\pi : P_1 \times P_2 \to Q$ is **separately countably continuous**, if for each $x_1 \in P_1$, the mapping $(x_2 \mapsto \pi(x_1, x_2))$ on P_2 is countably continuous, and symmetrically for each $x_2 \in P_2$. Prove that π is countably continuous if and only if it is separately countably continuous.

6.36. Definition. A point M is **maximal** in a subset S of a poset P if M is a member of S and no member of S is bigger,

$$M \in S \,\&\, (\forall x \in S)[M \leq x \Longrightarrow M = x].$$

A point m is **minimal** in S if it is a member of S and no member of S is smaller,

$$m \in S \,\&\, (\forall x \in S)[x \leq m \Longrightarrow x = m].$$

x6.6. Find in the poset of Figure **6.2** a subset S which has a maximal element but no maximum and another subset S' which has a minimal element but no minimum.

*__**x6.7.**__ Every finite, non-empty subset of an arbitrary poset P has at least one maximal and one minimal member.

*__**x6.8.**__ A finite poset P is inductive if and only if it has a least element.

An important notion in computer science is that of a *stream*, for example the stream of bytes in a file transmitted over the telephone lines to my home computer from the University of Athens CYBER. A stream is basically a sequence, but it may be *infinite*, in the idealized case; *terminated*, if after some stage an end-of-file signal comes and my machine knows that the transmission is done; or *stalled*, if after some stage the bytes stop coming, without warning, perhaps because the CYBER died or the telephone connection was interrupted.[16]

6.37. Definition. For each set A, we fix some $t \notin A$ (for example, the object $r(A)$ of (3-4)) and we define the **streams from** A by:

Streams(A)

$$=_{\mathrm{df}} \{\sigma : \mathbb{N} \rightharpoonup A \cup \{t\} \mid (\forall i < j)[\sigma(j)\downarrow \;\Longrightarrow\; [\sigma(i)\downarrow \,\&\, \sigma(i) \neq t]]\}.$$

We call a stream σ terminated or **convergent** if for some n, $\sigma(n) = t$, in which case, by the definition $\mathrm{Domain}(\sigma) = [0, n+1)$; **infinite** if $\mathrm{Domain}(\sigma) = \mathbb{N}$; and **stalled** if $\mathrm{Domain}(\sigma)$ is a finite, initial segment of $\mathbb{N}$ but σ does not take on the *terminating value t*. The infinite and stalled streams together are called **divergent**.

x6.9. For each set A, the set of streams Streams(A) is an inductive poset under the natural, partial ordering $\sqsubseteq$, where, as for strings,

$$\sigma \sqsubseteq \tau \iff_{\mathrm{df}} \sigma \subseteq \tau. \tag{6-22}$$

What are its maximal elements?

x6.10. The **concatenation** $\sigma \star \tau$ of two streams is defined so that if σ is divergent, then $\sigma \star \tau = \sigma$ and if σ is convergent with domain $[0, n+1)$, then

$$i < n \Longrightarrow (\sigma \star \tau)(i) = \sigma(i), \quad (\sigma \star \tau)(n + i) = \tau(i).$$

---

[16]Well, the CYBER has died completely since the first edition of these Notes, but the newer machines and more robust telephone lines still quit unexpectedly on some occasions!

Prove that $\star$ is a continuous function (of two variables) on Streams(A).

The full Least Fixed Point Theorem can be proved directly and easily for powersets:

*x6.11. Suppose $\pi : \mathcal{P}(A) \to \mathcal{P}(A)$ is a monotone mapping on a powerset. Prove that the set

$$A^{\star} = \bigcap \{X \mid \pi(X) \subseteq X\}$$

is the least fixed point of π, and

$$A_{\star} = \bigcup \{X \mid X \subseteq \pi(X)\}$$

is the largest fixed point of π.

The next few problems deal with "algorithmic" applications of the Least Fixed Point Theorem.

x6.12. For each relation $R \subseteq \mathbb{N} \times A$, there exists a least partial function $f : \mathbb{N} \times A \rightharpoonup \mathbb{N}$ such that

$$\begin{cases} R(n, x) \implies f(n, x) = n, \\ \neg R(n, x) \implies f(n, x) = f(n + 1, x), \end{cases}$$

It follows that

$$f(n, x){\downarrow} \iff (\exists m \geq n)[R(m, x),$$
$$f(n, x){\downarrow} \implies f(n, x) = \text{the least } m \geq n \text{ such that } R(m, x).$$

x6.13. For any three partial functions f_0, g, h with domains and ranges such that the identities below make sense, there exists a least partial function $f : \mathbb{N} \times A \rightharpoonup E$ which satisfies the identities

$$f(0, x) = f_0(x),$$
$$f(n + 1, x) = h(f(n, g(n, x)), n, x).$$

x6.14. Prove that there is exactly one total function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ which satisfies the identities

$$f(0, n) = f(n, 0) = 0,$$
$$f(n + 1, m + 1) = f(n, m) + 1.$$

Compute $f(5, 23)$ using these identities and "explain" what $f(n, m)$ is, for any n, m.

6.38. Definition. On the set E^* of strings (finite sequences) from a set E defined in 5.30, we define the partial functions

$$\text{head}(u) = u(0), \tag{6-23}$$
$$\text{tail}(u) = \langle u(1), \ldots, u(\text{lh}(u) - 1)\rangle. \tag{6-24}$$

Notice that $\text{head}(u){\downarrow}$ when $\text{lh}(u) > 0$, while $\text{tail}(u)$ is always defined, but it is the empty string when $\text{lh}(u) \leq 1$.

x6.15. Prove that there exists a unique, total function $r : E^* \to E^*$ which satisfies the identity

$$r(u) = \begin{cases} u, & \text{if } \mathrm{lh}(u) \leq 1, \\ r(\mathrm{tail}(u)) \star \langle \mathrm{head}(u) \rangle, & \text{if } \mathrm{lh}(u) > 1. \end{cases}$$

Compute $r(\langle a, b, c \rangle)$ and describe $r(u)$ in general.

x6.16. Prove that for each graph G, the relation $\Rightarrow_G$ is the least (under $\subseteq$) transitive relation on G which includes the edge relation $\to_G$.

x6.17. Prove that for every graph G with edge relation $\to_G$, the relation $\Rightarrow_G$ is the common least fixed point of the following monotone operators on the poset $\mathcal{P}(G \times G)$ of all binary relations on G:

$$\pi_1(R) = \{(x, y) \mid x \to_G y \vee (\exists z)[x \to_G z \,\&\, (z, y) \in R]\},$$
$$\pi_2(R) = \{(x, y) \mid x \to_G y \vee (\exists z)[(x, z) \in R \,\&\, z \to_G y]\},$$
$$\pi_3(R) = \{(x, y) \mid x \to_G y \vee (\exists z)[x \to_G z \to_G y]$$
$$\vee (\exists z, w)[x \to_G z \,\&\, (z, w) \in R \,\&\, w \to_G y]\}.$$

x6.18. Let P_1, P_2 be inductive posets and

$$\pi_1 : P_1 \times P_2 \to P_1,$$
$$\pi_2 : P_1 \times P_2 \to P_2$$

arbitrary countably continuous, monotone mappings, where $P_1 \times P_2$ is the product. Prove that there exist unique **least, mutual** or **simultaneous fixed points** x_1^*, x_2^* which are characterized by the properties:

$$\pi_1(x_1^*, x_2^*) = x_1^*, \quad \pi_2(x_1^*, x_2^*) = x_2^*,$$
$$\pi_1(y_1, y_2) \leq_1 y_1 \,\&\, \pi_2(y_1, y_2) \leq_2 y_2 \Longrightarrow x_1^* \leq_1 y_1 \,\&\, x_2^* \leq_2 y_2.$$

The next problem is an algorithmic version of the well-known number theoretic result, that for any two natural numbers $n, m \neq 0$, there exist (positive or negative) integers α, β such that

$$\gcd(n, m) = \alpha n + \beta m.$$

The proof uses some simple properties of the set of rational integers

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

*__x6.19.__ There exists exactly one pair of partial functions

$$\alpha : \mathbb{N} \times \mathbb{N} \rightharpoonup \mathbb{Z}, \quad \beta : \mathbb{N} \times \mathbb{N} \rightharpoonup \mathbb{Z},$$

with common domain of definition $\{(n, m) \mid n, m \neq 0\}$ which satisfy the following identities for all $n, m, k > 0$:

$$\text{if } n \neq m, \text{ then } \alpha(n, m) = \beta(m, n),$$
$$\alpha(n, n) = 1, \quad \beta(n, n) = 0,$$
$$\alpha(n, n + k) = \alpha(n, k) - \beta(n, k), \quad \beta(n, n + k) = \beta(n, k).$$

It follows that for any two natural numbers $n, m \neq 0$,

$$\gcd(n, m) = \alpha(n, m)n + \beta(n, m)m.$$

x6.20. Find integers $\alpha, \beta \in \mathbb{Z}$ such that

$$33 = 231\alpha + 165\beta.$$

Find also integers $\alpha_1, \beta_1 \in \mathbb{Z}$ such that

$$1 = 137\alpha_1 + 997\beta_1.$$

6.39. Definition. For each finite partial function $g : A \rightharpoonup E$, the **neighborhood** determined by g in the poset $(A \rightharpoonup E)$ is the set

$$N(g) =_{\text{df}} \{f : A \rightharpoonup E \mid g \subseteq f\}$$

of all extensions of g. *A set $G \subseteq (A \rightharpoonup E)$ is* **open in the topology of pointwise convergence** *if*

$$f \in G \Longrightarrow (\exists g, \text{finite})[f \in N(g) \subseteq G].$$

x6.21. Prove that the family of open sets in $(A \rightharpoonup E)$ defined in **6.39** is a topology by **4.30**, and a mapping

$$\pi : (A \rightharpoonup E) \to (B \rightharpoonup M)$$

is continuous in this topology by **6.25** if and only if it is continuous by **6.22**.

6.40. Definition. A subset $G \subseteq P$ of an inductive poset is **Scott open** if (1) it is upward closed, i.e.,

$$x \in G \,\&\, x \leq y \Longrightarrow y \in G,$$

and (2) for every non-empty chain $S \subseteq P$,

$$\sup S \in G \Longrightarrow (\exists x \in S)[x \in G].$$

*x6.22. Prove that the family of Scott open subsets of an inductive poset P is a topology.

*x6.23. Suppose P and Q are inductive posets and $\pi : P \to Q$ is a mapping. Prove that π is continuous in the relevant Scott topologies if and only if π is monotone and for every non-empty chain $S \subseteq P$,

$$\pi(\sup S) = \sup \pi[S].$$

HINT. Some find it helpful for this problem to first prove and use the fact that for every $c \in P$, the set $\{x \in P \mid x \leq c\}$ is Scott closed.

*x6.24. Suppose A is a countable set and $\pi : (A \rightharpoonup E) \to (B \rightharpoonup M)$ is a mapping. Show that π is continuous (by the definition in **6.22**) if and only if it is continuous with respect to the Scott topologies in the posets $(A \rightharpoonup E)$ and $(B \rightharpoonup M)$.

The Continuous Least Fixed Point Theorem is often formulated for the class of *directed-complete* posets, particularly in Computer Science texts.

6.41. Definition. A subset $S \subseteq P$ of a poset P is **directed** if any two members of S have an upper bound in S,

$$x, y \in S \Longrightarrow (\exists z \in S)[x \leq z \,\&\, y \leq z].$$

A poset P is **directed-complete** (a **dcpo**) if every directed $S \subseteq P$ has a least upper bound.

x6.25. Every chain in a poset is a directed set, hence, every dcpo is an inductive poset and the least fixed point theorems hold for directed-complete posets.

x6.26. For each A and E, the posets $(A \rightharpoonup E)$ and $(A \rightarrowtail E)$ are directed-complete.

x6.27. A mapping $\pi : (A \rightharpoonup E) \to (B \rightharpoonup M)$ is continuous if and only if for each directed $S \subseteq (A \rightharpoonup E)$,

$$\pi(\sup S) = \sup \pi[S].$$

x6.28. The product $P_1 \times P_2$ (Problem **x6.3**) of two directed-complete partial orderings is also directed-complete.

***x6.29.** Every countable, inductive poset is directed-complete.

Actually the notions of *inductive* and *directed-complete* are equivalent; for a monotone mapping $\pi : P \to Q$ on one inductive poset to another, the equation

$$\pi(\sup S) = \sup \pi[S] \qquad\qquad (6\text{-}25)$$

holds for all non-empty chains $S \subseteq P$ if and only if it holds for all non-empty directed sets $S \subseteq P$; and the characterization of Scott continuity in **x6.24** holds whether A is countable or not. The proofs of these results are not elementary and require the Axiom of Choice, see Problems **x9.22** – **x9.25**.

WELL ORDERED SETS

7.1. A **well ordered set**[17] is a poset

$$U = (\text{Field}(U), \leq_U),$$

where $\leq_U$ is a wellordering on $\text{Field}(U)$, i.e., a linear (total) ordering on $\text{Field}(U)$ such that every non-empty $X \subseteq \text{Field}(U)$ has a least member. Associated with U is its **strict ordering** $<_U$,

$$x <_U y \iff_{\text{df}} x \leq_U y \,\&\, x \neq y.$$

For example, the set $\mathbb{N}$ of natural numbers is well ordered by its natural ordering, and so is each of its finite initial segments

$$[0, n) = \{i \in \mathbb{N} \mid i < n\}.$$

A "longer" well ordered set is $(\mathbb{N} \cup \{\infty\}, \leq')$, where ∞ is some object not in $\mathbb{N}$ which we put after all the natural numbers, i.e.,

$$x \leq' y \iff y = \infty \vee [x, y \in \mathbb{N} \,\&\, x \leq y].$$

As we did with arbitrary posets in the preceding chapter, we will usually identify U with its field, talk about *the points* or *subsets* of U, meaning the members and subsets of $\text{Field}(U)$, skip the subscript in $\leq_U$ or $<_U$ when it is obvious from the context, etc.

The most basic results of the last two chapters were all proved by some combination of the coupling

$$\text{definition by recursion – proof by induction.} \qquad (7\text{-}1)$$

In the simplest case, some function $f : \mathbb{N} \to E$ is defined by recursion, some properties of f are proved by induction and these in turn imply the theorem we want. Typical are the Continuous Least Fixed Point and the Schröder-Bernstein theorems which say nothing (explicitly) about recursion, induction or any functions with domain $\mathbb{N}$, but whose proofs most assuredly use precisely these notions. We based the proof of the Recursion Theorem **5.6** directly on the Induction Axiom for the natural numbers. The key fact, however, which

---

[17]Do we dare call them *wosets*? It's not much worse than *posets* and it would sure save a lot of key strokes.

□   ○   ○   ⋯   ■   •   •   ⋯   ■   •   •   ⋯   ⋯

FIGURE **7.1**. The low end of a long wellordering.

can be generalized is that $\mathbb{N}$ is well ordered by its natural ordering. Here we will generalize **5.6** to a powerful TRANSFINITE RECURSION THEOREM **7.24** which justifies *definition by recursion* of functions $f : U \to E$ on any well ordered set U. Coupled with HARTOGS' THEOREM **7.34** which guarantees the existence of "arbitrarily large" well ordered sets, this makes it possible to apply the basic idea of (7-1) in situations far removed from the natural numbers. Typical applications are the FIXED POINT THEOREM **7.35** and its corollary, the LEAST FIXED POINT THEOREM **7.36**, which is just **6.21** without the countable continuity hypothesis.

7.2. A set A is **well orderable** if it admits a wellordering, so it is the field of some well ordered set $(A, \leq)$. One of the chief lessons of this chapter is that well orderable sets behave much better than arbitrary sets: for example any two of them are comparable in cardinality, either $A \leq_c B$ or $B \leq_c A$. In fact, *every set is well orderable*. Zermelo showed this in 1904, settling with one brilliant stroke the problem of Cardinal Comparability and a whole slew of related, regularity questions about arbitrary sets. We will prove Zermelo's Wellordering Theorem in the next chapter, after we introduce the Axiom of Choice on which it is based. It is worth pointing out here, however, that the mathematical content of this fundamental result is just the sum of the Transfinite Recursion and Hartogs' Theorems: the Axiom of Choice simply allows us to put the two together.

7.3. Exercise. *If C is well orderable and $A \leq_c C$, then A is well orderable.*

7.4. Exercise. *If C is well orderable and there exists a surjection $f : C \twoheadrightarrow A$, then $A \leq_c C$, and hence A is also well orderable.*

7.5. Successor and limit points. Every well ordered set U looks at its low end like an initial segment of $\mathbb{N}$. If it is not empty, it must have a least member which is typically denoted by 0 rather than $\perp$,

$$0 = 0_U =_{df} \text{ the least element of } U. \tag{7-2}$$

It is pictured by a hollow square, the first point in Figure **7.1**. Each $x \in U$ other than the maximum (which may or may not exist) has an element following it immediately,

$$S(x) = S_U(x) =_{df} \min_U \{y \in U \mid x < y\}. \tag{7-3}$$

The values of the partial function $S : U \rightharpoonup U$ are the **successor points** of U. In addition, U may have **limit points** which are above 0 but not the successor of anything:

$$\text{Limit}_U(x) \iff_{df} 0 < x \,\&\, (\forall u < x)(\exists v)[u < v < x]. \tag{7-4}$$

These are pictured by black boxes in Figure **7.1**. The first limit point of U is typically denoted by

$$\omega = \omega_U =_{\mathrm{df}} \min\{x \in U \mid \mathrm{Limit}(x)\}, \tag{7-5}$$

when it exists, the points below it are the **finite points** and the points above it (including ω) are the **infinite points** of U. If U is infinite, then the function $\pi : \mathbb{N} \rightarrowtail U$ defined by the recursion

$$\begin{aligned} \pi(0) &= 0_U = \text{ the least member of } U, \\ \pi(n+1) &= S_U(\pi(n)), \end{aligned} \tag{7-6}$$

is an order-preserving correspondence of $\mathbb{N}$ with the finite points of U.

7.6. Exercise. *For each subset $I \subseteq U$ of a well ordered set U, the restriction*

$$x \leq_I y \iff_{\mathrm{df}} x \leq_U y \,\&\, x, y \in I$$

of $\leq_U$ to I is a wellordering, so that I is a well ordered set in its own right with this ordering.

7.7. Definition. A well ordered set U is an **initial segment** of V if $\mathrm{Field}(U)$ is a downward closed subset of $\mathrm{Field}(V)$ and $\leq_U$ is the restriction of $\leq_V$ to $\mathrm{Field}(U)$:

$$\begin{aligned} U \sqsubseteq V \iff_{\mathrm{df}} \; &\mathrm{Field}(U) \subseteq \mathrm{Field}(V) \tag{7-7} \\ &\&\, (\forall y \in \mathrm{Field}(U))(\forall x \leq_V y)[x \in \mathrm{Field}(U)] \\ &\&\, (\forall x, y \in \mathrm{Field}(U))[x \leq_U y \iff x \leq_V y]. \end{aligned}$$

Clearly V is an initial segment of itself, the **trivial** one. With each $y \in V$ we associate the **proper initial segment** of points strictly below y,

$$\mathbf{seg}(y) = \mathbf{seg}_V(y) =_{\mathrm{df}} \{x \in V \mid x <_V y\} \subsetneq V. \tag{7-8}$$

More precisely, this is the field of $\mathbf{seg}(y)$, but the ordering is determined by V and we will talk about initial segments as if they were just sets, as usual.

7.8. Exercise. *If 0 is the least element of U, then $\mathbf{seg}(0) = \emptyset$, and if $x \in U$ has a successor, then*

$$\mathbf{seg}(S(x)) = \mathbf{seg}(x) \cup \{x\}.$$

7.9. Proposition. *A set I is an initial segment of a well ordered set U if and only if $I = U$ or for some $x \in U$, $I = \mathbf{seg}(x)$.*

PROOF. If $I \subsetneq U$, let $x = \min(U \setminus I)$ so that immediately,

$$y \in \mathbf{seg}(x) \Longrightarrow y < x \Longrightarrow y \in I,$$

and to verify that $I = \mathbf{seg}(x)$, it is enough to prove

$$y \in I \Longrightarrow y < x.$$

Towards a contradiction, if $y \in I$ but $y \nless x$, then we must have $x \leq y$, which implies $x \in I$ because I is downwards closed, contradicting the choice of x. The converse is trivial. $\dashv$

7.10. Exercise. *The family of initial segments of a well ordered set U is well ordered by the relation $\sqsubseteq$.*

The general idea is to view a well ordered set U as a generalization of the natural number sequence $0, 1, 2, \ldots$, possibly shorter than or of equal length to $\mathbb{N}$, typically much longer. The particular members of U will be of little consequence: it is the *length* of the sequence in which we will be interested. We introduce here the general notion of isomorphism which relates posets with the same *shape*, the shape of a well ordered set being just a "length".

7.11. Definition. A function $\pi : P \to Q$ from one poset into another is **order-preserving** if for all $x, y \in P$,

$$x \leq_P y \iff \pi(x) \leq_Q \pi(y);$$

a **similarity** is an order-preserving bijection $\pi : P \rightarrowtail\!\!\!\rightarrow Q$, and if one such exists we call P and Q **similar**, **order isomorphic** or **copies** of each other. We write

$$P =_o Q \iff_{\text{df}} (\exists \pi : P \rightarrowtail\!\!\!\rightarrow Q)[\pi \text{ is a similarity}].$$

The subscript o in $=_o$ stands for "order type", a fancier expression for "shape". Notice that by our general convention of talking about a poset as if it were its field, we write $\pi : P \rightarrowtail\!\!\!\rightarrow Q$ for similarities instead of the more explicit $\pi : \text{Field}(P) \rightarrowtail\!\!\!\rightarrow \text{Field}(Q)$.

7.12. Exercise. *Every order-preserving $\pi : P \to Q$ from one poset to another is monotone; but there exist monotone mappings which are not order-preserving.*

7.13. Exercise. *If P and Q are linear posets, then a function $f : P \to Q$ is order-preserving if and only if it is **strictly monotone**, i.e.,*

$$x <_P y \Longrightarrow f(x) <_Q f(y).$$

In particular, order-preserving functions on well ordered sets are strictly monotone, and hence one-to-one.

7.14. Exercise. *For all posets P, Q, R,*

$$P =_o P,$$
$$P =_o Q \Longrightarrow Q =_o P,$$
$$P =_o Q \,\&\, Q =_o R \Longrightarrow P =_o R.$$

7.15. Lemma. *If a poset P is similar to a well ordered set U, then it is also well ordered.*

PROOF. Given $\emptyset \neq X \subseteq P$, let $p \in U$ be the $\leq_U$-least element of the image $\pi[X]$ and verify (easily) that $x = \pi^{-1}(p)$ is $\leq_P$-least in X, because π preserves the orderings. ⊣

We can construct explicitly some fairly long wellorderings by starting with $\mathbb{N}$ and its finite initial segments and applying repeatedly several natural operations on posets which yield well ordered sets on well ordered arguments. Here we look at just one of these, leaving the rest for the problems.
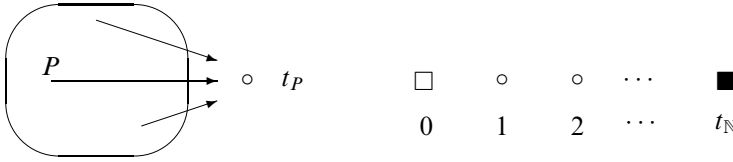
Figure 7.2. The successor poset to P and Succ($\mathbb{N}$).

7.16. The **successor** Succ(P) of a poset P is obtained by adding a new point above all the members of P. To be specific, we can choose to add to the field of P the object

$$t_P =_{\text{df}} \mathbf{r}(\text{Field}(P)) \tag{7-9}$$

which is guaranteed by **3.11** to be a new element, and we set

$$x \leq_{\text{Succ}(P)} y \iff_{\text{df}} x \leq_P y \vee [x \in P \,\&\, y = t_P] \vee x = y = t_P. \tag{7-10}$$

If P is finite with n elements, then Succ(P) has $n + 1$ elements, in fact, easily Succ($[0, n)$) $=_o [0, (n+1))$. On the other hand, Succ($\mathbb{N}$) is countably infinite, but with a different, "longer" ordering than $\mathbb{N}$, as it has a maximum element which comes after all the natural numbers.

7.17. Exercise. If $P =_o Q$, then Succ(P) $=_o$ Succ(Q).

7.18. Exercise. If U is well ordered, so is Succ(U).

Using this successor operation on posets, we can view each well ordered set U as a proper initial segment of another,

$$U = \mathbf{seg}_{\text{Succ}(U)}(t_U) \underset{\neq}{\sqsubseteq} \text{Succ}(U). \tag{7-11}$$

7.19. Definition. A mapping $\pi : P \to P$ on a poset to itself is **expansive**, if for all $x \in P$, $x \leq \pi(x)$.

7.20. Theorem. *Every order-preserving injection $\pi : U \rightarrowtail U$ of a well ordered set into itself is expansive.*

Proof. Towards a contradiction, assume that $\pi : U \rightarrowtail U$ is order-preserving but that for some $x \in U$, $\pi(x) < x$, and let

$$x^* = \min\{x \in U \mid \pi(x) < x\}.$$

Thus, $\pi(x^*) < x^*$, and so $\pi(\pi(x^*)) < \pi(x^*)$ since π is an order-preserving injection, which contradicts the choice of x^*. ⊣

7.21. Corollary. *No well ordered set is similar with one of its proper initial segments, and hence no two distinct initial segments of a well ordered set are similar.*

Proof. Every similarity $\pi : U \rightarrowtail \mathbf{seg}(x)$ is (in particular) an order-preserving injection of U into U, so we cannot have $\pi(x) < x$, by the theorem.⊣

Because a well ordered set may have limit points in addition to its 0 and its successor points, it is easiest to generalize the principles of proof by *complete induction* and definition by *complete recursion*.

7.22. Transfinite Induction Theorem. *For every well ordered set U and every unary definite condition P,*

$$\text{if } (\forall y \in U)[(\forall x < y)P(x) \Longrightarrow P(y)] \text{ then } (\forall y \in U)P(y).$$

PROOF. Assuming the opposite, towards a contradiction, let

$$y^* =_{\mathrm{df}} \min\{y \in U \mid (\forall x < y)P(x) \,\&\, \neg P(y)\};$$

the hypothesis yields $P(y^*)$, which contradicts the choice of y^*.   ⊣

In specific cases, it is often just as easy to prove $(\forall y \in U)P(y)$ by contradiction rather than appeal to **7.22**, in effect repeating this little argument. It depends on the statement to be proved and how much one is annoyed by dealing with negative statements. We will illustrate both styles. Incidentally, the term "transfinite" is used because U may be longer than $\mathbb{N}$, but the theorem also holds, of course, when U is finite or similar with $\mathbb{N}$.

The next lemma is the key step in the proof of the fundamental theorem which follows it.

7.23. Lemma. *Suppose U is a well ordered set and $h : (U \rightharpoonup E) \times U \to E$ maps the partial functions from U to E and U into E. It follows that for every $t \in U$, there exists exactly one function*

$$\sigma_t : \mathbf{seg}(t) \to E$$

which satisfies the identity

$$\sigma_t(x) = h(\sigma_t \restriction \mathbf{seg}(x), x) \quad (x < t). \tag{7-12}$$

PROOF. By Transfinite Induction, assume that for each $u < t$ there exists exactly one function $\sigma_u : \mathbf{seg}(u) \to E$ such that

$$\sigma_u(x) = h(\sigma_u \restriction \mathbf{seg}(x), x) \quad (x < u). \tag{7-13}$$

The induction hypothesis gives us nothing if $t = 0_U$ is the least point in U, but the required conclusion is trivial in this case taking $\sigma_0 = \emptyset$. If $t = Sv$ is a successor point in U, we set

$$\sigma_t = \sigma_v \cup \{(v, h(\sigma_v, v))\};$$

now (7-13) holds for $x < v$ by the induction hypothesis and it holds for $x = v$ by the definition (since $\sigma_t \restriction \mathbf{seg}(t) = \sigma_v$). For the last case, when t is a limit point, we need a

Sublemma. *The set of functions $\{\sigma_u \mid u < t\}$ is a chain under $\subseteq$, i.e.,*

$$x < u < v < t \Longrightarrow \sigma_u(x) = \sigma_v(x). \tag{7-14}$$

Proof. Assume not and let x be least such that (7-14) fails for some $u > x$, $u < v < t$. This means that

$$\sigma_u \restriction \mathbf{seg}(x) = \sigma_v \restriction \mathbf{seg}(x),$$

and then by the identity which σ_u, σ_v satisfy,

$$\begin{aligned}
\sigma_u(x) &= h(\sigma_u \restriction \mathbf{seg}(x), x) \\
&= h(\sigma_v \restriction \mathbf{seg}(x), x) \\
&= \sigma_v(x),
\end{aligned}$$

which contradicts the choice of x.   ⊣ (Sublemma)

We now take

$$\sigma_t = \bigcup\{\sigma_u \mid u < t\};$$

this is a function with domain $\mathbf{seg}(t)$ by the Sublemma, and it satisfies (7-12), since for each $x < t$,

$$\begin{aligned}
\sigma_t(x) &= \sigma_u(x) &&\text{for some } u \text{ such that } x < u < t, \\
&= h(\sigma_u \restriction \mathbf{seg}(x), x) &&\text{by ind. hyp., since } u < t, \\
&= h(\sigma_t \restriction \mathbf{seg}(x), x) &&\text{since } \sigma_u \restriction \mathbf{seg}(x) = \sigma_t \restriction \mathbf{seg}(x) \\
& &&\text{by the definition of } \sigma_t.
\end{aligned}$$

This completes the proof of existence of σ_t, and its uniqueness is easily verified by Transfinite Induction, using (7-12).   ⊣

7.24. Transfinite Recursion Theorem. *For each well ordered set U and each function $h : (U \rightharpoonup E) \times U \to E$, there exists exactly one function $f : U \to E$ which satisfies the identity*

$$f(x) = h(f \restriction \mathbf{seg}(x), x) \quad (x \in U). \tag{7-15}$$

PROOF. Consider the well ordered set $\mathrm{Succ}(U)$ which has some point $t = t_U$ on top of U, and the extension $h' : (\mathrm{Succ}(U) \rightharpoonup E) \times \mathrm{Succ}(U) \to E$ of h defined by

$$h'(\sigma, x) = \begin{cases} h(\sigma \restriction U, x), & \text{if } x \in U, \\ e^*, & \text{otherwise, i.e., if } x = t, \end{cases}$$

where e^* is some arbitrary member of E, of no consequence. The function h' has the correct domain for applying the Lemma to $\mathrm{Succ}(U)$ and h', because $\mathbf{seg}_{\mathrm{Succ}(U)}(t) = U$. For the top point t, the Lemma gives a unique function $f = \sigma_t : U \to E$ which satisfies (7-12) for all $x \in U$.   ⊣

Perhaps the simplest, non-trivial application of Transfinite Recursion is the definition of transfinite orbits for mappings of an inductive poset into itself. We consider first the basic case of *expansive* mappings, defined in **7.19**. Expansive mappings are related to the monotone mappings we studied in the last chapter, but the two notions do not coincide; for example, the constant mapping $X \mapsto \emptyset$ on $\mathcal{P}(\mathbb{N})$ is obviously monotone but not expansive, while

$$\pi(X) = \begin{cases} X \cup \{1\} & \text{if } 0 \in X, \\ X \cup \{2\} & \text{if } 0 \notin X \end{cases}$$

is expansive but (easily) not monotone. It turns out, however, that results about expansive mappings can often be translated into similar results about monotone mappings.
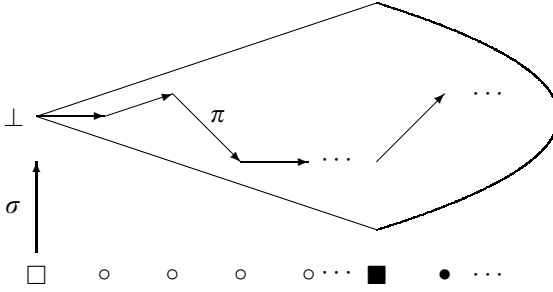
FIGURE **7.3**. The transfinite orbit of an expansive mapping.

7.25. Iteration Lemma. *Suppose* $\pi : P \to P$ *is an expansive mapping on an inductive poset and* U *is a well ordered set. There exists a unique function* $\sigma : U \to P$ *which satisfies the following conditions*:

$$\sigma(0) = \perp,$$
$$\text{if } x = S(y), \text{ then } \sigma(x) = \pi(\sigma(y)), \qquad (7\text{-}16)$$
$$\text{if } \text{Limit}(x), \text{ then } \sigma(x) = \sup_P \{\sigma(y) \mid y < x\}.$$

In addition, this σ *is monotone from* U *to* P, *i.e.*,

$$x \le y \Longrightarrow \sigma(x) \le_P \sigma(y). \qquad (7\text{-}17)$$

PROOF. The conditions in (7-16) just about give a definition of σ by transfinite recursion, except that there is a problem in the limit case if the set $\{\sigma(y) \mid y < x\}$ is not a chain in P. To account for this possibility, we define σ by appealing to Theorem **7.24** so that it satisfies the following:

$$\sigma(x) = \begin{cases} \perp, & \text{if } x = 0, \\ \pi(\sigma(y)), & \text{if } x = S(y) \text{ for some } y, \\ \sup_P \{\sigma(y) \mid y < x\}, & \text{if } \text{Limit}(x) \\ \qquad \& \ (\forall x_1 < x_2 < x)[\sigma(x_1) \le_P \sigma(x_2)], \\ \perp, & \text{otherwise}, \end{cases}$$

where $\le \ = \ \le_U$ is the wellordering of U, as in the statement of the theorem.

Sublemma. *For each* $x \in U$,

$$x_1 < x_2 \le x \Longrightarrow \sigma(x_1) \le_P \sigma(x_2). \qquad (7\text{-}18)$$

Proof. Assume not and let x be least in U such that (7-18) fails. Since (7-18) holds vacuously when $x = 0$ is the least element in U, we need consider only two cases.

CASE 1. $x = S(y)$ is a successor point. Assume $x_1 < x_2 \le x$. If $x_2 \le y$, we get $\sigma(x_1) \le_P \sigma(x_2)$ by the choice of x. The only other possibility is that $x_2 = x$, but then $\sigma(x_1) \le_P \sigma(y)$ by the choice of x, and $\sigma(y) \le \pi(\sigma(y)) = \sigma(x)$ by the expansiveness of π. Thus, (7-18) holds for x, which is a contradiction.

CASE 2. x is limit. By the choice of x,

$$x_1 < x_2 < x \Longrightarrow \sigma(x_1) \leq \sigma(x_2), \qquad\qquad (7\text{-}19)$$

and so in order to get a contradiction, we only need show that

$$x_1 < x \Longrightarrow \sigma(x_1) \leq_P \sigma(x).$$

This holds because (7-19) and the definition of σ also imply immediately that $\sigma(x) = \sup_P \{\sigma(y) \mid y < x\}$. $\dashv$ (Sublemma)

The result now follows directly, since the Sublemma implies in particular that the "otherwise" case in the definition of σ never comes up. $\dashv$

The transfinite orbit $\sigma : U \to P$ of a mapping $\pi : P \to P$ guaranteed by the Iteration Lemma is obviously an extension of the orbit $(n \mapsto x_n)$ which we defined in the proof of the Continuous Least Fixed Point Theorem **6.21**, at least if the well ordered set U is longer than $\mathbb{N}$. It is one of the tools we will use in the proof of the Least Fixed Point Theorem, as follows.

7.26. Plan for a proof. Suppose that for the given inductive poset P, we can construct a well ordered set U such that *there exists no injection $\sigma : U \rightarrowtail P$.* In particular, the transfinite orbit $\sigma : U \to P$ of **7.25** cannot be an injection, and so there must exist $x < y$ such that $\sigma(x) = \sigma(y)$. The monotonicity of σ implies that

$$x \leq u \leq y \Longrightarrow \sigma(x) = \sigma(u);$$

x has a successor since it is not maximum in U, $x < Sx \leq y$; and, hence,

$$\sigma(x) = \sigma(Sx) = \pi(\sigma(x)),$$

in other words, the point $\sigma(x)$ is a fixed point of π.

Thus, to prove that every expansive mapping $\pi : P \to P$ on an inductive poset has a fixed point, it is sufficient to show that *for each set P, there exists some well ordered set U which cannot be injected into P.* This is precisely Hartogs' Theorem, for which we aim next. To show it, we must study in some detail the question of comparability of well ordered sets as to length.

The picture of the typical well ordered set in Figure **7.1** suggests that we should be able to compare any two of them, line them up side-by-side, the least element 0_U of one facing the least element 0_V of the other, the next $S_U(0_U)$ facing $S_V(0_V)$, the first limit point ω_U (if it exists) facing ω_V, etc. until we run out of elements in either U or V. The precise version of this fact is a generalization of the Uniqueness Theorem for the natural numbers **5.4**.

7.27. Definition. An **initial similarity**

$$\pi : U \rightarrowtail \pi[U] \sqsubseteq V$$

from one well ordered set into another is a similarity of U with an initial segment of V. If such an initial similarity exists, we say that U is **less than or equal to V in length**, in symbols:

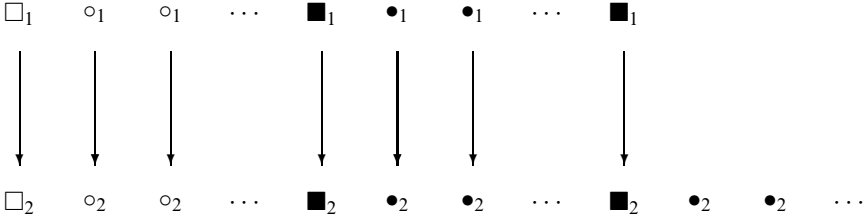$$U \leq_o V \iff_{\mathrm{df}} (\exists I \sqsubseteq V)[U =_o I]. \qquad\qquad (7\text{-}20)$$

FIGURE 7.4. Portrait of an initial similarity.

We also write

$$U <_o V \iff_{df} U \leq_o V \ \& \ U \neq_o V. \qquad (7\text{-}21)$$

By **7.9**, every initial similarity $\pi : U \rightarrowtail V$ is either a similarity with V, or one with a proper initial segment of V, so that

$$U <_o V \iff (\exists x \in V)[U =_o \mathbf{seg}_V(x)]. \qquad (7\text{-}22)$$

7.28. Exercise. *If $\pi : U \rightarrowtail V$ and $\rho : V \rightarrowtail W$ are initial similarities, then so is their composition $\rho\pi : U \rightarrowtail W$.*

7.29. Proposition. *For all well ordered sets U, V, W,*

$$U \leq_o U,$$
$$\text{if } [U \leq_o V \ \& \ V \leq_o W], \text{ then } U \leq_o W,$$
$$\text{if } [U \leq_o V \ \& \ V \leq_o U], \text{ then } U =_o V.$$

PROOF. Only the third of these assertions needs proof and it follows from **7.21**. The composition $\rho\pi$ of the initial similarities $\pi : U \rightarrowtail V$, $\rho : V \rightarrowtail U$ witnessing the hypothesis is an initial similarity $\rho\pi : U \rightarrowtail U$, which if it were not onto, would witness that U is similar with one of its proper initial segments; so it is a bijection, and then π must also be a bijection. ⊣

7.30. Theorem. *A function $\pi : U \to V$ is an initial similarity of a well ordered set into another if and only if it satisfies the identity*

$$\pi(x) = \min_V\{y \in V \mid (\forall u <_U x)[\pi(u) <_V y]\} \quad (x \in U). \qquad (7\text{-}23)$$

PROOF. If $\pi : U \rightarrowtail V$ is an initial similarity, then it is order-preserving and one-to-one, so it satisfies

$$(\forall u <_U x)[\pi(u) <_V \pi(x)], \qquad (7\text{-}24)$$

and hence

$$z = \min_V\{y \in V \mid (\forall u <_U x)[\pi(u) <_V y]\} \leq_V \pi(x).$$

Since π is initial and $z \leq_V \pi(x)$, there exists some $u \in U$ such that $\pi(u) = z$. Assuming towards a contradiction that $z = \pi(u) <_V \pi(x)$, we infer that $u <_U x$ because π is an order-preserving injection; and hence $\pi(u) <_V z$ by the definition of z, which is absurd since $z = \pi(u)$.

Conversely, if $\pi : U \to V$ satisfies (7-23), then it is an order-preserving injection, since by (7-23), $u <_U x \implies \pi(u) <_V \pi(x)$. Suppose the image

$\pi[U]$ is not an initial segment of V and choose x least in U such that there exists some $y <_V \pi(x)$, $y \notin \pi[U]$. Now $\pi[\mathbf{seg}_U(x)] \sqsubseteq V$ by the choice of x; it is a proper initial segment since it does not contain y; so $\pi[\mathbf{seg}_U(x)] = \mathbf{seg}_V(z)$ for some $z \in V$, and (7-23) yields $\pi(x) = z$. Thus $y <_V z$ and $y \in \mathbf{seg}_V(z) = \pi[\mathbf{seg}_U(x)]$, which is absurd. $\dashv$

7.31. Theorem (Comparability of well ordered sets). *For any two well ordered sets U, V, either $U \leq_o V$ or $V \leq_o U$.*

PROOF. The result is trivial if $V = \emptyset$, so we may assume the minimum 0_V exists. By the Transfinite Recursion Theorem **7.24**, there exists a function $\pi : U \to V$ which satisfies the identity

$$\pi(x) = \begin{cases} \min_V \{y \in V \mid (\forall u <_U x)[\pi(u) <_V y]\}, \\ \qquad \text{if } (\exists y \in V)(\forall u <_U x)[\pi(u) <_V y], \qquad (7\text{-}25) \\ 0_V, \text{ otherwise.} \end{cases}$$

In pedantic detail, we are applying here Theorem **7.24** with the mapping $h : (U \rightharpoonup E) \times U \to E$, defined by

$$h(p, x) = \begin{cases} \min_V \{y \in V \mid (\forall u <_U x)[p(u) <_V y]\}, \\ \qquad \text{if } (\exists y \in V)(\forall u <_U x)[p(u) <_V y], \\ 0_V, \text{ otherwise.} \end{cases}$$

We now distinguish two possibilities.

CASE 1. *For every $x \neq 0_U, \pi(x) \neq 0_V$.* This means that the second case in (7-25) never applies, π satisfies the identity (7-23) and it must be an initial similarity by Theorem **7.30**.

CASE 2. *For some $a \in U, a \neq 0_U$, we have $\pi(a) = 0_V$.* Let $a \neq 0_U$ be least in U and such that $\pi(a) = 0_V$, and consider the restriction

$$\rho = (\pi \upharpoonright \mathbf{seg}_U(a)) : \mathbf{seg}_U(a) \to V.$$

Now ρ satisfies (7-23), so by Theorem **7.30**, it is an initial similarity of $\mathbf{seg}_U(a)$ into V. In particular, the image $\rho[\mathbf{seg}_U(a)] = \pi[\mathbf{seg}_U(a)]$ is an initial segment of V; if it were proper, then $\pi[\mathbf{seg}_U(a)] = \mathbf{seg}_V(z)$ for some $z \in V$ and (7-25) would yield $\pi(a) = z \neq 0_V$, contradicting the choice of a; hence, $\pi[\mathbf{seg}_U(a)] = V$. Thus, $V =_o \mathbf{seg}_U(a)$, which gives us an initial similarity of V into U. $\dashv$

This fundamental theorem has a host of corollaries, some of which are worth listing immediately. The first one gives us an easier way to compare well ordered sets.

7.32. Corollary. *For all well ordered sets U, V,*

$$U \leq_o V \iff (\exists \pi : U \rightarrowtail V)[\pi \text{ is order-preserving}].$$

PROOF. Suppose $\pi : U \rightarrowtail V$ is order-preserving but $U \not\leq_o V$, so that $V <_o U$. It follows that $V =_o \mathbf{seg}_U(x)$ for some x by (7-22), and composing the order-preserving injections we get an injection $\rho : U \rightarrowtail \mathbf{seg}_U(x)$ which is still order-preserving and violates **7.20**. $\dashv$

7.33. Corollary (Wellfoundedness of $\leq_o$). *Every non-empty class $\mathscr{E}$ of well ordered sets has a $\leq_o$-least member, i.e., for some $U_0 \in \mathscr{E}$ and all $U \in \mathscr{E}$, $U_0 \leq_o U$.*

PROOF. The hypothesis gives us some $W \in \mathscr{E}$, and if W is $\leq_o$-least in $\mathscr{E}$, there is nothing to prove. If not, then **7.31** implies that there exists well ordered sets in $\mathscr{E}$ which are similar with proper initial segments of W, so the set

$$J =_{\mathrm{df}} \{x \in W \mid (\exists U \in \mathscr{E})[U =_o \mathbf{seg}_W(x)]\} \tag{7-26}$$

is non-empty and it has a $\leq_W$-least element x. By the definition of J, there exist some $U_0 \in \mathscr{E}$ such that $U_0 =_o \mathbf{seg}_W(x)$ and we claim that this U_0 is $\leq_o$-least in $\mathscr{E}$. To prove it, assume towards a contradiction that for some $U \in \mathscr{E}$, $U_0 \not\leq_o U$; hence $U <_o U_0 =_o \mathbf{seg}_W(x)$; hence $U =_o \mathbf{seg}_W(y)$ for some $y <_W x$, contradicting the choice of x. ⊣

Most often this is applied when $\mathscr{E}$ is actually a set, a family of well ordered sets, but occasionally it is convenient to cite it more generally for classes. For example, there exists a $\leq_o$-least well ordered set which has a limit point—namely Succ($\mathbb{N}$).

After all this work, still we have not constructed any uncountable well ordered sets and it might appear that all our results apply only to peculiar, long reshufflings of $\mathbb{N}$. Next comes the second basic theorem of this chapter which rectifies the situation.

7.34. Hartogs' Theorem. *There is a definite operation $\chi(A)$ which associates with each set A, a well ordered set*

$$\chi(A) = (h(A), \leq_{\chi(A)}),$$

such that $h(A) \not\leq_c A$, i.e., there exists no injection $\pi : h(A) \rightarrowtail A$. Moreover, $\chi(A)$ is $\leq_o$-minimal with this property, i.e., for every well ordered set W,

$$\text{if } W \not\leq_c A, \text{ then } \chi(A) \leq_o W. \tag{7-27}$$

PROOF. First set

$$\mathrm{WO}(A) =_{\mathrm{df}} \{U \mid U = (\mathrm{Field}(U), \leq_U) \text{ is a well ordered set}$$
$$\text{with } \mathrm{Field}(U) \subseteq A\}, \tag{7-28}$$

and let $\sim_A$ be the restriction of the definite condition $=_o$ to $\mathrm{WO}(A)$,

$$U \sim_A V \iff_{\mathrm{df}} U, V \in \mathrm{WO}(A) \,\&\, U =_o V.$$

Clearly $\sim_A$ is an equivalence relation on $\mathrm{WO}(A)$, and we set

$$h(A) =_{\mathrm{df}} [\![\mathrm{WO}(A)/\!\sim_A]\!] \subseteq \mathcal{P}(\mathrm{WO}(A)). \tag{7-29}$$

We order the equivalence classes in $h(A)$ by their "representatives",

$$[U/\!\sim_A] \leq_{\chi(A)} [V/\!\sim_A] \iff_{\mathrm{df}} U \leq_o V; \tag{7-30}$$

this makes sense because if

$$[U/\!\sim_A] = [U'/\!\sim_A], \; [V/\!\sim_A] = [V'/\!\sim_A], \text{ and } U \leq_o V,$$

then $U' =_o U \leq_o V =_o V'$. The fact that $\leq_{\chi(A)}$ is a wellordering of $h(A)$ follows easily from the general properties of $\leq_o$, **7.31** and **7.33**. Taking the negation of both sides of (7-30) we infer its strict version,

$$V <_o U \iff [V/\sim_A] <_{\chi(A)} [U/\sim_A] \quad (U, V \in \mathrm{WO}(A)). \qquad (7\text{-}31)$$

The basic properties of the Hartogs operation are embodied in the following

Lemma. *For every $\alpha = [U/\sim_A] \in h(A)$,*

$$\mathbf{seg}_{\chi(A)}(\alpha) = \{[\mathbf{seg}_U(x)/\sim_A] \mid x \in U\} =_o U.$$

In particular, every proper initial segment of $\chi(A)$ is similar with some well ordered set $U \in \mathrm{WO}(A)$, and every $U \in \mathrm{WO}(A)$ is similar with a proper, initial segment of $\chi(A)$.

Proof. We verify first the identity

$$\mathbf{seg}_{\chi(A)}(\alpha) = \{[\mathbf{seg}_U(x)/\sim_A] \mid x \in U\}.$$

If $\beta = [V/\sim_A] <_{\chi(A)} \alpha$, then $V <_o U$ from (7-31), and hence $V =_o \mathbf{seg}_U(x)$ for some $x \in U$, so that $\beta = [\mathbf{seg}_U(x)/\sim_A]$. Conversely, for each $x \in U$, $\mathbf{seg}_U(x) <_o U$, hence, $[\mathbf{seg}_V(x)/\sim_A] <_{\chi(A)} [U/\sim_A] = \alpha$, again by (7-31).

To show the similarity

$$U =_o \mathbf{seg}_{\chi(A)}(\alpha) = \{[\mathbf{seg}_U(x)/\sim_A] \mid x \in U\},$$

define $\rho : U \to h(A)$ by

$$\rho(x) = [\mathbf{seg}_U(x)/\sim_A], \quad (x \in U);$$

now ρ is a similarity of U with the image $\rho[U]$, because

$$x <_U y \iff \mathbf{seg}_U(x) \subsetneq \mathbf{seg}_U(y)$$
$$\iff \mathbf{seg}_U(x) <_o \mathbf{seg}_U(y)$$
$$\iff [\mathbf{seg}_U(x)/\sim_A] <_{\chi(A)} [\mathbf{seg}_U(y)/\sim_A]. \quad \dashv \text{(Lemma)}$$

Suppose now, towards a contradiction, that there exists an injection

$$\pi : h(A) \rightarrowtail A,$$

and let $B = \pi[h(A)] \subseteq A$ be its image. The injection π copies the wellordering of $h(A)$ to a wellordering of B,

$$x \leq_B y \iff_{\mathrm{df}} \pi^{-1}(x) \leq_{\chi(A)} \pi^{-1}(y) \quad (x, y \in B),$$

so that $U = (B, \leq_B)$ is a well ordered subset of A, and by its definition,

$$U =_o \chi(A). \qquad (7\text{-}32)$$

But U is similar with a proper initial segment of $\chi(A)$ by the Lemma, and hence $U <_o \chi(A)$, which contradicts $U =_o \chi(A)$.

To show the minimality of $\chi(A)$, notice that for any well ordered set W, if $W <_o \chi(A)$, then $W =_o \mathbf{seg}_{\chi(A)}(\alpha)$ for some $\alpha = [U/\sim_A]$, so that $W =_o U$ by the Lemma. Thus,

$$\text{if } W <_o \chi(A), \text{ then } W \leq_c A, \qquad (7\text{-}33)$$

since (the field of) U is a subset of A and similarities are injections. Taking the negation of both sides,

$$\text{if } W \not\lesssim_c A, \text{ then } \neg[W <_o \chi(A)], \text{ hence } \chi(A) \leq_o W. \qquad\qquad \dashv$$

Of course, we would like to prove that $A <_c \chi(A)$ instead of the timid $\chi(A) \not\lesssim_c A$, and this is certainly true, but its proof depends on the Axiom of Choice. Wait for a bit until we finally bring the Deus ex Machina onto the stage.

The annoying details of this proof are forced on us by the fact that the restriction $\lesssim_A$ of the definite condition $\leq_o$ to $\mathrm{WO}(A)$ is not a wellordering, for the trivial reason that it is not antisymmetric: there may exist distinct, similar $U, V \in \mathrm{WO}(A)$, in fact they always do, if A has more than one element. This is why we were forced to take $h(A)$ as a set of equivalence classes rather than simply set $h(A) = \mathrm{WO}(A)$. Technically, $\lesssim_A$ is a *prewellordering* (ugh!) and it is worth recasting the argument in a different form, after introducing this notion. See Problems **x7.17 – x7.20**.

The Hartogs operation can be used to construct general *infimum* and *supremum* operation for families of well ordered sets (Problems **x7.26** and **x7.27**), and it has many other interesting properties. We use it next to extend the Continuous Least Fixed Point Theorem **6.21** to discontinuous mappings. Let us first put down, for the record, the Fixed Point Theorem for expansive mappings, which we have already discussed.

7.35. Fixed Point Theorem (Zermelo).[18] *Every expansive mapping $\pi : P \to P$ on an inductive poset has at least one fixed point, i.e., some $x^* \in P$ satisfies the equation*

$$x^* = \pi(x^*).$$

Proof. The argument given in **7.26** needs only some well ordered set U which cannot be injected into P, and $U = \chi(P)$ does it. $\dashv$

7.36. Least Fixed Point Theorem. *Every monotone mapping $\pi : P \to P$ on an inductive poset has a (unique) least fixed point, i.e., for some $x^* \in P$,*

$$\pi(x^*) = x^*,$$
$$(\forall y \in P)[\pi(y) = y \Longrightarrow x^* \leq y].$$

In fact, x^* satisfies the following, strong minimality property:

$$(\forall y \in P)[\pi(y) \leq y \Longrightarrow x^* \leq y]. \qquad\qquad (7\text{-}34)$$

Proof. A careful examination of the proof of the Iteration Lemma **7.25** and the proof of the Fixed Point Theorem **7.26** reveals that exactly the same construction of the fixed point for an expansive mapping works and yields

---

[18]Zermelo did not formulate the Fixed Point Theorem in this generality, which is why it and many of its Corollaries have been attributed at various times to later mathematicians. But the famous "first proof" of the *Wellordering Theorem* which Zermelo gave in 1904 proves exactly this result, trivially restricted to the special case which interested him.

the least fixed point of a monotone mapping. However, it is not necessary
to do this, as the Least Fixed Point Theorem is an easy consequence of the
Fixed Point Theorem. The basic idea is to observe that the given monotone
mapping π is necessarily expansive on some inductive *sub-poset* of P.

Let

$$Q = \{x \in P \mid x \le \pi(x) \,\&\, (\forall y)[\pi(y) \le y \Longrightarrow x \le y]\}$$

and observe first that the restriction

$$\le_Q = \{(x, y) \mid x, y \in Q \,\&\, x \le_P y\}$$

of $\le_P$ to Q is also a partial ordering—this is automatically true for the re-
striction of $\le_P$ to any subset of P. (We skip the subscripts $_P$ and $_Q$ for the
remaining of the argument.) In addition, $\pi[Q] \subseteq Q$, because

$$x \le \pi(x) \Longrightarrow \pi(x) \le \pi(\pi(x)),$$

and for every y,

$$\pi(y) \le y \,\&\, x \le y \Longrightarrow \pi(x) \le \pi(y) \le y,$$

by the monotonicity of π. It follows that the restriction

$$\pi_Q = \{(x, \pi(x)) \mid x \in Q\}$$

of π to Q is a mapping on Q, it continues to be monotone (of course) and
it is also expansive, because of the definition of Q. To apply the Fixed Point
Theorem **7.35** to Q and π_Q we need the following.

Lemma. *The poset Q is inductive.*

Proof. It is enough to show that for every chain $S \subseteq Q$, the least upper
bound $M = \sup S$ (which exists in P because P is inductive) is a member of
Q, i.e., (1) $M \le \pi(M)$, and (2) for every y, $\pi(y) \le y \Longrightarrow M \le y$. For (1) we
compute:

$$
\begin{aligned}
x \in S \implies & x \le M, && \text{because } M \text{ is an upper bound of } S, \\
\implies & \pi(x) \le \pi(M), && \text{because } \pi \text{ is monotone,} \\
\implies & x \le \pi(x) \le \pi(M), && \text{because } x \in S \subseteq Q,
\end{aligned}
$$

and therefore $\pi(M)$ is an upper bound of S and we have $M = \sup S \le \pi(M)$.
(2) follows from the observation that every y such that $\pi(y) \le y$ is an upper
bound of Q (from Q's definition), and therefore an upper bound of the smaller
set $S \subseteq Q$, so that $M = \sup S \le y$. $\dashv$ (Lemma)

By the Fixed Point Theorem **7.35** now, there exists some $x^* \in Q$, such that
$\pi(x^*) = x^*$ and (7-34) holds simply because $x^* \in Q$.

That there is at most one least fixed point is obvious: if y^* is also a least
fixed point, then $x^* \le y^*$ and $y^* \le x^*$, so $x^* = y^*$. $\dashv$

The full Least Fixed Point Theorem frees us from the necessity to check
continuity in the applications of least fixed points to computer science, the
fixpoint theory of programs. This is nice, but not very important, since (as
we observed in Chapter **6**) the mappings which come up in algorithmic appli-
cations are typically continuous "by inspection". However, the theorem has
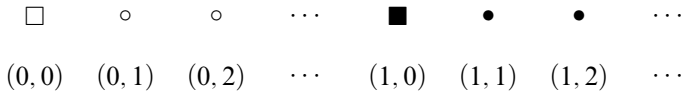
FIGURE 7.5. The sum $\mathbb{N} +_o \mathbb{N}$.

more significant, deeper applications to the general theory of sets, particularly in the study of *definability* in set theory as well as the construction of examples and counterexamples with specified properties. We will encounter several of these in the chapters which follow.

Problems for Chapter 7

x7.1. Every linear ordering of a finite set is a wellordering. (See the related Problem **x6.7.**)

7.37. The **sum** $P +_o Q$ of two posets P and Q is obtained by placing disjoint copies of P and Q side-by-side, every point of P preceding every point of Q. Formally, we set $P +_o Q = R$, where

$$\text{Field}(R) =_{\text{df}} (\{0\} \times \text{Field}(P)) \cup (\{1\} \times \text{Field}(Q)), \qquad (7\text{-}35)$$

and for $(i, x), (j, y) \in \text{Field}(R)$,

$$(i, x) \leq_R (j, y) \iff_{\text{df}} i < j \vee [i = j = 0 \,\&\, x \leq_U y] \\ \vee [i = j = 1 \,\&\, x \leq_V y]. \qquad (7\text{-}36)$$

The idea is that P is similar with the set $\{0\} \times \text{Field}(P)$ partially ordered by its second elements, by the obvious similarity $(x \mapsto (0, x))$, and (again) $Q =_o \{1\} \times \text{Field}(Q)$.

x7.2. If $P =_o P'$ and $Q =_o Q'$, then $P +_o Q =_o P' +_o Q'$.

x7.3. For all posets P, $\text{Succ}(P) =_o P +_o [0, 1)$.

x7.4. For all posets P, Q, R,

$$P +_o (Q +_o R) =_o (P +_o Q) +_o R.$$

x7.5. If U and V are well ordered sets, then so is their sum $U +_o V$.

x7.6. Prove that $[0, 1) +_o \mathbb{N} =_o \mathbb{N} \neq_o \mathbb{N} +_o [0, 1)$, so that the addition operation on well ordered sets is not commutative.

7.38. The **product** $P \cdot_o Q$ of two posets is obtained by replacing each point of Q by a copy of P. Formally, we let $P \cdot_o Q = R$, where

$$\text{Field}(R) = \text{Field}(P) \times \text{Field}(Q), \qquad (7\text{-}37)$$

and $\leq_R$ is the *inverse lexicographic* ordering of pairs i.e., we compare the second members first: for $(x_1, y_1), (x_2, y_2) \in \text{Field}(R)$,

$$(x_1, y_1) \leq_R (x_2, y_2) \iff_{\text{df}} y_1 <_Q y_2 \vee [y_1 = y_2 \,\&\, x_1 \leq_P x_2]. \qquad (7\text{-}38)$$

A :   ● ● ⋯

$A \times B$:   ● ● ⋯   ● ● ⋯   ● ● ⋯   ● ● ⋯   ⋯

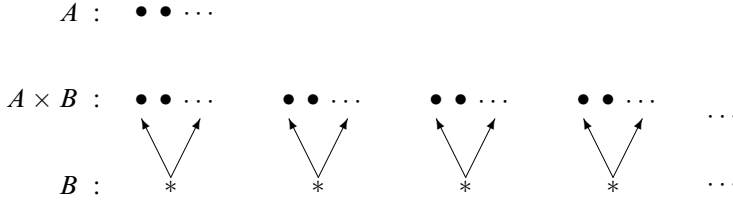B :   *   *   *   *   ⋯

FIGURE **7.6**. The product of two well ordered sets.

There is no special reason for ordering pairs by looking at their second members first, it is just that Cantor chose to do it this way and it has stuck.

x7.7. If $P =_o P'$ and $Q =_o Q'$, then $P \cdot_o Q =_o P' \cdot_o Q'$.

x7.8. Prove that $P \cdot_o [0, 2) =_o P +_o P$, but $[0, 2) \cdot_o \mathbb{N} =_o \mathbb{N} \neq_o \mathbb{N} \cdot_o [0, 2)$, so that multiplication of well ordered sets is not commutative.

x7.9. For all posets P, Q, R,

$$P \cdot_o (Q \cdot_o R) =_o (P \cdot_o Q) \cdot_o R.$$

x7.10. The product of two well ordered sets is well ordered.

x7.11. For each well ordered set U, there exists exactly one function

$$\text{Parity} : U \to \mathbb{N},$$

such that $\text{Parity}(y) = 0$ if $y = 0$ or y is a limit point, and at successor points,

$$\text{Parity}(S(x)) = 1 - \text{Parity}(x).$$

x7.12. Every point y in a well ordered set U can be expressed uniquely in the form

$$y = S^n(x), \tag{7-39}$$

where (1) x is either the minimum 0 or a limit point, (2) n is a natural number, and (3) the function $(i, x) \mapsto S^i(x)$ is defined by the recursion

$$S^0(x) = x, \quad S^{i+1}(x) = S(S^i(x)).$$

x7.13. For any two well ordered sets U, V, there exists at most one initial similarity $\pi : U \rightarrowtail \pi[U] \sqsubseteq V$.

x7.14. For all well ordered sets U, V, W,

$$U <_o V \ \& \ V \leq_o W \implies U <_o W,$$
$$U \leq_o V \ \& \ V <_o W \implies U <_o W.$$

x7.15. If κ and λ are well orderable cardinal numbers, then either $\kappa \leq_c \lambda$ or $\lambda \leq_c \kappa$.

*__x7.16.__ If κ is a well orderable, infinite cardinal number, then $\kappa + 1 =_c \kappa$.

	(0, 0)	(1, 0)	(2, 0)	$\cdots$	(0, 1)	(1, 1)	(2, 1)	$\cdots$
$N \cdot_o [0, 2):$	□	○	○	$\cdots$	■	●	●	$\cdots$

	(0, 0)	(1, 0)	(0, 1)	(1, 1)	(0, 2)	$\cdots$
$[0, 2) \cdot_o N:$	□	○	○	○	○	$\cdots$

FIGURE 7.7. Multiplication of well ordered sets is not commutative.

7.39. Definition. A **prewellordering** on a set A is any relation $\precsim \subseteq A \times A$ which is reflexive, transitive, connected (total) and grounded. "Connected" means that any two points in A are comparable,

$$(\forall x, y \in A)[x \precsim y \lor y \precsim x],$$

and "grounded" means that every non-empty $X \subseteq A$ has a $\precsim$-least member,

$$(\forall X \subseteq A, X \neq \emptyset)(\exists x \in X)(\forall y \in X)[x \precsim y].$$

A prewellordering would be a wellordering, if only it were antisymmetric.

x7.17. For each set A, consider the set

$$B = \{X \subseteq A \mid X \text{ is finite}\}$$

of all finite subsets of A, and set on B

$$X \precsim_B Y \iff_{df} X \leq_c Y.$$

Prove that $\precsim_B$ is a prewellordering.

x7.18. A relation $\precsim \subseteq A \times A$ is a prewellordering if and only if there exists a well ordered set $U = (\text{Field}(U), \leq_U)$ and a surjection $\pi : A \twoheadrightarrow \text{Field}(U)$ such that

$$x \precsim y \iff \pi(x) \leq_U \pi(y) \quad (x, y \in A).$$

x7.19. For each set A, the relation

$$U \precsim_A V \iff_{df} U, V \in \text{WO}(A) \,\&\, U \leq_o V$$

is a prewellordering of $\text{WO}(A)$.

x7.20. Rework the proof of Hartogs' Theorem by applying the preceding two problems.

x7.21. For every set A, there exists a well ordered set V such that there exists no surjection $\pi : A \twoheadrightarrow V$.

x7.22. Prove that if $A \leq_c B$, then $\chi(A) \leq_o \chi(B)$.

x7.23. Prove that $\chi([0, n)) =_o [0, n + 1)$.

x7.24. If W is a well ordered set and $W \leq_c A$, then $W <_o \chi(A)$.

x7.25. For each set A and each well ordered set U,
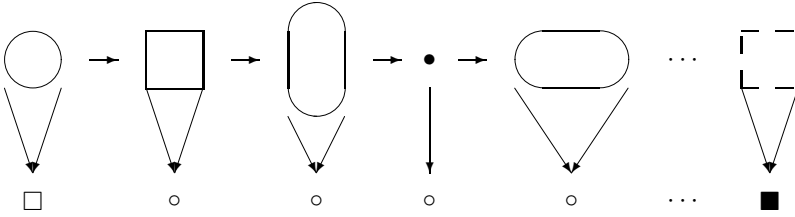
$$U <_o \chi(A) \iff \text{Field}(U) \leq_c A.$$

FIGURE **7.8**. Portrait of a prewellordering.

The operation $\chi(A)$ is definite, as we gave an explicit definition of the field $h(A)$ and the wellordering $\leq_{\chi(A)}$ of $\chi(A)$ from A. We can use it to define some related, "infinitary" operations on families of well ordered sets.

*__x7.26.__ Define a definite operation $\inf(\mathscr{E})$, such that for every non-empty family $\mathscr{E}$ of well ordered sets, $\inf(\mathscr{E})$ has the following properties:

(1) $\inf(\mathscr{E})$ is a well ordered set.
(2) For some $U \in \mathscr{E}$, $\inf(\mathscr{E}) =_o U$.
(3) For every $U \in \mathscr{E}$, $\inf(\mathscr{E}) \leq_o U$.

HINT. Look for $\inf(\mathscr{E})$ in the initial segments of $\chi(\bigcup \mathscr{E})$.

*__x7.27.__ Define a definite operation $\sup(\mathscr{E})$, such that for every non-empty family $\mathscr{E}$ of well ordered sets, $\sup(\mathscr{E})$ has the following properties:

(1) $\sup(\mathscr{E})$ is a well ordered set.
(2) If $U \in \mathscr{E}$, then $U \leq_o \sup(\mathscr{E})$.
(3) If W is a well ordered set and for each $U \in \mathscr{E}$, $U \leq_o W$, then $\sup(\mathscr{E}) \leq_o W$.

*__x7.28.__ Let $\leq$ be a linear ordering of a set A and define on the poset $\mathcal{P}(A)$ the mapping
$$\pi(X) =_{\mathrm{df}} \{y \in A \mid (\forall x < y)[x \in X]\}.$$
Verify that $\pi : \mathcal{P}(A) \to \mathcal{P}(A)$ is monotone and give an example where it is not countably continuous. Prove that if A_w is the least fixed point of π, then
$$x \in A_w \iff \{(s, t) \in A \times A \mid s \leq t < x\} \text{ is a wellordering.}$$

There are situations where it is easier to use *the proof* of the Fixed Point Theorem **7.35** rather than its statement.

*__x7.29. Detailed Fixed Point Theorem.__ For each expansive or monotone mapping $\pi : P \to P$ on an inductive poset P, there exists a subset $D \subseteq P$ with the following properties:

1. D is a well ordered chain in P.
2. Every member of D is determined from its predecessors by the formula
$$x = \pi(\sup\{y \in D \mid y < x\}).$$
3. No point in D is a fixed point of π.

4. The point $\pi(\sup D)$ is a fixed point of π,

$$\pi(\pi(\sup D)) = \pi(\sup D).$$

5. If π is monotone, then $\pi(\sup D)$ is the least fixed point of π.

Prove also that these conditions determine D uniquely.

*$\mathbf{x7.30.}$ Suppose P and Q are inductive posets and $\pi : P \times Q \to P$ is a monotone mapping on the product and define the mapping $\rho : Q \to P$ by appealing to Problem $\mathbf{x6.4}$ and the Least Fixed Point Theorem $\mathbf{7.36}$,

$$\begin{aligned}\rho(y) &= (\mu x \in P)[\pi(x, y) = x]\\&= \text{the least fixed point of } \pi(x, y) = x.\end{aligned} \tag{7-40}$$

Prove that ρ is a monotone mapping, and if π is countably continuous, then so is ρ.

*$\mathbf{x7.31.}$ **Bekič-Scott Rule**. Suppose P_1, P_2 are inductive posets, and

$$\pi_1 : P_1 \times P_2 \to P_1, \quad \pi_2 : P_1 \times P_2 \to P_2$$

are monotone mappings. Using the μ-notation for least fixed points of (7-40), let

$$\rho(x_2) = (\mu x_1 \in P_1)[\pi_1(x_1, x_2) = x_1],$$

let

$$\overline{x}_2 = (\mu x_2 \in P_2)[\pi_2(\rho(x_2), x_2) = x_2]$$

be the least fixed point of the mapping $x_2 \mapsto \pi_2(\rho(x_2), x_2)$ (which is monotone by $\mathbf{x7.30}$) and finally let

$$(x_1^*, x_2^*) = (\mu(x_1, x_2) \in P_1 \times P_2)[(\pi_1(x_1, x_2), \pi_2(x_1, x_2)) = (x_1, x_2)]$$

be the least fixed point in the product poset. Prove that

$$x_2^* = \overline{x}_2.$$

The problem insures that we can compute simultaneous least fixed points by iterating the least fixed point operation $(\mu x \in P)$ on one inductive poset at a time.

CHAPTER 8

CHOICES

8.1. The Axiom of Choice, **AC**. For any two sets A, B and any binary relation $P \subseteq (A \times B)$,

$$(\forall x \in A)(\exists y \in B)P(x, y) \Longrightarrow (\exists f : A \to B)(\forall x \in A)P(x, f(x)). \quad (8\text{-}1)$$

This is the last and most controversial axiom of Zermelo. To understand how such an axiom might be needed, consider the classical example of Russell, where A is a set of pairs of shoes, $B = \bigcup A$ and

$$P(x, y) \iff y \in x.$$

The function

$$f(x) =_{\mathrm{df}} \text{the left shoe in } x, \quad (x \in A)$$

obviously selects a shoe from each pair, in symbols $(\forall x \in A)P(x, f(x))$. If, however, A is a set of pairs of *socks*, then there is no obvious way to *define* a function $f : A \to \bigcup A$ which selects one sock $f(x) \in x$ from each pair, because (as we stipulate for the example), a pair of socks comprises precisely two perfectly identical objects. We can still prove that a selector function f exists when A is finite, by induction on the number of elements in A (Problem **x8.1**). But in mathematics we can imagine infinite sets of pairs of socks, and in that case we need something like the Axiom of Choice to guarantee the existence of such a function.

Less amusing but more significant for mathematics is the proof of the basic theorem **2.10**, where we considered a sequence A_0, A_1, ... of countable sets and began with the phrase

> It is enough to prove the theorem in the special case where none of the A_n is empty, in which case we can find for each A_n an enumeration $\pi_n : \mathbb{N} \twoheadrightarrow A_n$.

Perhaps for each n "we can find" (i.e., "there exists") some enumeration π of A_n, but the rest of the proof needs a *function* $(n \mapsto \pi_n)$ which associates a specific enumeration π_n with each n: which of the axioms $(\mathbf{I}) - (\mathbf{VI})$ can be used to prove the existence of such a function? Here $A = \mathbb{N}$, $B = (\mathbb{N} \to \bigcup_{n=0}^{\infty} A_n)$ and

$$P(n, \pi) \iff \pi : \mathbb{N} \twoheadrightarrow A_n,$$

so that $(\forall n \in \mathbb{N})(\exists \pi \in B)P(n, \pi)$ from the hypothesis that each A_n is non-empty and countable: and the Axiom of Choice guarantees precisely that there
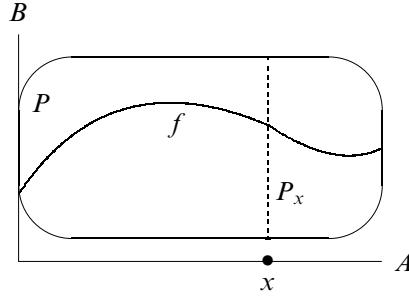
FIGURE **8.1**. A selector for $P \subseteq A \times B$.

exists a function $f : \mathbb{N} \to B$ such that for each $n \in \mathbb{N}$, the value $f(n) = \pi_n$ satisfies $P(n, \pi_n)$, i.e., it enumerates A_n. Such "silent" appeals to the Axiom of Choice, masked by the notation, are very common in mathematics and especially in analysis, where the classical theory of limits and continuous functions cannot be developed in a satisfactory way without choices.

If we picture $P \subseteq A \times B$ as a subset of the product space, then the hypothesis $(\forall x \in A)(\exists y \in B) P(x, y)$ means that the *fiber* or *section*

$$P_x =_{\mathrm{df}} \{y \in B \mid P(x, y)\} \tag{8-2}$$

above each $x \in A$ is non-empty; the Axiom of Choice guarantees the existence of a *selector* for P, a function $f : A \to B$ which assigns to each $x \in A$ exactly one point in the fiber above it. There are two other, simple reformulations of the axiom which express in different ways the process of "collecting into a whole" any number of unrestricted, non-conflicting choices.

8.2. Definition. A set S is a **choice set** for a family of sets $\mathscr{E}$ if (1) $S \subseteq \bigcup \mathscr{E}$, and (2) for every $X \in \mathscr{E}$, the intersection $S \cap X$ is a singleton. A choice set S *selects* from each $X \in \mathscr{E}$ the unique member of the intersection $S \cap X$.

8.3. Exercise. *If $\emptyset \in \mathscr{E}$, then $\mathscr{E}$ does not admit a choice set. Also, if $a \neq b$, then the family $\mathscr{E} = \{\{a\}, \{a, b\}, \{b\}\}$ does not admit a choice set.*

8.4. Theorem. *The Axiom of Choice is equivalent to the following proposition: every family $\mathscr{E}$ of non-empty and pairwise disjoint sets admits a choice set.*

This is the version of **AC** postulated by Zermelo.

PROOF. Assume first the Axiom of Choice and let $U = \bigcup \mathscr{E}$ be the union of the given family of pairwise disjoint, non-empty sets, which means that

$$(\forall X \in \mathscr{E})(\exists x \in U)[x \in X].$$

The Axiom of Choice guarantees that there exists a function $f : \mathscr{E} \to U$, such that

$$(\forall X \in \mathscr{E})[f(X) \in X];$$

we set $S = f[\mathscr{E}] = \{f(X) \mid X \in \mathscr{E}\}$, and the fact that the members of $\mathscr{E}$ are pairwise disjoint implies easily that S intersects every member of $\mathscr{E}$ in a singleton.

For the converse, assume

$$(\forall x \in A)(\exists y \in B)P(x, y),$$

set for each $x \in A$

$$U_x = \{(t, y) \in P \mid t = x\},$$

and let

$$\mathscr{E} = \{U_x \mid x \in A\}.$$

Each member of $\mathscr{E}$ is non-empty by the hypothesis and it is determined by the constant, first member of the pairs in it, so any two members of $\mathscr{E}$ are disjoint. If S is a choice set for this $\mathscr{E}$, then the function

$$f(x) = \text{the unique } y \text{ such that } (x, y) \in S$$

easily satisfies the conclusion of the Axiom of Choice. $\dashv$

8.5. Definition. A **choice function** for a set A is any partial function

$$\varepsilon : \mathcal{P}(A) \rightharpoonup A,$$

such that

$$\emptyset \neq X \subseteq A \Longrightarrow \varepsilon(X){\downarrow} \,\&\, \varepsilon(X) \in X.$$

8.6. Lemma. *The Axiom of Choice is equivalent to the assertion that every set admits a choice function.*

PROOF. For every A, obviously $(\forall X \in \mathcal{P}(A) \setminus \{\emptyset\})(\exists y \in A)[y \in X]$; and so, directly from the Axiom of Choice, there must exist some function $\varepsilon : \mathcal{P}(A) \setminus \{\emptyset\} \to A$ such that

$$(\forall X \in \mathcal{P}(A) \setminus \{\emptyset\})[\varepsilon(X) \in X].$$

The converse is easy enough to leave for an exercise. $\dashv$

8.7. Exercise. *If every set admits a choice function, then the Axiom of Choice is true.*

8.8. But is it true? (1) Naively understood, the Axiom of Choice asserts that if each of a set of non-conflicting choices is possible, then they can all be made independently and their results collected into a completed whole, a set. By this understanding it is quite obvious, it can be justified by the natural interpretation we would give to the Powerset Axiom: when we grant sethood to the class $\{X \mid X \subseteq A\}$ of all subsets of A, we truly mean *all* subsets of A, including those for which the membership criterion is not determined by some explicit law but by free choice, by chance if you will.

The Axiom of Choice is different in form from the earlier "constructive" axioms (**II**) – (**VI**), because it *postulates directly* the existence of a set for which it does not supply a definition. Each of (**II**) – (**VI**) grants sethood to a specific, explicitly defined collection of objects, it legitimizes a special case of the most appealing (if false) General Comprehension Principle **3.3**. *The Axiom of Choice is the only Zermelo axiom other than Extensionality which is not a special case of the General Comprehension Principle.* This is misstated on occasion, to make the claim that the Axiom of Choice is the only one which

demands the existence of objects for which it does not supply a definition, which is not true: the Extensionality and Powerset Axioms do the same, in a more fundamental if indirect manner.

Zermelo introduced the Axiom of Choice explicitly in 1904, in a brief paper in which he used it to prove that every set is well orderable. This was a long-standing conjecture, and Cantor had outlined a proof of it in a letter to Dedekind, then still unpublished. His proof, however (and the related proof of the Cardinal Comparability Hypothesis), depended on intuitions about sets which were not sufficiently explained. In contrast to this, Zermelo made it clear, from the start, that his own detailed proof depended on the Axiom of Choice, and he was immediately attacked for this by some of the leading mathematicians of the time, for introducing a questionable method to derive an implausible conclusion. Given the fact that choice principles were by no means new to mathematics and that they permeate Cantor's earlier reasoning, it is fair to say that the shock was caused more by the realization of the power of the axiom than by its meaning.

In the next, fundamental result, we establish the somewhat surprising equivalence of the Axiom of Choice with two very different set theoretic claims. We have kept the traditional names for these propositions—Axiom, Hypothesis, Theorem—which have been attached to them by the historical accident of when and how they were introduced in the mathematical literature.

8.9. Theorem. *The following propositions are all equivalent.*

(1) **Axiom of Choice**: *Every set admits a choice function.*

(2) **Hypothesis of Cardinal Comparability**: *For any two sets A, B, either $A \leq_c B$ or $B \leq_c A$.*

(3) **Wellordering Theorem**: *Every set is well orderable.*

Proof. We verify, round-robin style, that each of the first two propositions implies the next, and finally $(3) \implies (1)$.

$(1) \implies (2)$. By Exercise **6.17**, the poset $(A \rightarrowtail B)$ of partial injections from A to B is inductive. The idea is to define (using **AC**) an expansive mapping on $(A \rightarrowtail B)$ which extends properly each partial injection $p : A \rightarrowtail B$ whose domain does not exhaust A and whose range does not exhaust B; any fixed point of that mapping then will either be defined on all of A, witnessing that $A \leq_c B$, or its image will exhaust B, witnessing that $B \leq_c A$.

In detail, let

$$\varepsilon_A : \mathcal{P}(A) \rightharpoonup A, \quad \varepsilon_B : \mathcal{P}(B) \rightharpoonup B$$

be choice functions on A and B, supplied by the Axiom of Choice, and define π on $(A \rightarrowtail B)$ as follows:

$$\pi(p) = \begin{cases} p \cup \{(\varepsilon_A(A \setminus \mathrm{Domain}(p)), \varepsilon_B(B \setminus \mathrm{Image}(p)))\}, \\ \qquad \text{if } \mathrm{Domain}(p) \subsetneq A \,\&\, \mathrm{Image}(p) \subsetneq B, \\ p, \qquad \text{otherwise.} \end{cases}$$

This mapping is obviously expansive, and so by the Fixed Point Theorem **7.35**, it has a fixed point $p^* : A \rightharpoonup B$, for which

$$\text{either Domain}(p^*) = A \text{ or Image}(p^*) = B.$$

CASE 1, Domain$(p^*) = A$. Now $p^* : A \rightarrowtail B$ is an injection of A into B, and so $A \leq_c B$.

CASE 2, Image$(p^*) = B$. In this case, p^* is a bijection of Domain(p^*) with B, and so $B =_c \text{Domain}(p^*) \subseteq A$ and hence $B \leq_c A$.

(2) $\implies$ (3). By the Cardinal Comparability Hypothesis, either $A \leq_c h(A)$ or $h(A) \leq_c A$, where $h(A)$ is the Hartogs set for A; but $h(A) \not\leq_c A$ by Hartogs' Theorem **7.34**, and so $A \leq_c h(A)$; in addition, $h(A)$ is well orderable, and hence A is also wellorderable by Exercise **7.3**.

(3) $\implies$ (1). If $\leq$ is a wellordering of A, then the partial function

$$\varepsilon(X) = \text{the } \leq\text{-least member of } X$$

is a choice function for A. ⊣

8.10. But is it true? (2) The import of this theorem is that if we accept the basic, constructive first six axioms of Zermelo, then the Axiom of Choice, the Hypothesis of Cardinal Comparability and the Wellordering Theorem express in three different ways the same set theoretic principle. No doubt, the Axiom of Choice is the most direct and intuitive formulation of this principle, the one which makes it most obvious that it is true. The Cardinal Comparability Hypothesis is certainly easy to understand and plausible, but few would propose it as an axiom: it has the feel of a mathematical claim which ought to be proved. Finally, the Wellordering Theorem is crystal clear in its meaning and it gives a mechanism for making choices which "explains" in some way the Axiom of Choice; but far from being obvious, it raises a flag of caution. For example, what does a wellordering of the powerset of the natural numbers $\mathcal{P}(\mathbb{N})$ look like? Without some thought, it is not even obvious that $\mathcal{P}(\mathbb{N})$ admits linear orderings (see Problem **x8.9**). It is quite difficult to imagine the structure of the beast, and this naturally casts doubt on the truth of the axiom which implies its existence. It is hardly surprising that the commotion about the Axiom of Choice was caused by Zermelo's proof of the implication (1) $\implies$ (3), whose conclusion is still thought by many to be counterintuitive.

In addition to the Cardinal Comparability Hypothesis and the Wellordering Theorem which are fundamental for the development of set theory, the Axiom of Choice is equivalent with a host of other propositions which are important in other areas of mathematics. We include here just two of them, which are closest to our subject and easy to prove by the methods we have been studying, but there are many others.[19]

---

[19]Kenneth Hoffman once declared in a class that the Tychonoff Theorem of general topology is "obviously" equivalent to the Axiom of Choice, "since all fishy, general principles are equivalent to **AC**".

8.11. Theorem. *The Axiom of Choice is equivalent with the following two propositions*:

(1) **Maximal Chain Principle**: *Every poset P has a chain $S \subseteq P$ which is maximal, in the sense that for every other chain S', $S \subseteq S' \implies S = S'$.*

(2) **Zorn's Lemma**: *If every chain in a poset P has an upper bound, then P has at least one maximal element.*

Proof. **AC** $\implies$ (1). Assume **AC**, let $(\mathrm{Chains}(P), \subseteq)$ be the poset of all chains in P, which is inductive by Proposition **6.14**, and assume towards a contradiction that there is no maximal chain, so that by **AC**,

$$(\forall S \in \mathrm{Chains}(P))(\exists S' \in \mathrm{Chains}(P))[S \subsetneq S'];$$

now **AC** gives us a mapping $\pi : \mathrm{Chains}(P) \to \mathrm{Chains}(P)$ which is expansive but has no fixed point, contradicting the Fixed Point Theorem **7.35**.

(1) $\implies$ (2). If S is a maximal chain in P and M is an upper bound of S, both guaranteed by the hypotheses, then M is maximal in P—because if $M <_P M'$, then $S \cup \{M'\}$ would be a chain which extends S strictly.

(2) $\implies$ **AC**. For any two sets A, B, consider the poset $(A \rightarrowtail B)$ of partial injections. This is inductive, by Exercise **6.17**, and so every chain has an upper bound in it; by Zorn's Lemma then, $(A \rightarrowtail B)$ has a maximal element f, which (as in the proof of (1) $\implies$ (2) in Theorem **8.9**) establishes that either $A \leq_c B$ or $B \leq_c A$, which in turn implies **AC**. $\dashv$

We now consider two easy corollaries of the Axiom of Choice which express simpler principles of choice.

8.12. Countable Principle of Choice, $\mathbf{AC}_\mathbb{N}$. For each set B and each binary relation $P \subseteq \mathbb{N} \times B$ between natural numbers and members of B,

$$(\forall n \in \mathbb{N})(\exists y \in B)P(n, y) \implies (\exists f : \mathbb{N} \to B)(\forall n \in \mathbb{N})P(n, f(n)).$$

8.13. (**VII**) **Axiom of Dependent Choices**, **DC**. For each set A and each relation $P \subseteq A \times A$,

$$a \in A \ \& \ (\forall x \in A)(\exists y \in A)P(x, y)$$
$$\implies (\exists f : \mathbb{N} \to A)[f(0) = a \ \& \ (\forall n \in \mathbb{N})P(f(n), f(n+1))].$$

In contrast to the full Axiom of Choice which demands the existence of choice functions $f : A \to B$ for arbitrary A, B, the Countable Principle of Choice $\mathbf{AC}_\mathbb{N}$ justifies only a sequence of independent choices from an arbitrary set B which successively satisfy the conditions

$$P(0, f(0)), \ P(1, f(1)), \ P(2, f(2)), \cdots$$

The Axiom of Dependent Choices **DC** also justifies only a sequence of choices, where, however, each of them may depend on the previous one, since they must now satisfy the conditions

$$P(f(0), f(1)), \ P(f(1), f(2)), \ P(f(2), f(3)), \cdots$$

It is easily equivalent to the following, seemingly stronger principle which allows each choice to depend on all the preceding ones.

8.14. Proposition. *The Axiom of Dependent Choices is equivalent to the following proposition*: *for every set A and every relation $P \subseteq A^* \times A$ between strings from A and members of A*,

$$(\forall u \in A^*)(\exists x \in A)P(u, x) \Longrightarrow (\exists f : \mathbb{N} \to A)(\forall n)P(\overline{f}(n), f(n)).$$

PROOF. The implication from this version of **DC** to the "official" one is easy and we leave it for an exercise. Assuming now **DC** and the hypothesis of the seemingly stronger version, define on A^* the relation

$$Q(u, v) \Longleftrightarrow_{\text{df}} (\exists x \in A)[v = u \star \langle x \rangle \,\&\, P(u, x)];$$

we obviously have $(\forall u \in A^*)(\exists v \in A^*)Q(u, v)$, **DC** gives us a function $g : \mathbb{N} \to A^*$ such that $g(0) = \emptyset$ and $(\forall n)Q(g(n), g(n+1))$, and the function we need is $f = \bigcup g$, for which $f(n) = g(n+1)(n)$ and $\overline{f}(n) = g(n)$. $\dashv$

8.15. Exercise. *Show the direction of Proposition* **8.14** *omitted in the given proof.*

8.16. Theorem. (1) *The Axiom of Choice implies the Axiom of Dependent Choices.*

(2) *The Axiom of Dependent Choices implies the Countable Principle of Choice.*

PROOF. (1) Let $\varepsilon : \mathcal{P}(A) \setminus \{\emptyset\} \to A$ be a choice function for A and assume the hypothesis of **DC**. The function $f : \mathbb{N} \to A$ that we need for the conclusion is defined by the recursion

$$f(0) = a,$$
$$f(n+1) = \varepsilon(\{y \in A \mid P(f(n), y)\}).$$

(2) Assume the hypothesis of the Countable Principle of Choice, let $A = \mathbb{N} \times B$, let $a = (0, b)$ where $b \in B$ is any point satisfying $P(0, b)$, and define on A the relation

$$Q((n, x), (m, y)) \Longleftrightarrow_{\text{df}} m = n + 1 \,\&\, P(m, y).$$

The function $f : \mathbb{N} \to \mathbb{N} \times B$ supplied by **DC** for this a and Q takes pairs as values, so $f(n) = (g(n), h(n))$, $g(0) = 0$, $h(0) = b$ for suitable functions g, h, and for every n, $g(n+1) = g(n) + 1$, $P(g(n+1), h(n+1))$. It follows that for every n, $g(n) = n$ and $P(n, h(n))$, as required by the conclusion of the Countable Principle of Choice. $\dashv$

We need a definition to formulate the most useful version of the Axiom of Dependent Choices.

8.17. Definition. A graph $(G, \to_G)$ is **grounded** or **well founded** if every non-empty subset of G has a *minimal member*, i.e.,

$$\text{if } \emptyset \neq X \subseteq G, \text{ then } (\exists m \in X)(\forall x \in X)[m \not\to_G x]. \tag{8-3}$$

A poset $(P, \leq)$ is **grounded** if the associated "inverse strict graph" $(P, >)$ is grounded, which means that for every X,

$$\text{if } \emptyset \neq X \subseteq P, \text{ then } (\exists m \in X)(\forall x \in X)[x \leq m \Longrightarrow x = m]. \tag{8-4}$$

8.18. Exercise. *Prove that a linear ordering* $(P, \leq)$ *is grounded if and only if it is a wellordering.*

8.19. Proposition. *The Axiom of Dependent Choices is equivalent to the following proposition: a graph* G *is grounded if and only if it has no* **infinite**, ***descending chains***, *i.e., there exists no function* $f : \mathbb{N} \to G$ *such that for all* n, $f(n) \to_G f(n+1)$,

$$f(0) \to_G f(1) \to_G f(2) \to_G \ldots .$$

Proof. First assume **DC**. If $f : \mathbb{N} \to G$ is an infinite, descending chain, then the set $\{f(n) \mid n \in \mathbb{N}\}$ has no minimal element, so G is not grounded. Conversely, if G has a non-empty subset X with no minimal element, then $(\forall x \in X)(\exists y \in X)[x \to_G y]$, and then **DC** gives us an infinite descending chain, starting with some arbitrary $a \in X$.

Assume now that every graph which has no infinite descending chains is grounded and the hypotheses $a \in A$ and

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

of **DC** holds. Consider the graph $(A, \to_A)$ where

$$x \to_A y \iff_{\text{df}} P(x, y) \quad (x, y \in A).$$

The conclusion of **DC** is exactly the statement that $(A, \to_A)$ has an infinite descending chain which starts with a, so if it fails, there must exist some minimal $m \in A$; this means precisely that $(\forall y \in A)\neg P(m, y)$, which contradicts the hypothesis of **DC**.   ⊣

Grounded graphs have many of the properties of well ordered sets: in particular, we can prove propositions by induction and define functions by recursion over them, cf. Problems **x8.10**, **x8.11** and Theorem **11.5**. The easy direction of this result makes **DC** particularly useful in studying them, as it is often simpler to verify that a given graph G has no infinite descending chains than to prove directly that it is grounded.

8.20. But is it true? (3) We have remarked that before it was formulated precisely by Zermelo, the Axiom of Choice had been used many times "silently" in classical mathematics, and in particular in analysis. *These classical applications, however, can all be justified on the basis of the Axiom of Dependent Choices*—in fact most of them need only the weaker Countable Principle of Choice. This will become clear in Chapter **10** and Appendix **A**. Zermelo assumed the full Axiom of Choice because it is a natural hypothesis in the context of Cantor's set theory; because it is needed in the proofs of the Wellordering Theorem and the Cardinal Comparability Hypothesis; and because it is indispensable for the development of cardinal arithmetic. This difference between the choice principles needed for classical mathematics and those required by Cantor's new theory of sets explains in part the strident reaction to the axioms of Zermelo by the distinguished analysts of his time (including the great Borel), who had used choice principles routinely in their work—and continued using them, as they denounced general set theory and

called it an illusion: in the context of 19th century classical analysis, the Axiom of Dependent Choices is natural and necessary, while the full Axiom of Choice is unnecessary and even has some counterintuitive consequences, including certainly the Wellordering Theorem.

We should also mention here that even in general set theory where the full Axiom of Choice is routinely accepted as obvious, many of the basic theorems do not need it, and in particular *all the results of Chapter **2** can be based axiomatically on the Axiom of Dependent Choices*. Notice also that we proved all the basic facts about well ordered sets in the preceding chapter with no appeal to choice principles whatsoever. For this reason, we will deviate technically from Zermelo and we will put in our basic system the Axiom of Dependent Choices instead of the full Axiom of Choice. "Technically", because we adopt the view that there is no doubt about the truth of the Axiom of Choice and we will never hesitate to appeal to it when it is needed: we will simply include it (discreetly) among the hypotheses.

8.21. The axiomatic theory ZDC. The axiomatic system **ZDC** comprises the constructive axioms $(\textbf{I}) - (\textbf{VI})$ of Chapter **3** and the Axiom $(\textbf{VII})$ of Dependent Choices **8.13**, symbolically

$$\textbf{ZDC} = (\textbf{I}) - (\textbf{VI}) + \textbf{DC} = (\textbf{I}) - (\textbf{VII}).$$

From now on and until Chapter **11**, we will use in proofs the axioms of **ZDC** without explicit mention. When the Axiom of Choice is required, we will make a note of the fact by annotating the relevant proposition with the mark $(\textbf{AC})$. In Chapter **11** we will complete our axiomatization by adding to **ZDC** the *Axiom of Replacement*.

8.22. Consistency and independence results. Could we settle the controversy about the Axiom of Choice by simply *proving* or *refuting* **AC** from the constructive axioms $(\textbf{I}) - (\textbf{VI})$? Neither possibility seems likely. On the one hand, **AC** is probably true, as are axioms $(\textbf{I}) - (\textbf{VI})$, and we cannot refute a true statement on the basis of true assumptions. On the other hand, **AC** appears to be a genuinely new set theoretic principle, and we cannot expect to prove it from the other ones, by logic alone. As a matter of fact, it can be shown rigorously that the Axiom of Choice can neither be proved nor refuted from axioms $(\textbf{I}) - (\textbf{VI})$.

The most direct way to show that a certain proposition ϕ cannot be proved in a certain axiomatic system **T** is to produce a **model** of **T** in which ϕ is false. Consider the classical problem about plane Euclidean geometry, whether the *Parallel Axiom*[20] can be deduced or not from the others. To show that it cannot, we declare that by "plane" we will mean the two-dimensional sphere, the surface of the unit ball, and by "line" we will mean any great circle on the

---

[20]The Parallel (fifth) Axiom of Euclid is equivalent to the assertion that given a line L and a point P not on it, there exists exactly one line L' through P and having no points in common with L.

sphere. The remaining primitive notions of plane Euclidean geometry can be defined naturally in this interpretation, and it is not hard to verify that the basic, simple axioms of Euclid are true with these definitions; thus, we have a model of plane geometry in which the Parallel Axiom fails, simply because any two great circles intersect. We conclude that the Parallel Axiom cannot be proved from the others "by logic alone", because then it would be true in every interpretation which makes the other axioms true, and we have found one where it is false.

To define a model for an axiomatic theory, in general, one needs to specify a domain of objects and interpret on it the primitives of the theory, so that the axioms are true. For a theory about sets, this means we must define *sethood* and *membership* on some domain, and we must also identify which conditions and operations on the domain will be considered *definite*. Now, **ZDC** is a very strong theory, and models for it do not come cheap; we will study some very special ones ("set universes") in Appendix **B**, but the most interesting constructions require delicate methods from *mathematical logic* which are outside the scope of these Notes. Here we will just state and discuss some of the many famous consistency and independence results of the subject as they become relevant in what follows.

We have assumed at the outset, in **3.6**, that our theory has a model, the standard universe of objects $\mathcal{W}$, in which axioms (**I**) – (**VI**) (at least) are true. This assumption is natural and even necessary if our lives as set theorists are to have any meaning, but it is not included among the axioms of **ZDC** or any of the stronger theories we will consider.[21] It is almost never needed, except when we assert the existence of models of extensions of **ZDC**: to construct those, we have to start with something, and that is always the assumed, standard model of our theory.

8.23. Proviso for model existence assertions. Without further mention, all claims in these Notes of existence of models, consistency of theories and independence of propositions are based on the existence of a model which satisfies axioms (**I**) – (**VI**) and (**VIII**), the *Axiom of Replacement*, which we will introduce in Chapter **11**.

8.24. The consistency of the Axiom of Choice (Gödel, 1939). *Zermelo's theory* **ZDC** + **AC** *with the full Axiom of Choice has a model*, and hence (**I**) – (**VI**) do not refute **AC**, or **AC** is *consistent with* (**I**) – (**VI**). Gödel's famous model L of *constructible sets* has many more canonical properties and it witnesses the consistency of **AC** with theories much stronger than (**I**) – (**VI**). We will come back to it on several occasions.

---

[21] In fact it is not possible to assume such an axiom: adding the existence of a model of **ZDC** to the axioms of **ZDC** creates a new and stronger theory **ZDC′** and the further problem whether **ZDC′** has a model. In the most famous result of Mathematical Logic, Gödel proved (rigorously) that this conundrum cannot be avoided: *there exists no axiomatic theory* (consistent and worth studying) *which includes among its axioms or theorems the assertion that it possesses a model.*

8.25. The independence of the Axiom of Choice (Fraenkel-Mostowski, 1939, Cohen, 1963). *Each of the theories*

$$(\mathbf{I}) - (\mathbf{VI}) + \neg\mathbf{AC}_{\mathbb{N}}, \ (\mathbf{I}) - (\mathbf{VI}) + \mathbf{AC}_{\mathbb{N}} + \neg\mathbf{DC}, \ \mathbf{ZDC} + \neg\mathbf{AC}$$

has a model. This means that we cannot prove $\mathbf{AC}_{\mathbb{N}}$ from the constructive axioms $(\mathbf{I}) - (\mathbf{VI})$, we cannot prove $\mathbf{DC}$ from the constructive axioms and $\mathbf{AC}_{\mathbb{N}}$, and we cannot prove $\mathbf{AC}$ in $\mathbf{ZDC}$: each of these three choice principles is stronger than the preceding ones. The early model constructions of Fraenkel and Mostowski either contained atoms or had some other, technical defects which limited the possibility of generalizing them. Cohen constructed his models by his famous *forcing method*, which he (and others) also used to establish many more unprovability results. We will refer to it several times in the remainder of these Notes.

Problems for Chapter 8

Let us call two propositions ϕ and ψ **constructively equivalent** if their equivalence $\phi \iff \psi$ can be established on the basis of the constructive axioms $(\mathbf{I}) - (\mathbf{VI})$, i.e., without appealing to any choice principle whatsoever.

x8.1. Prove the Axiom of Choice (8-1) for finite A.

x8.2. The Axiom of Choice is constructively equivalent with the following proposition: for every $A \neq \emptyset$ and every $f : A \to B$, there exists some $g : B \to A$ such that for all $x \in A$, $f(g(f(x))) = f(x)$.

x8.3. The Axiom of Choice is constructively equivalent with the following proposition: for each I and each indexed family of sets $(i \mapsto A_i)$ on I,

$$(\forall i \in I)[A_i \neq \emptyset] \implies \prod_{i \in I} A_i \neq \emptyset.$$

The Countable Principle of Choice is constructively equivalent with the proposition: for every sequence of sets $(n \mapsto A_n)$, $n \in \mathbb{N}$,

$$(\forall n \in \mathbb{N})[A_n \neq \emptyset] \implies \prod_{n \in \mathbb{N}} A_n \neq \emptyset.$$

Combined with Problem **x5.28**, the next problem gives a formulation of the Countable Principle of Choice $\mathbf{AC}_{\mathbb{N}}$ directly in terms of the membership relation, with no reference to $\mathbb{N}$ or the concept of "function".

x8.4. The Countable Principle of Choice $\mathbf{AC}_{\mathbb{N}}$ is constructively equivalent with the following proposition: every countable, infinite family $\mathscr{E}$ of non-empty and pairwise disjoint sets admits a choice set.

In the next four problems we establish that the Axiom of Dependent Choices is equivalent with several seemingly weaker principles of choice.

*x8.5. The Axiom of Dependent Choices is constructively equivalent with the following proposition: for every non-empty A and every relation $P \subseteq A \times A$,

if $(\forall x \in A)(\exists y \in A)P(x, y)$,

$\quad$ then $(\exists B \subseteq A)[B \neq \emptyset \,\&\, (\exists f : B \to B)(\forall x \in B)P(x, f(x))]$.

*x8.6. The Axiom of Dependent Choices is constructively equivalent with the following proposition: for every relation $P \subseteq A \times A$ and $a \in A$,

if $(\forall x \in A)(\exists y \in A)P(x, y)$,

$\quad$ then $(\exists B \subseteq A)[a \in B \,\&\, (\exists f : B \to B)(\forall x \in B)P(x, f(x))]$.

*x8.7. The Axiom of Dependent Choices is constructively equivalent with the following proposition: a poset P is grounded if and only if it has no **infinite, descending chains**, i.e., if for every $f : \mathbb{N} \to P$,

$$(\forall n \in \mathbb{N})[f(n+1) \leq f(n)] \implies (\exists n)[f(n+1) = f(n)].$$

It is also possible to formulate the Axiom of Dependent Choices directly in terms of the membership relation, but not in a very pretty manner.

*x8.8. Prove that the following proposition is constructively equivalent with the Axiom of Dependent Choices: for every set A and every binary definite condition P,

if $\Big[\emptyset \in A \,\&\, (\forall u, v \in A)[P(u, v) \implies (\exists!x)[v = u \cup \{x\}]]$

$$\&\, (\forall u \in A)(\exists v \in A)P(u, v)\Big],$$

$\quad$ then $(\exists B \subseteq A)[\emptyset \in B \,\&\, (\forall u \in B)(\exists!v \in B)P(u, v)]$.

*x8.9. Prove that the following, lexicographic ordering on $(\mathbb{N} \to \mathbb{N})$ is indeed a linear ordering but not a wellordering:

$$f \leq g \iff_{\mathrm{df}} f = g \vee (\exists n \in \mathbb{N})[(\forall i < n)[f(i) = g(i)] \,\&\, f(n) < g(n)].$$

Infer that $\mathcal{P}(\mathbb{N})$ admits a linear ordering.

x8.10. **Grounded induction**. For every grounded graph G and each unary definite condition P,

$\quad$ if $(\forall x \in G)[(\forall y)(x \to_G y \implies P(y)) \implies P(x)]$, then $(\forall x \in G)P(x)$.

*x8.11. For every grounded graph G and every function

$$h : (G \to E) \times G \to E,$$

there exists a unique (total) function $f : G \to E$ which satisfies the identity

$$f(x) = h(f \restriction \{y \in G \mid x \to_G y\}, x) \quad (x \in G).$$

HINT. Rework the proof of Theorem **7.24**, using functions

$$\sigma_t : \{x \in G \mid t \Rightarrow_G x\} \to E$$

defined on "initial segments" of the transitive closure $\Rightarrow_G$ of $\to_G$.

CHAPTER 9

CHOICE'S CONSEQUENCES

We will begin this chapter with a few results about countability whose proofs illustrate the difference between $\mathbf{AC}_{\mathbb{N}}$, $\mathbf{DC}$ and $\mathbf{AC}$, but our main task is to establish some important consequences of the full Axiom of Choice, including the basic laws of *cardinal arithmetic*. The telltale mark ($\mathbf{AC}$) will grace practically all the numbered propositions.

9.1. Theorem. *Every infinite set has a countable, infinite subset, and so for every cardinal number κ, either $\kappa <_c \aleph_0$ or $\aleph_0 \leq_c \kappa$.*

PROOF. If A is infinite, obviously

$$(\forall u \in A^*)(\exists y \in A)(\forall i < \mathrm{lh}(u))[u(i) \neq y].$$

It follows from $\mathbf{DC}$ that there exists a sequence $f : \mathbb{N} \to A$ such that

$$(\forall n)(\forall i < n)[f(i) \neq f(n)],$$

and the image $f[\mathbb{N}]$ is a countable, infinite subset of A. The second assertion is trivial, taking cases on whether κ is finite or infinite. ⊣

The point of the second assertion of the theorem is that while the general property of Cardinal Comparability requires the full Axiom of Choice, the special (and significant) case of comparability with $\aleph_0$ is a theorem of $\mathbf{ZDC}$. In fact, it is possible to prove Theorem **9.1** using the Countable Principle of Choice $\mathbf{AC}_{\mathbb{N}}$ instead of $\mathbf{DC}$, but the proof is somewhat more technical. cf. Problem **x9.1**. This is a general fact about the relation between $\mathbf{DC}$ and $\mathbf{AC}_{\mathbb{N}}$: many results whose natural proofs call for $\mathbf{DC}$ follow from the weaker principle, with some additional effort.

Theorem **9.1** also settles the relation between infinite and Dedekind-infinite sets.

9.2. Corollary. *A set A is finite if and only if it is Dedekind-finite by* **4.31**, *i.e., if there exists no injection $\pi : A \rightarrowtail B \subsetneq A$ from A into one of its proper subsets.*

PROOF. Finite sets are Dedekind-finite by the Pigeonhole Principle, **5.27**. If A is infinite, let $f : \mathbb{N} \rightarrowtail A$ enumerate without repetitions some infinite, countable subset of it. The injection

$$\pi(x) = \begin{cases} f(n+1) & \text{if for some } n, \ x = f(n), \\ x & \text{if } x \notin f[\mathbb{N}] \end{cases}$$
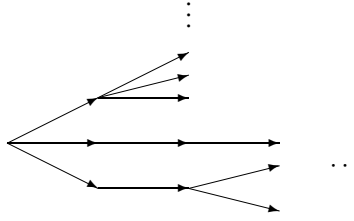
FIGURE **9.1**. Non-empty tree.

witnesses that A is Dedekind-infinite, since $\pi[A] = A \setminus \{f(0)\}$.   ⊣

Next we consider an elementary but very useful result about trees, whose proof offers an additional illustration of the use of **DC** and its relation to **AC**$_{\mathbb{N}}$.

9.3. Definition. A **tree**[22] on a set E is any set $T \subseteq E^*$ of strings from E which is closed under the relation of initial segment,

$$u \sqsubseteq v \in T \Longrightarrow u \in T.$$

By (5-16), for strings, $u \sqsubseteq v \iff u \subseteq v$.

A lot of terms are used in the study of trees, most of them deriving from our picturing trees as, well, trees. The members of T are its **nodes** or **finite branches**, and every non-empty tree has $\emptyset$ as its least node, the **root**. If $u \star \langle x \rangle \in T$, then u is a **parent** of $u \star \langle x \rangle$ and $u \star \langle x \rangle$ is a **child** of u in T. Each node other than the root has exactly one parent, but may have many children; if it has none, it is a **terminal node** or **leaf**. With each node u we associate the **subtree**

$$T_u =_{\mathrm{df}} \{w \in T \mid w \sqsubseteq u \vee u \sqsubseteq w\} \tag{9-1}$$

of nodes comparable with u. Easily,

$$T_u = \{w \mid w \sqsubseteq u\} \cup \bigcup \{T_v \mid v \text{ is a child of } u\}. \tag{9-2}$$

9.4. Exercise. *Show* (9-2).

The **infinite branches** of a tree are its infinite sequences, and we collect them in the **body** of T,

$$[T] =_{\mathrm{df}} \{f : \mathbb{N} \to E \mid (\forall n)[\overline{f}(n) \in T]\}. \tag{9-3}$$

Every infinite branch of a tree involves an infinite number of distinct nodes, so finite trees have empty bodies. It is also easy to construct infinite trees with empty bodies:

9.5. Exercise. *Show that the tree*

$$T = \{u \in \mathbb{N}^* \mid (\forall i < \mathrm{lh}(u), \ i > 0)[u(i-1) > u(i)]\}$$

on the natural numbers is infinite but has no infinite branch.

---

[22]Trees occur in many branches of mathematics, differently defined depending on the special needs of the field. The present definition is the most general we will need in these Notes.

9.6. Definition. A tree T is **finitely branching** if every node of T has at most finitely many children.

Notice that the tree in Exercise **9.5** is not finitely branching (at the root), and it could not be, by the following, basic result.

9.7. König's Lemma. *Every infinite, finitely branching tree has at least one infinite branch.*

PROOF. Suppose $T \subseteq E^*$ is infinite, finitely branching, and let

$$S =_{\mathrm{df}} \{u \in T \mid T_u \text{ is infinite}\}.$$

This is the subtree of those nodes in T which are comparable with infinitely many nodes. Since $T_\emptyset = T$ is infinite by hypothesis, the root $\emptyset \in S$, and (9-2) implies that S has no leaves,

$$(\forall u \in S)(\exists x \in T)[u \star \{x\} \in S],$$

because each u has at most finitely many children and the infinite set S_u cannot be a finite union of finite sets. By the strong version of **DC** in Proposition **8.14**, there exists some $f : \mathbb{N} \to E$ such that for every n, $\overline{f}(n) \in S$ and $f(n+1)$ is a child of $\overline{f}(n)$, so that f is an infinite branch of S—and hence of T. ⊣

König's Lemma is very useful, especially in the following, more "constructive" version.

9.8. Definition. A set of nodes $B \subseteq T$ is a **bar** in a tree T, if every infinite branch of T passes through at least one node of B,

$$(\forall f \in [T])(\exists n)[\overline{f}(n) \in B].$$

9.9. Fan Theorem. *If T is a finitely branching tree and B is a bar for T, then there exists a finite subset*

$$B_0 = \{u_1, \dots, u_n\} \subseteq B$$

which is also a bar of T.

PROOF. Let B_0 comprise the minimal members of the bar B,

$$B_0 =_{\mathrm{df}} \{u \in B \mid (\forall v \subsetneq u)[v \notin B]\},$$

and notice that B_0 is also a bar, because if $f \in [T]$ and n is least such that $\overline{f}(n) \in B$, then $\overline{f}(n) \in B_0$. Let S be the tree of all initial segments of the nodes in B_0,

$$S =_{\mathrm{df}} \{v \in T \mid (\exists u \in B_0)[v \sqsubseteq u]\}.$$

Now S is a finitely branching tree (a subtree of T), and its leaves are precisely the nodes in B_0, because no member of B_0 is a proper initial segment of another. Thus, S cannot have an infinite branch, since B_0 is a bar for T. By König's lemma then, S is finite, and so its subset B_0 is also finite. ⊣

The surprisingly simple proof of König's Lemma is typical of arguments from **DC**, partly because its basic structure calls for **DC**, but also because of the following two reasons:

(1) König's Lemma can be proved for every tree T on a well orderable set E with no use of choice principles, Problem **x9.3**. In many applications, $E = \mathbb{N}$ or E is finite, and then we need no choice hypotheses whatsoever.

(2) Like **9.1**, König's Lemma can be proved by appealing to $\mathbf{AC_\mathbb{N}}$ rather than **DC**, Problem **x9.4**.

Many of the applications of the full Axiom of Choice have the following form: first we state and prove in **ZDC** (or even with no choice at all) some interesting proposition about well orderable sets, and then we infer the result we want for all sets by appealing to the Wellordering Theorem. Typical is the following generalization of the Hypothesis of Cardinal Comparability where (for the first and last time) we will state separately the corollary about all sets.

9.10. Theorem. Well foundedness of $\leq_c$. (1) *For every non-empty class $\mathscr{E}$ of well orderable sets, there exists some $A_0 \in \mathscr{E}$ such that for every $A \in \mathscr{E}$, $A_0 \leq_c A$.*

(2) (**AC**) *Every non-empty class $\mathscr{E}$ of sets has a $\leq_c$-least member.*

PROOF. By **7.33**, let $U_0 = (A_0, \leq_0)$ be a $\leq_o$-least well ordered set with field in $\mathscr{E}$. If $A \in \mathscr{E}$, then there exists some wellordering $\leq$ of A and by the choice of U_0, $(A_0, \leq_0) \leq_o (A, \leq)$, so that, in particular, $A_0 \leq_c A$ since every initial similarity is an injection.   ⊣

9.11. Lemma. The next cardinal. *For every well orderable cardinal number κ, the cardinal*

$$\kappa^+ =_{\mathrm{df}} |\chi(\kappa)| \tag{9-4}$$

is also well orderable and it is least among the well orderable cardinals bigger than κ, i.e.,

$$\kappa <_c \kappa^+, \quad \kappa <_c \lambda \Longrightarrow \kappa^+ \leq_c \lambda, \tag{9-5}$$

for every well orderable cardinal λ. Here $\chi(\kappa)$ is the Hartogs well ordered set of κ, defined in **7.34**.

PROOF. Since κ^+ is well orderable, it is comparable with κ, and it cannot be $\leq_c \kappa$ by Hartogs' Theorem **7.34**, so $\kappa <_c \kappa^+$. The minimality of $\chi(\kappa)$ implies the rest.   ⊣

We set

$$\aleph_1 =_{\mathrm{df}} \aleph_0{}^+, \ \aleph_2 =_{\mathrm{df}} \aleph_1^+, \ldots . \tag{9-6}$$

9.12. Exercise. (**AC**) *Since* (*with* **AC**) *every two cardinal numbers are comparable, the Continuum Hypothesis* **CH** *and the Generalized Continuum Hypothesis* **GCH** *can also be expressed by simple equations of cardinal arithmetic:*

$$\mathbf{CH} \iff 2^{\aleph_0} =_c \aleph_1, \quad \mathbf{GCH} \iff (\forall \kappa \geq_c \aleph_0)[2^\kappa =_c \kappa^+]. \tag{9-7}$$

This, unfortunately, does not help their resolution.

The next Lemma is often useful in arguments about well orderable sets.

9.13. Definition. A **best wellordering** of a set A is any wellordering $\leq$ of A in which every initial segment is smaller in cardinality than A,

$$(\forall x \in A)[\mathbf{seg}(x) <_c A].$$

9.14. Lemma. (1) *Every well orderable set admits a best wellordering.*

(2) *If $\leq_A$, $\leq_B$ are best wellorderings of A and B, then*

$$\text{if } A =_c B, \text{ then } (A, \leq_A) =_o (B, \leq_B).$$

In particular, any two best wellorderings of the same set are similar.

PROOF. (1) Let $U = (A, \leq)$ be $\leq_o$-least in the class of all well ordered sets with field A, and suppose (towards a contradiction) that there exists some $x \in A$, $A \leq_c \mathbf{seg}_U(x)$. This yields an injection $\pi : A \rightarrowtail \mathbf{seg}_U(x)$ and the relation

$$u \leq' v \iff_{\text{df}} \pi(u) \leq \pi(v) \quad (u, v \in A)$$

is evidently a wellordering of A which is $\leq_o \mathbf{seg}_U(x)$ by **7.32**, hence $<_o U$, contrary to the choice of U. (2) Suppose $U = (A, \leq_A)$, $V = (B, \leq_B)$ and (towards a contradiction) $U =_o \mathbf{seg}_V(x)$ for some $x \in B$. The similarity $\pi : A \rightarrowtail \mathbf{seg}_V(x)$ witnesses that $A =_c \mathbf{seg}_V(x) <_c B$, which is contrary to the hypothesis $A =_c B$.   ⊣

Every best wellordering of a countable, infinite set is similar with the natural wellordering of $\mathbb{N}$, and we can use best wellorderings to show that many properties of countable sets hold for all well orderable sets. Typical is the next result, which generalizes the identity $\aleph_0{}^2 =_c \aleph_0$ and shows that the transfinite arithmetic of binary addition and multiplication is trivial.

9.15. Lemma. *For every infinite, well orderable set C, $C \times C =_c C$.*

PROOF. Assume the contrary towards a contradiction, let C be a $\leq_c$-least counterexample by **9.10**, and let $\leq$ be a best wellordering of C. By the choice of C, for every infinite point $x \in C$,

$$|\mathbf{seg}(x)| + |\mathbf{seg}(x)| =_c 2 \cdot |\mathbf{seg}(x)|$$

$$\leq_c |\mathbf{seg}(x)| \cdot |\mathbf{seg}(x)| =_c \mathbf{seg}(x) <_c C. \quad (9\text{-}8)$$

The key step in the proof is the following definition of a new wellordering of the product $C \times C$, due to Gödel, which we have already met (somewhat concealed) in the proof of **5.32**. We set

$$(x_1, y_1) \leq_g (x_2, y_2) \iff_{\text{df}} [\max(x_1, y_1) < \max(x_2, y_2)] \quad (9\text{-}9)$$

$$\vee [\max(x_1, y_1) = \max(x_2, y_2) \, \& \, x_1 < x_2]$$

$$\vee [\max(x_1, y_1) = \max(x_2, y_2) \, \& \, x_1 = x_2$$

$$\& \, y_1 \leq y_2].$$

The maxima here are obviously computed relative to the ordering $\leq$.

Sublemma. *The relation $\leq_g$ is a wellordering of $C \times C$.*

Proof. It is easy to verify that $\leq_g$ is a linear ordering. To show that it is a wellordering, suppose $X \subseteq C \times C$ is not empty: let w^* be $\leq$-least such that for some $(x, y) \in X$, $\max(x, y) = w^*$; next let x^* be $\leq$-least such that for some y, $(x^*, y) \in X$ and $\max(x^*, y) = w^*$; and finally let y^* be $\leq$-least such that $(x^*, y^*) \in X$, $\max(x^*, y^*) = w^*$. It follows that (x^*, y^*) is the least member of X.   ⊣ (Sublemma)
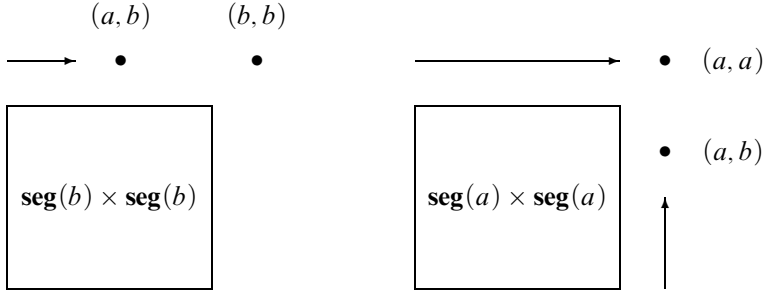
FIGURE **9.2**. Initial segments of the Gödel wellordering.

The well ordered sets $(C, \leq)$ and $(C \times C, \leq_g)$ are $\leq_o$-comparable and by the choice of $(C, \leq)$, $(C \times C, \leq_g) \leq_o (C, \leq)$ is not possible, so we must have $C <_o C \times C$; thus there exists some pair (a, b) of members of C such that

$$(C, \leq) =_o \mathbf{seg}_{C \times C}((a, b)) = \mathbf{seg}_g((a, b)),$$

and we will reach the desired contradiction if we can show that the initial segment $\mathbf{seg}_g((a, b)) <_c C$. We consider the possibilities arising from the relative positions of a and b in $\leq$, and we use the fact that the point $\max(a, b)$ must be infinite.

CASE 1, $a = b$. From the definition of the Gödel wellordering,

$$(u, v) <_g (a, a) \iff [u < a \,\&\, v < a] \vee [u < a \,\&\, v = a] \vee [u = a \,\&\, v < a],$$

so that

$$\mathbf{seg}_g((a, a)) = (\mathbf{seg}(a) \times \mathbf{seg}(a)) \cup (\mathbf{seg}(a) \times \{a\}) \cup (\{a\} \times \mathbf{seg}(a)),$$

and by repeated applications of (9-8),

$$|\mathbf{seg}_g((a, a))| \leq_c |\mathbf{seg}(a)|^2 + |\mathbf{seg}(a)| \cdot 2 \leq_c |\mathbf{seg}(a)| \cdot 3 <_c C.$$

CASE 2, $a < b$. Now $\max(a, b) = b$,

$$(u, v) <_g (a, b) \iff [u < b \,\&\, v < b] \vee [u < a \,\&\, v = b],$$

so $\mathbf{seg}((a, b)) = (\mathbf{seg}(b) \times \mathbf{seg}(b)) \cup (\mathbf{seg}(a) \times \{b\})$ and a similar computation shows again that $|\mathbf{seg}_g((a, b))| <_c C$.

CASE 3, $a > b$. This time

$$(u, v) <_g (a, b) \iff [u < a \,\&\, v < a] \vee [u < a \,\&\, v = a] \vee [u = a \,\&\, v < b],$$

from which we reach a contradiction as in the preceding cases.   ⊣

9.16. Theorem (The absorption laws). *If κ and λ are non-zero, well orderable cardinals and at least one of them is infinite, then*

$$\kappa + \lambda =_c \kappa \cdot \lambda =_c \max(\kappa, \lambda).$$

PROOF. Assuming that $0 <_c \kappa \leq_c \lambda$ and using the result $\lambda \cdot \lambda =_c \lambda$ from the Lemma, we compute

$$\lambda \leq_c \kappa + \lambda \leq_c \kappa \cdot \lambda \leq_c \lambda \cdot \lambda =_c \lambda. \qquad \dashv$$

9.17. Corollary. (**AC**) *For every indexed family of sets* $(i \mapsto \kappa_i)_{i \in I}$ *and every infinite* κ, *if* $|I| \leq_c \kappa$ *and for each* $i \in I$, $\kappa_i \leq_c \kappa$, *then* $\sum_{i \in I} \kappa_i \leq_c \kappa$.

PROOF. Using **AC** and the hypothesis, choose for each $i \in I$ some injection $\pi_i : \kappa_i \rightarrowtail \kappa$, so that the mapping $((i, x) \mapsto (i, \pi_i(x)))$ is an injection of $\{(i, x) \mid i \in I \,\&\, x \in \kappa_i\}$ into $I \times \kappa$. Thus

$$\sum_{i \in I} \kappa_i =_c \{(i, x) \mid i \in I \,\&\, x \in \kappa_i\} \leq_c |I \times \kappa| =_c |I| \cdot |\kappa| =_c \kappa. \qquad \dashv$$

To find interesting problems and results in cardinal arithmetic we must consider operations with infinitely many arguments, of which the simplest are the following.

9.18. Cardinal Minimum Lemma. *There is a definite operation* $\inf_c(\mathscr{E})$, *such that for each non-empty family* $\mathscr{E}$ *of well orderable sets, the value* $\kappa = \inf_c(\mathscr{E})$ *has the following properties*:

(1) κ *is a well orderable cardinal number.*
(2) *For some* $A \in \mathscr{E}$, $\kappa =_c A$.
(3) *For every* $B \in \mathscr{E}$, $\kappa \leq_c B$.

In addition, these conditions determine the value $\inf_c(\mathscr{E})$ *up to* $=_c$, *i.e., if* κ *is any set which satisfies* (1) – (3), *then* $\kappa =_c \inf_c(\mathscr{E})$.

PROOF. If the cardinal assignment $|X|$ is strong by **4.21**, then **9.10** implies that there exists exactly one cardinal number which satisfies the condition

$$\text{Least}(\mathscr{E}, \kappa) \iff (\exists A \in \mathscr{E})[(\forall B \in \mathscr{E})[A \leq_c B] \,\&\, \kappa = |A|],$$

and we can set

$$\inf_c(\mathscr{E}) = \min(\mathscr{E}) = \text{the unique } \kappa \text{ such that } \text{Least}(\mathscr{E}, \kappa).$$

We need to do more, since we have only assumed that $|X|$ is a weak cardinal assignment and there may well exist many values of κ which satisfy $\text{Least}(\mathscr{E}, \kappa)$.

By the Lemma in the proof of Hartogs' Theorem **7.34**, if $A \subseteq \bigcup \mathscr{E}$ and $\leq$ is a wellordering of A, then the well ordered set $U = (A, \leq)$ is similar with some proper initial segment of $W = \chi(\bigcup \mathscr{E})$, and hence every $A \in \mathscr{E}$ is equinumerous with some proper initial segment of W. Thus we can set

$$w =_{\text{df}} \text{the least } x \in W \text{ such that } (\exists A \in \mathscr{E})[A =_c \mathbf{seg}_W(x)],$$
$$\inf_c(\mathscr{E}) =_{\text{df}} |\mathbf{seg}_W(w)|.$$

Verification of the required properties of $\inf_c(\mathscr{E})$ is quite easy. $\dashv$

9.19. Exercise. *Show the part of the theorem which follows the "in addition".*

9.20. Cardinal Supremum Lemma. *There is a definite operation* $\sup_c(\mathscr{E})$, *such that for every non-empty family* $\mathscr{E}$ *of well orderable sets, the value* $\kappa = \sup_c(\mathscr{E})$ *has the following properties*:

(1) κ *is a well orderable cardinal.*

(2) *For every $A \in \mathscr{E}$, $A \leq_c \kappa$.*

(3) *If B is well orderable and for all $A \in \mathscr{E}$, $A \leq_c B$, then $\kappa \leq_c B$.*

In addition, these conditions determine the value $\sup_c(\mathscr{E})$ up to $=_c$, i.,e. if κ is any set which satisfies $(1) - (3)$, then $\kappa =_c \sup_c(\mathscr{E})$.

PROOF. Let $C = h(\bigcup \mathscr{E})$ be the Hartogs set for the union of $\mathscr{E}$, which by Hartogs' Theorem **7.34** is well orderable and greater in cardinality than every well orderable subset of $\bigcup \mathscr{E}$, including every $A \in \mathscr{E}$. We set

$$\sup_c(\mathscr{E}) =_{\mathrm{df}} \inf_c(\{B \subseteq C \mid (\forall A \in \mathscr{E})[A \leq_c B]\})$$

and verify easily the conclusion of the Lemma.   ⊣

Infinite sums and products were defined in **4.21**. We cannot say much about these, because infinite sums are as trivial as the finite ones (Problem **x9.15**), and infinite products are as complex as the Generalized Continuum Hypothesis, because

$$2^\kappa =_c \prod_{i \in \kappa} 2.$$

There is, however, a very interesting inequality relating the two.

9.21. König's Theorem. (**AC**) *For any two families of sets $(i \mapsto A_i)$ and $(i \mapsto B_i)$ on the same index set $I \neq \emptyset$,*

$$\textit{if } (\forall i \in I)[A_i <_c B_i], \textit{ then } \bigcup_{i \in I} A_i <_c \prod_{i \in I} B_i. \qquad (9\text{-}10)$$

In particular, for families of cardinals, $(i \mapsto \kappa_i)$ and $(i \mapsto \lambda_i)$,

$$\textit{if } (\forall i \in I)[\kappa_i <_c \lambda_i], \textit{ then } \sum_{i \in I} \kappa_i <_c \prod_{i \in I} \lambda_i. \qquad (9\text{-}11)$$

PROOF. The hypothesis and **AC** yield for each i an injection $\pi_i : A_i \rightarrowtail B_i$; and since π_i cannot be a bijection, there also exists a function $c : I \to \bigcup_{i \in I} B_i$ such that for each i, $c(i) \in B_i \setminus \pi_i[A_i]$. We set

$$f(x, i) = \begin{cases} \pi_i(x), & \text{if } x \in A_i, \\ c(i), & \text{if } x \notin A_i, \end{cases}$$
$$g(x) = (i \mapsto f(x, i)).$$

If $x \neq y$ and x, y belong to the same A_i for some i, then

$$g(x)(i) = \pi_i(x) \neq \pi_i(y) = g(y)(i),$$

because π_i is an injection, and hence $g(x) \neq g(y)$. If no A_i contains both x and y, suppose $x \in A_i$, $y \notin A_i$; it follows that $g(x)(i) = \pi_i(x) \in \pi_i[A_i]$ and $g(y)(i) = c(i) \in B_i \setminus \pi_i[A_i]$ so that again $g(x) \neq g(y)$. We conclude that the mapping $g : \bigcup_{i \in I} A_i \rightarrowtail \prod_{i \in I} B_i$ is an injection, and hence

$$\bigcup_{i \in I} A_i \leq_c \prod_{i \in I} B_i.$$

Suppose, towards a contradiction that there existed a correspondence

$$h : \bigcup_{i \in I} A_i \longrightarrow\!\!\!\!\rightarrow \prod_{i \in I} B_i,$$

so that these two sets are equinumerous. For every i, the function

$$h_i(x) =_{\mathrm{df}} h(x)(i) \quad (x \in A_i)$$

is (easily) a function of A_i into B_i and by the hypothesis it cannot be a surjection; hence by **AC** there exists a function ε which selects in each B_i some element not in the image, i.e.,

$$\varepsilon(i) \in B_i \setminus h_i[A_i], \quad (i \in I).$$

By its definition, $\varepsilon \in \prod_{i \in I} B_i$, so there must exist some $x \in A_j$, for some j, such that $h(x) = \varepsilon$; this yields

$$\varepsilon(j) = h(x)(j) = h_j(x) \in h_j[A_j],$$

contrary to the characteristic property of ε.

The cardinal version (9-11) follows by applying (9-10) to $A_i = \{i\} \times \kappa_i$ and $B_i = \lambda_i$. $\dashv$

9.22. Exercise. (**AC**) *König's Theorem applies to the case $I = \kappa$, $A_i = \{i\}$ and $B_i = 2$ and yields*

$$\kappa = \bigcup_{i \in \kappa} \{i\} <_c \prod_{i \in \kappa} 2 =_c 2^\kappa,$$

i.e., the theorem of Cantor.

Despite its simplicity, König's Theorem implies immediately a non-trivial inequality about the cardinal number $\mathfrak{c}$ of the continuum which goes beyond Cantor's $\aleph_0 <_c \mathfrak{c}$. It is most naturally expressed using *cofinalities*.

9.23. Definition. The **cofinality** of a well orderable, infinite cardinal number κ is the least well orderable cardinal λ such that κ is the union of λ-many sets with cardinality smaller than κ. Precisely:

$$\mathrm{cf}(\kappa) =_{\mathrm{df}} \inf_c(\{I \subseteq \kappa \mid \text{for some indexed family } (i \mapsto K_i)_{i \in I},$$
$$(\forall i \in I)[K_i <_c \kappa] \,\&\, \kappa = \bigcup_{i \in I} K_i\}).$$

Notice that the family of well orderable index sets whose $\inf_c$ we take is not empty; it contains κ, since

$$\kappa = \bigcup_{i \in \kappa} \{i\}. \tag{9-12}$$

The general properties of $\inf_c$ imply the following basic properties of the cofinality operation:

(1) $\mathrm{cf}(\kappa) \leq_c \kappa$.
(2) $\kappa = \bigcup_{i \in \mathrm{cf}(\kappa)} K_i$ for some indexed family $(i \mapsto K_i)$ such that $(\forall i \in \mathrm{cf}(\kappa))[K_i <_c \kappa]$.
(3) If λ is well orderable, $(\forall i \in \lambda)[L_i <_c \kappa]$ and $\kappa = \bigcup_{i \in \lambda} L_i$, then $\mathrm{cf}(\kappa) \leq_c \lambda$.

Moreover, these conditions characterize $\mathrm{cf}(\kappa)$ up to $=_c$.

A well orderable, infinite cardinal κ is **regular** if $\mathrm{cf}(\kappa) =_c \kappa$, otherwise it is **singular**. It is convenient to define the operation $\mathrm{cf}(\kappa)$ and the regularity condition for infinite, well orderable κ without assuming the full Axiom of Choice, but most results about these notions require **AC**.

9.24. Exercise. $\aleph_0$ *is regular, because every finite sum of finite cardinals is finite.*

9.25. Corollary. (**AC**) *For each infinite cardinal number* κ,

$$\mathrm{cf}(2^\kappa) >_c \kappa,$$

and in particular, $\mathrm{cf}(\mathfrak{c}) >_c \aleph_0$, *i.e., the continuum* $\mathfrak{c}$ *is not the union of countably many sets of cardinality* $<_c \mathfrak{c}$.

PROOF. By König's Theorem, if $K_i <_c 2^\kappa$ for every $i \in \lambda$ with $\lambda \leq_c \kappa$, then

$$\bigcup_{i\in\lambda} K_i <_c \prod_{i\in\lambda} 2^\kappa =_c (2^\kappa)^\lambda =_c 2^{\kappa\cdot\lambda} =_c 2^\kappa,$$

which contradicts $\mathrm{cf}(2^\kappa) \leq_c \kappa$.   ⊣

9.26. Gödel's model L of the constructible sets satisfies the Generalized Continuum Hypothesis, so for each κ, $2^\kappa =_c \kappa^+$ is regular by Problem **x9.19**. Using Cohen's forcing method, it is possible to construct models of Zermelo's theory **ZDC** + **AC** in which $\mathfrak{c}$ is singular, with cofinality $\mathrm{cf}(\mathfrak{c})$ some regular cardinal between $\aleph_0$ and $\mathfrak{c}$, for example $\aleph_1$.

We have left the basic properties of cofinalities for the problems, as they are very simple. We should remark, however, that it is not possible to study the topic seriously now, because without the Axiom of Replacement, it is not even possible to prove that singular cardinals *exist*!

Problems for Chapter 9

****x9.1.** Show Theorem **9.1** using only the constructive axioms (**I**) – (**VI**) and the Countable Principle of Choice **AC**$_\mathbb{N}$.

x9.2. Consider a system of airline routes which connects the (possibly infinitely many) cities of some world and assume the following. (1) From each city, there are only finitely many cities to which one can fly non-stop. (2) It is possible to travel by air from every city to every other city. (3) It is not possible to keep flying forever without visiting the same Airport twice. Show that this world has only finitely many cities.

x9.3. Show König's Lemma **9.7** for the case where T is a tree on a well orderable set E, with no appeal to choice principles.

****x9.4.** Show König's Lemma **9.7** using only the constructive axioms (**I**) – (**VI**) and the Countable Principle of Choice **AC**$_\mathbb{N}$.

x9.5. Suppose T is a finitely splitting tree and B is a bar for T. Show that there exists some integer k, such that for all infinite branches $f \in [T]$, some $\overline{f}(i) \in B$, with $i \leq k$.

****x9.6.** Suppose C is a well orderable set, $f : C \times C \to C$, $A \subseteq C$, and let

$$A_f =_{\mathrm{df}} \bigcap \{X \subseteq C \mid A \subseteq X \,\&\, f[X \times X] \subseteq X\}$$

be the **closure** of A under f. Define the sets $\{A_n\}_{n \in \mathbb{N}}$ by the recursion

$$A_0 = A, \quad A_{n+1} = A_n \cup f[A_n \times A_n],$$

and show that

$$A_f = \bigcup_{n \in \mathbb{N}} A_n.$$

Show also that if A is infinite, then $A_f =_c A$.

x9.7. Show that if C is well orderable, then so is the set C^* of all words (finite sequences) from C.

x9.8. If you used $\mathbf{AC_\mathbb{N}}$ or $\mathbf{DC}$ in Problems **x9.6** and **x9.7**, do them again, using no choice principles at all.

x9.9. Every Hartogs set $h(A)$ is best wellordered by $\leq_{\chi(A)}$.

x9.10. If $(n \mapsto \kappa_n)_{n \in \mathbb{N}}$ and $(n \mapsto \lambda_n)_{n \in \mathbb{N}}$ are sequences of cardinal numbers, and for every n, $\kappa_n \leq_c \lambda_n$, then

$$\sum_{n \in \mathbb{N}} \kappa_n \leq_c \sum_{n \in \mathbb{N}} \lambda_n, \qquad \prod_{n \in \mathbb{N}} \kappa_n \leq_c \prod_{n \in \mathbb{N}} \lambda_n.$$

x9.11. (**AC**) If $(i \mapsto \kappa_i)_{i \in I}$ and $(i \mapsto \lambda_i)_{i \in I}$ are families of cardinal numbers on the same index set I and for all $i \in I$, $\kappa_i \leq_c \lambda_i$, then

$$\sum_{i \in I} \kappa_i \leq_c \sum_{i \in I} \lambda_i, \ \prod_{i \in I} \kappa_i \leq_c \prod_{i \in I} \lambda_i.$$

x9.12. (**AC**) For every indexed family of sets $(i \mapsto A_i)_{i \in I}$,

$$\left| \prod_{i \in I} A_i \right| =_c \prod_{i \in I} |A_i|$$

and the same for sums, with "disjoint union" on the left.

x9.13. Explain the notation and prove the identity

$$\prod_{i \in I} \prod_{j \in J(i)} \kappa_{ij} =_c \prod_{\{(i,j) \mid i \in I \ \& \ j \in J(i)\}} \kappa_{ij}.$$

x9.14. (**AC**) Prove the characterization of $\sup_c(\mathscr{E})$ claimed in **9.20**.

x9.15. (**AC**) For every family of infinite cardinal numbers $(i \mapsto \kappa_i)$ on a non-empty index set I,

$$\sum_{i \in I} \kappa_i =_c \max(|I|, \sup_c(\{\kappa_i \mid i \in I\})).$$

x9.16. Prove that for every infinite, well orderable cardinal κ,

$$\mathrm{cf}(\kappa) = \inf_c \Big(\{I \subseteq \kappa \mid \text{for some indexed family } (i \mapsto K_i')_{i \in I},$$
$$(\forall i, j \in I)[i \neq j \implies K_i' \cap K_j' = \emptyset] \ \& \ (\forall i \in I)[K_i' <_c \kappa]$$
$$\& \ \kappa = \bigcup_{i \in I} K_i'\} \Big).$$

*$\mathbf{x9.17.}$ (**AC**) Prove that for every infinite cardinal κ,

$$\mathrm{cf}(\kappa) = \inf_c \Big(\{ I \subseteq \kappa \mid \text{for some indexed family of cardinals } (i \mapsto \kappa_i)_{i \in I},$$

$$(\forall i \in I)[\kappa_i <_c \kappa] \,\&\, \kappa =_c \sum\nolimits_{i \in I} \kappa_i \} \Big).$$

Hint. It helps to use Problem $\mathbf{x9.16}$.

*$\mathbf{x9.18.}$ Show that for every infinite cardinal κ, $\mathrm{cf}(\mathrm{cf}(\kappa)) =_c \mathrm{cf}(\kappa)$, and hence $\mathrm{cf}(\kappa)$ is always a regular cardinal

$\mathbf{x9.19.}$ (**AC**) For each infinite cardinal κ, the next cardinal κ^+ is regular. Hint. Problem $\mathbf{x9.17}$ simplifies the proof.

$\mathbf{x9.20.}$ (**AC**) Show that for each infinite cardinal κ, $\kappa <_c \kappa^{\mathrm{cf}(\kappa)}$.

*$\mathbf{x9.21.}$ (**AC**) Every partial ordering $\leq$ on a set P has a **linearization**, i.e., some linear ordering $\leq'$ of P exists such that $x \leq y \Longrightarrow x \leq' y$.

The next problem gives the basic fact which relates inductive and directed-complete posets. Notice that by "chain" in a poset P we mean any subset $C \subseteq P$ which is linearly ordered by the poset partial ordering $\leq_P$; C is a *well ordered chain* if in addition, the restriction of $\leq_P$ to C is a wellordering. When we say that "S is well orderable" for some $S \subseteq P$, we mean that S *admits some wellordering* $\leq$, which may be (and typically is) totally unrelated with the given partial ordering $\leq_P$ of P.

*$\mathbf{x9.22.}$ If every well ordered chain in a poset $(P, \leq_P)$ has a least upper bound, then for every well orderable, directed subset S of P there exists a well ordered chain C with the following two properties.
   (1) S is dominated by C, i.e., for each $x \in S$ there exists some $y \in C$ such that $x \leq_P y$.
   (2) For each $y \in C$, there exists a directed subset $C_y \subseteq S$ such that $|C_y| <_c |S|$ and $y = \sup C_y$.
   Notice that C may satisfy these conditions without being a subset of S. Hint (W. Allen). Towards a contradiction, let S a be well orderable, directed and $\leq_c$-least counterexample to the conclusion, verify first that S must be uncountable, and let $\leq$ be a best wellordering of S. Define the function $f : S \times S \to S$ so that $x, y \in S \Longrightarrow x, y \leq_P f(x, y)$, and for every $x \in S$, set

$$C_x =_{\mathrm{df}} \mathbf{seg}(x)_f,$$

with the notation of Problem $\mathbf{x9.6}$. Show that this is directed, that $\sup C_x$ exists for each $x \in S$, and that

$$C =_{\mathrm{df}} \{\sup C_x \mid x \in S\}$$

is a well ordered chain in P which has properties (1) and (2) for S.

*$\mathbf{x9.23.}$ (**AC**) The following three conditions are equivalent, for every poset P:
   1. Every directed set in P has a least upper bound.
   2. Every chain in P has a least upper bound.

3. Every well ordered chain in P has a least upper bound.

In particular: (**AC**) *A poset is inductive if and only if it is directed-complete, a dcpo.*

****x9.24.** (**AC**) Show that a monotone mapping $\pi : P \to Q$ on one inductive poset to another satisfies the identity

$$\pi(\sup S) = \sup \pi[S] \tag{9-13}$$

for every non-empty chain $S \subseteq P$, if and only if it satisfies (9-13) for every non-empty directed $S \subseteq P$.

x9.25. Prove that the characterization of continuity for mappings of the form $\pi : (A \rightharpoonup E) \to (B \rightharpoonup M)$ in **x6.24** holds for all sets A, E, B, M.

x9.26. (**AC**) **Finite Basis Lemma**. Let $\mathscr{I}$ be a non-empty family of subsets of some set V, such that

$$X \in \mathscr{I} \iff (\forall Y \subseteq X)[Y \text{ finite} \Longrightarrow Y \in \mathscr{I}].$$

Show that $\mathscr{I}$ has a maximal member (under $\subseteq$).

***x9.27.** Let $\mathscr{I}$ be a family with the finite basis property as in **x9.26** and assume in addition that V is well orderable; show (without **AC**) that $\mathscr{I}$ has a maximal member.

x9.28. (**AC**) If you know what *vector spaces* are and the basic facts about *linear independence*, prove that every vector space has a basis. Prove also without **AC**, that every well orderable vector space has a basis. HINT. Apply **x9.26** or **x9.27** to the family of all linearly independent subsets of the given space.

***x9.29.** (**AC**) If you know something about fields and algebraic extensions, prove that every field has an algebraic closure. HINT. The usual argument for this runs as follows. We consider the class

$$\mathscr{A} =_{df} \{F \mid F \text{ is an algebraic extension of } K\} \tag{9-14}$$

partially ordered by

$$F_1 \sqsubseteq F_2 \iff_{df} F_1 \text{ is a subfield of } F_2,$$

we notice that it is an inductive poset, so that it has a maximal element $\overline{K}$, and we verify that this $\overline{K}$ is algebraically closed. The argument is defective, because the class $\mathscr{A}$ in (9-14) is not a set. To correct it, in the interesting case where K is infinite, we need to notice that *every algebraic extension of K is isomorphic with some field $F =_c K$*, so we can replace $\mathscr{A}$ in (9-14) by

$$\mathscr{A}' =_{df} \{F \subseteq E \mid F \text{ is an algebraic extension of } K\}, \tag{9-15}$$

where E is some superset of K with cardinality greater than K.

*$\mathbf{x9.30.}$ Prove that every well orderable field (and in particular, every countable field) has an algebraic closure. HINT. The idea is to avoid $\mathbf{AC}$ by using Transfinite Recursion to construct the closure explicitly. You still need the trick suggested in the previous problem. Do the countable case first, it clarifies which algebraic results are needed.

CHAPTER 10

BAIRE SPACE

Next to the natural numbers, perhaps the most fundamental object of study of set theory is **Baire space**,

$$\mathcal{N} =_{\mathrm{df}} (\mathbb{N} \to \mathbb{N}), \tag{10-1}$$

the set of all number theoretic sequences. If we let

$$\mathcal{C} =_{\mathrm{df}} (\mathbb{N} \to \{0, 1\}) \tag{10-2}$$

be the **Cantor set**[23] of all infinite, binary sequences, then

$$\mathcal{C} \subseteq \mathcal{N} \subseteq \mathcal{P}(\mathbb{N} \times \mathbb{N}),$$

and with now familiar computations,

$$\mathfrak{c} =_c 2^{\aleph_0} =_c |\mathcal{P}(\mathbb{N})| =_c |\mathcal{C}| \leq_c |\mathcal{N}| \leq_c |\mathcal{P}(\mathbb{N} \times \mathbb{N})| =_c |\mathcal{P}(\mathbb{N})| = \mathfrak{c}.$$

Since $\mathcal{N} =_c \mathbb{R}$ will follow as in Chapter **2** from the proper definitions in Appendix **A**, the Continuum Hypothesis **3.2** is equivalent to the proposition

(**CH**) $\qquad\qquad (\forall X \subseteq \mathcal{N})[X \leq_c \mathbb{N} \vee X =_c \mathcal{N}].$

In fact, there is such a tight connection between $\mathcal{N}$, $\mathcal{C}$ and $\mathbb{R}$ that practically every interesting property of one of these spaces translates immediately to a related, interesting property of the others. In the problems we will make this precise for $\mathcal{N}$ and $\mathcal{C}$ and in Appendix **A** for $\mathbb{R}$, where we will also draw the consequences of the results of this chapter for the real numbers.[24]

The material in this chapter is not necessary for the comprehension of the two chapters which follow, and the exposition is more condensed and requires more effort from the reader than the rest of these Notes. In a first reading, it may be best to skip it and come back to it after Chapters 11 and 12 have been mastered.

[23] It is traditional to use the same name for this subset of $\mathcal{N}$ and the set of real numbers defined in the proof of **2.14**. Figure **2.4** explains vividly the reason for this, and nobody has ever been confused by it.

[24] One may think of $\mathcal{N}$ as a "discrete", "digital", or "combinatorial" version of the "continuous" or "analog" $\mathbb{R}$. A real number x is completely determined by a decimal expansion $x(0) . x(1)x(2) \dots$, where $(n \mapsto x(n)) \in \mathcal{N}$, *but* two distinct decimal expansions may compute to the same real number. This is a big "but", it is the key fact behind the so-called *topological connectedness* of the real line which is of interest in analysis, to be sure, but of little set theoretic consequence. We may view Baire space as a "digital version" of $\mathbb{R}$ because it does not make any such identifications, each point $x \in \mathcal{N}$ determines unambiguously its "digits" $x(0), x(1), \dots$
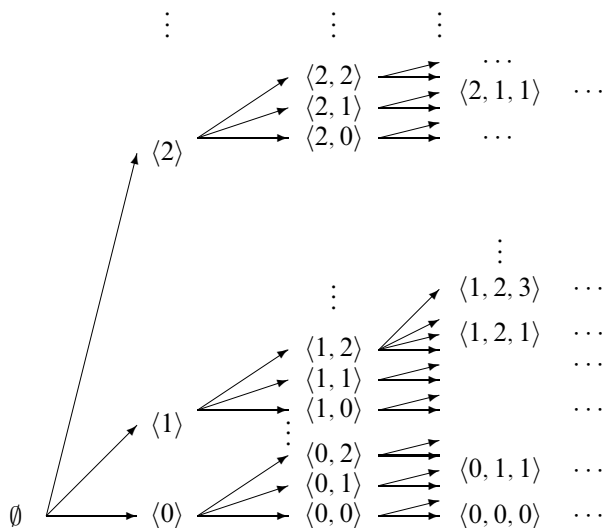
Figure **10.1**. A small part of Baire space.

Our aim here is to establish some elementary facts about $\mathcal{N}$ which bear on the Continuum Problem. We will define the family of ANALYTIC SUBSETS OF $\mathcal{N}$ and prove that every analytic set satisfies the Continuum Hypothesis, in the sense that it must be either countable or equinumerous with $\mathcal{N}$. This PERFECT SET THEOREM **10.20** is significant because essentially every set of interest in classical analysis is analytic, including all the BOREL SETS which play a fundamental role in measure theory and integration; by SUSLIN's THEOREM **10.31**, a set $A \subseteq \mathcal{N}$ is Borel exactly when both A and $\mathcal{N} \setminus A$ are analytic. On the other hand, we will show in Theorem **10.32** that the basic and natural method of proving the Continuum Hypothesis for analytic sets cannot be extended to solve the full Continuum Problem, which remains open. In addition to their applications in analysis, Theorems **10.20** and **10.32** are of substantial foundational interest, as their proofs illustrate beautifully the role of choice principles in classical mathematics.

10.1. The structure of $\mathcal{N}$. Our intuitions about $\mathcal{N}$ come from picturing it as the body of the largest tree on $\mathbb{N}$ in the terminology of **9.3** and (9-3),

$$\mathcal{N} = [\mathbb{N}^*].$$

We will refer to subsets of Baire space as **pointsets**, the term "point" temporarily reserved for members of $\mathcal{N}$, infinite branches in $\mathbb{N}^*$. By the **complement** of a pointset, we will mean its complement in $\mathcal{N}$,

$$cA =_{\mathrm{df}} \mathcal{N} \setminus A. \tag{10-3}$$

It is convenient to extend the initial segment notation on strings,

$$u \sqsubseteq x \iff_{\mathrm{df}} u \subset x \quad (u \in \mathbb{N}^*, x \in \mathcal{N}), \tag{10-4}$$

to indicate that a finite sequence u is an initial segment of the point x, an **approximation** of x which determines the first $\mathrm{lh}(u)$ values of x. For each $u \in \mathbb{N}^*$, the set

$$\mathcal{N}_u =_{\mathrm{df}} \{x \in \mathcal{N} \mid u \sqsubseteq x\} = [\mathbb{N}^*_u] \qquad (10\text{-}5)$$

of points in $\mathcal{N}$ which extend u is the **neighborhood** determined by u in $\mathcal{N}$.

10.2. Exercise. *For all $u, v \in \mathbb{N}^*$,*

$$u \sqsubseteq v \iff \mathcal{N}_v \subseteq \mathcal{N}_u.$$

10.3. Exercise. *The family of neighborhoods is countable.*

10.4. Definition. A pointset G is **open** if it is a union of neighborhoods, so that

$$x \in G \iff (\exists u)[x \in \mathcal{N}_u \ \& \ \mathcal{N}_u \subseteq G]; \qquad (10\text{-}6)$$

closed if its complement is open; and **clopen** if it is both closed and open.

Open sets are often defined by the following, easily equivalent condition:

10.5. Exercise. *A set $G \subseteq \mathcal{N}$ is open if and only if for every $x \in G$, there exists some neighborhood $\mathcal{N}_u$ such that $x \in \mathcal{N}_u \subseteq G$.*

10.6. Proposition. (1) $\emptyset$, $\mathcal{N}$ *and all neighborhoods are clopen.*

(2) *Every singleton $\{x\}$ is closed but not open.*

(3) *Every non-empty open pointset is the union of a sequence of neighborhoods.*

(4) *The union $\bigcup \mathcal{G}$ of a family $\mathcal{G}$ of open pointsets is open and, dually, the intersection $\bigcap \mathcal{F}$ of a family $\mathcal{F}$ of closed pointsets is closed.* (We assume $\bigcap \emptyset = \mathcal{N}$, so the intersection operation is defined for every family of pointsets.)

(5) *The intersection $G_1 \cap G_2$ of two open pointsets is open, and dually, the union $F_1 \cup F_2$ of two closed pointsets is closed.*

PROOF. (1) Each neighborhood is open, since $\mathcal{N}_u = \bigcup\{\mathcal{N}_u\}$, and, in particular, $\mathcal{N} = \mathcal{N}_{\langle\,\rangle}$ is open. Neighborhoods are also closed by Exercise **10.5**: if $x \notin \mathcal{N}_u$, then $u \not\sqsubseteq x$, so there exists some $i < \mathrm{lh}(u)$ such that $x(i) \neq u(i)$ and then $x \in \mathcal{N}_{\langle x(0),\dots,x(i)\rangle}$ while $\mathcal{N}_{\langle x(0),\dots,x(i)\rangle} \cap \mathcal{N}_u = \emptyset$. The empty set is the union of the empty (!) family of neighborhoods, formally

$$\emptyset = \bigcup\{\mathcal{N}_u \mid \mathcal{N}_u \subseteq \emptyset\}.$$

(2) A singleton $\{x\}$ is not open, because it does not contain any neighborhood, and so it cannot be a union of neighborhoods. Its complement is open, however, since

$$\mathcal{N} \setminus \{x\} = \bigcup\{\mathcal{N}_u \mid u \not\sqsubseteq x\}.$$

(3) If G is open and non-empty, then the set $\{u \mid \mathcal{N}_u \subseteq G\}$ is non-empty and countable, and so it can be enumerated.

(4) This is immediate from the characterization of open sets in Exercise **10.5**.

(5) If G_1, G_2 are open and $x \in G_1 \cap G_2$, then there exist $u, v \sqsubseteq x$ such that $\mathcal{N}_u \subseteq G_1$ and $\mathcal{N}_v \subseteq G_2$. The finite sequences u, v are comparable since they are both initial segments of x, so suppose $u \sqsubseteq v$, the argument being the same

in the opposite case: now $\mathcal{N}_u \supseteq \mathcal{N}_v$, so $\mathcal{N}_v \subseteq G_1 \cap G_2$, as required. The dual property for closed sets follows by taking complements.   ⊣

Baire space is a *topological space* by the classical definition recounted in **4.30**, but a very special one, because of the next, basic connection between its topology and the combinatorial structure of the tree $\mathbb{N}^*$. In proving it—and in the sequel, routinely—we will use the following trivial equivalence relating a tree T and its body:

$$x \in [T] \iff (\forall u \sqsubseteq x)[u \in T]. \tag{10-7}$$

It follows immediately from the definition of $[T]$, (9-3).

10.7. Proposition. *A pointset F is closed if and only if it is the body of a tree T on $\mathbb{N}$, $F = [T]$.*

Proof. If $x \notin [T]$, then for some $u \sqsubseteq x$, $u \notin T$, and then $\mathcal{N}_u \cap [T] = \emptyset$, so $\mathcal{N}_u \subseteq c[T]$; thus $c[T]$ is open and $[T]$ is closed. Conversely, if we associate with each pointset F the tree

$$T^F =_{df} \{u \in \mathbb{N}^* \mid (\exists x \in F)[u \sqsubseteq x]\}, \tag{10-8}$$

then obviously

$$F \subseteq [T^F].$$

If F is closed, we also have $[T^F] \subseteq F$: because if $x \notin F$, then for some $u \sqsubseteq x$, $\mathcal{N}_u \cap F = \emptyset$ by the openness of the complement cF, hence $u \notin T^F$ and $x \notin [T^F]$ by (10-7).   ⊣

This basic characterization allows us to classify closed pointsets by the combinatorial properties of the trees which define them. It is not wrong to think of the cluster of combinatorial notions to come as the **combinatorial geometry** of $\mathcal{N}$, although it is not a "geometry" by any standard, classical definition of this term.

10.8. Definition. Set

$$u \mid v \iff_{df} u, v \text{ are incompatible} \tag{10-9}$$
$$\iff (\exists i < \text{lh}(u), \text{lh}(v))[u(i) \neq v(i)],$$

and, by extension,

$$u \mid x \iff_{df} \neg[u \sqsubseteq x] \iff (\exists v \sqsubseteq x)[u \mid v].$$

A string u **splits** in a tree T if it has incompatible extensions in T and a tree T is **splitting** if every $u \in T$ splits in T,

$$u \in T \implies (\exists u_1, u_2 \in T)[u \sqsubseteq u_1 \ \& \ u \sqsubseteq u_2 \ \& \ u_1 \mid u_2].$$

Notice that a splitting tree has no terminal nodes.

*A pointset P is **perfect** if it is the body of a splitting tree.* Perfect sets are automatically closed.

10.9. Exercise. *Each neighborhood $\mathcal{N}_u$ is perfect.*

10.10. Proposition. *Every non-empty, perfect pointset P has cardinality $\mathfrak{c}$.*

PROOF. Suppose $P = [T]$ with T non-empty, splitting, and choose functions

$$l : T \to T, \quad r : T \to T$$

which witness the splitting property for T, i.e., for each $u \in T$,

$$u \sqsubseteq l(u), \quad u \sqsubseteq r(u), \quad l(u) \mid r(u).$$

By the String Recursion Theorem **5.33**, there is a function

$$\sigma : \{0, 1\}^* \to T$$

from the tree of all binary strings into T which satisfies the identities

$$\sigma(\emptyset) = \emptyset, \quad \sigma(u \star \langle 0 \rangle) = l(\sigma(u)), \quad \sigma(u \star \langle 1 \rangle) = r(\sigma(u)).$$

Thus $\sigma(u \star \langle i \rangle)$ is a proper extension of $\sigma(u)$ for $i = 0, 1$, so σ is strictly monotone,

$$u \subsetneqq v \Longrightarrow \sigma(u) \subsetneqq \sigma(v),$$

and we can define a function $\pi : \mathcal{C} \to [T]$ by

$$\pi(x) =_{\mathrm{df}} \sup \{\sigma(u) \mid u \sqsubseteq x\}. \tag{10-10}$$

The key property of σ is that it also preserves incompatibility,

$$u \mid v \Longrightarrow \sigma(u) \mid \sigma(v). \tag{10-11}$$

To see this, let i be least such that $u(i) \neq v(i)$, so for some w we have

$$w \star \langle 0 \rangle \sqsubseteq u, \quad w \star \langle 1 \rangle \sqsubseteq v$$

(or the other way around); now $\sigma(w \star \langle 0 \rangle)$ and $\sigma(w \star \langle 1 \rangle)$ are incompatible by their definition, and the monotonicity property implies that $\sigma(u)$, $\sigma(v)$ extend them, so they are incompatible too. Finally, (10-11) implies that π is an injection, and this establishes that $\mathcal{C} \leq_c [T]$, which is all we need.   ⊣

This simple abstraction of Cantor's proof of the uncountability of the reals (**2.14**) suggests an attack on the Continuum Problem: to prove that an uncountable pointset has cardinality $\mathfrak{c}$, it is enough to show that it contains a non-empty, perfect subset. This is trivially true of open sets (because each $\mathcal{N}_u$ is perfect) and it is also true of closed sets, less trivially.

10.11. Cantor-Bendixson Theorem. *Every closed subset F of $\mathcal{N}$ can be decomposed uniquely into two disjoint subsets*

$$F = P \cup S, \quad P \cap S = \emptyset, \tag{10-12}$$

*where P, the **kernel** of F, is perfect and S, the **scattered part** of F, is countable.*

It follows that every uncountable, closed pointset has a non-empty, perfect kernel and hence has cardinality $\mathfrak{c}$.

PROOF. Let $T = T^F$ as in (10-8), so that T has no terminal nodes and $F = [T]$, and set

$$S =_{\mathrm{df}} \bigcup \{[T_u] \mid u \in T \ \& \ |[T_u]| \leq_c \aleph_0\},$$
$$P =_{\mathrm{df}} F \setminus S.$$

By its definition, S is the union of countably many, countable sets, so it is countable (note the use of $\mathbf{AC}_\mathbb{N}$ here), and it remains to show that P is perfect.

The set of strings

$$kT = \{u \in T \mid |[T_u]| >_c \aleph_0\}$$

is easily a tree, and

$$x \in S \iff x \in F \;\&\; (\exists u \sqsubseteq x)[u \notin kT]$$

is another way to read the definition of S. Since $P = F \setminus S$,

$$
\begin{aligned}
x \in P &\iff x \in F \;\&\; [x \notin F \vee (\forall u \sqsubseteq x)[u \in kT]] \\
&\iff (\forall u \sqsubseteq x)[u \in T] \;\&\; (\forall u \sqsubseteq x)[u \in kT] \\
&\iff (\forall u \sqsubseteq x)[u \in kT] \\
&\iff x \in [kT],
\end{aligned}
$$

and it is enough to prove that kT is splitting. Suppose, towards a contradiction that some $u \in kT$ does not split. This means that all extensions of u in kT are compatible and they define a single point

$$x = \sup\{v \in kT \mid u \sqsubseteq v\}.$$

Since every extension of u in kT approximates x,

$$[T_u] = \{x\} \cup \bigcup\{[T_v] \mid u \sqsubseteq v \in T \;\&\; |[T_v]| \leq_c \aleph_0\};$$

this, however, implies that $[T_u]$ is a countable union of countable sets, which is absurd.

We leave the uniqueness of the decomposition (10-12) for the problems, **x10.2**.   ⊣

10.12. Definition. A family Γ of pointsets has **property P** if every uncountable set in Γ contains a non-empty, perfect subset. In this classical terminology,[25] the family $\mathcal{F}$ of closed pointsets has property **P**, or (more simply) *every closed pointset has property* **P**.

10.13. Exercise. *If a family of pointsets Γ has property* **P***, then the family Γ_σ of all countable unions of sets in Γ also has property* **P**.

Thus every $\mathcal{F}_\sigma$ pointset, of the form

$$A = \bigcup\nolimits_{n \in \mathbb{N}} F_n \tag{10-13}$$

---

[25]The classical terminology in question is quite absurd, but so well established that it would be folly to change it or bypass it. In any topological space, closed sets are $\mathcal{F}$-sets, from the French *fermèt*; open sets are $\mathcal{G}$-sets, from the German *Gebiete* (it means *region*); countable unions of Γ-sets are Γ_σ-sets, and countable intersections of Γ-sets are Γ_δ-sets, from the German words *Summe* and *Durchschnit* for *union* and *intersection*, respectively. We will only use this terminology in passing references to $\mathcal{F}_\sigma$ and $\mathcal{G}_\delta$ pointsets.

with each F_n closed has property **P**. The same is true of every $\mathcal{G}_\delta$ set, of the form

$$A = \bigcap_{n \in \mathbb{N}} G_n \tag{10-14}$$

with each G_n open, but the proof is not that simple, and it is ultimately easier to establish directly the property **P** for the much larger family of analytic pointsets.

10.14. Definition. Recall from **6.25** that a function $f : X \to Y$ on one topological space to another is **continuous** if the inverse image $f^{-1}[G]$ of every open set in Y is open in X. *A pointset $A \subseteq \mathcal{N}$ is* **analytic** *or* **Suslin***, if either $A = \emptyset$ or A is the image of Baire space under a continuous function, in symbols,*

$$\mathcal{A} =_{\mathrm{df}} \{A \subseteq \mathcal{N} \mid A = \emptyset \vee (\exists \text{ continuous } f : \mathcal{N} \to \mathcal{N})[A = f[\mathcal{N}]]\}.$$

Continuity in $\mathcal{N}$ has a simple, combinatorial interpretation which is the key to its applications.

10.15. Theorem. *A function $f : \mathcal{N} \to \mathcal{N}$ is continuous if and only if there exists a monotone function $\tau : \mathbb{N}^* \to \mathbb{N}^*$ on strings, such that*

$$\begin{aligned} f(x) &= \sup\{\tau(u) \mid u \sqsubseteq x\} \quad (x \in \mathcal{N}) \\ &= \lim_n \tau(\overline{x}(n)). \end{aligned} \tag{10-15}$$

When $\tau : \mathbb{N}^* \to \mathbb{N}^*$ is monotone and (10-15) holds, we say that τ **computes** the function f.

PROOF. If f satisfies (10-15), then

$$f(x) \in \mathcal{N}_v \iff (\exists u \sqsubseteq x)[v \sqsubseteq \tau(u)],$$

so each inverse image of a neighborhood

$$f^{-1}[\mathcal{N}_v] = \bigcup \{\mathcal{N}_u \mid v \sqsubseteq \tau(u)\}$$

is a union of neighborhoods and f is continuous.

For the more difficult converse, suppose f is continuous and let

$$S(u) =_{\mathrm{df}} \{v \in \mathbb{N}^* \mid f[\mathcal{N}_u] \subseteq \mathcal{N}_v\} \quad (u \in \mathbb{N}^*).$$

Each $S(u) \neq \emptyset$, since the root $\emptyset \in S(u)$; $v \sqsubseteq v' \in S(u) \Longrightarrow v \in S(u)$; and

$$v, v' \in S(u) \Longrightarrow f[\mathcal{N}_u] \subseteq \mathcal{N}_v \cap \mathcal{N}_{v'} \Longrightarrow [v \sqsubseteq v' \vee v' \sqsubseteq v],$$

since $v \mid v' \Longrightarrow \mathcal{N}_v \cap \mathcal{N}_{v'} = \emptyset$. Thus, there are two possibilities:

CASE 1. There is some $v \in S(u)$ such that $\mathrm{lh}(v) = \mathrm{lh}(u)$. In this case we set

$$\tau(u) =_{\mathrm{df}} v = \text{ the unique string in } S(u) \text{ such that } \mathrm{lh}(v) = \mathrm{lh}(u).$$

CASE 2. There is no $v \in S(u)$ such that $\mathrm{lh}(v) = \mathrm{lh}(u)$. Now we set

$$\tau(u) =_{\mathrm{df}} \sup\{v \mid v \in S(u)\}.$$

The monotonicity of τ follows easily from the implications

$$u_1 \sqsubseteq u_2 \Longrightarrow f[\mathcal{N}_{u_1}] \supseteq f[\mathcal{N}_{u_2}] \Longrightarrow S(u_1) \subseteq S(u_2),$$

considering the possibilities in the definitions of $\tau(u_1)$ and $\tau(u_2)$. To prove (10-15), notice first that because $\tau(u) \in S(u)$,

$$u \sqsubseteq x \in \mathcal{N} \Longrightarrow f(x) \in \mathcal{N}_{\tau(u)} \Longrightarrow \tau(u) \sqsubseteq f(x).$$

Moreover, by the continuity of f, if $v \sqsubseteq f(x)$, then there is some $u \sqsubseteq x$ such that $f[\mathcal{N}_u] \subseteq \mathcal{N}_v$, hence $v \in S(u)$ and either immediately $v \sqsubseteq \tau(u)$, if $\tau(u)$ is defined by Case 2, or there is some u' extending u, with $\mathrm{lh}(u') = \mathrm{lh}(v)$ such that $v = \tau(u')$ in the other case. ⊣

It is useful to think of (10-15) as a computational characterization of continuity: the string function $\tau(u)$ gives us better and better approximations $\tau(u) \sqsubseteq f(x)$ to the value of f, as we feed into it successively finer approximations $u \sqsubseteq x$ to the argument. We can turn this picture into a precise and elegant result, in terms of the notions introduced in Chapter **6**.

10.16. Corollary. *A function $f : \mathcal{N} \to \mathcal{N}$ is continuous if and only if it is the restriction to $\mathcal{N}$ of some monotone, continuous mapping*

$$\pi : (\mathbb{N} \rightharpoonup \mathbb{N}) \to (\mathbb{N} \rightharpoonup \mathbb{N})$$

on the inductive poset $(\mathbb{N} \rightharpoonup \mathbb{N})$. By Definition **6.22**, a monotone mapping $\pi : (\mathbb{N} \rightharpoonup \mathbb{N}) \to (\mathbb{N} \rightharpoonup \mathbb{N})$ is continuous if it satisfies the equivalence

$$\pi(x)(i) = w \iff (\exists u \in \mathbb{N}^*)[u \sqsubseteq x \;\&\; \pi(u)(i) = w].$$

Here we are using the fact that $\mathcal{N}$ is a subset of $(\mathbb{N} \rightharpoonup \mathbb{N})$, consisting precisely of all its maximal elements, and the basic observation is the decomposition

$$(\mathbb{N} \rightharpoonup \mathbb{N}) = \mathbb{N}^* \cup \mathcal{N}, \quad \mathbb{N}^* \cap \mathcal{N} = \emptyset. \tag{10-16}$$

Proof. If $f : \mathcal{N} \to \mathcal{N}$ is continuous, let τ compute it by the Theorem and take (literally)

$$\pi = \tau \cup f,$$

i.e., $\pi(u) = \tau(u)$ for $u \in \mathbb{N}^*$ and $\pi(x) = f(x)$ for $x \in \mathcal{N}$. The continuity of π is trivial. The converse is very easy. ⊣

10.17. Exercise. *Prove the "easy converse", i.e., that if $f : \mathcal{N} \to \mathcal{N}$ is the restriction to $\mathcal{N}$ of some continuous $\pi : (\mathbb{N} \rightharpoonup \mathbb{N}) \to (\mathbb{N} \rightharpoonup \mathbb{N})$, then f is continuous.*

The Corollary makes it possible to recognize continuity of specific functions on Baire space instantly, by inspection, simply noticing that every digit $f(x)(i)$ of each value $f(x)$ can be computed using only finitely many values of x. As in Chapter **6**, a passing remark that some function or other is "evidently continuous" accompanied by no proof typically means an appeal to this result.

10.18. Definition. A pointset $K \subseteq \mathcal{N}$ is **compact** if $K = [T]$ with a tree T on $\mathbb{N}$ which is finitely branching. In particular, every compact pointset is closed, and $\mathcal{C}$ is compact.

Some cheating is involved in adopting this as the definition of compactness for pointsets, since there is a perfectly general definition of compactness for sets in arbitrary topological spaces, by which **10.18** is a theorem. Without

comment, we did the same for perfection, which is also a general, topological notion. What we need here are the combinatorial properties of these pointsets specific to Baire space and we have relegated their topological characterizations to the problems, **x10.17** and **x10.21**.

10.19. Proposition. (1) *The image $f[K]$ of a compact pointset K by a continuous function $f : \mathcal{N} \to \mathcal{N}$ is compact.*

(2) *The image $f[K]$ of a compact and perfect pointset K by a continuous injection $f : \mathcal{N} \rightarrowtail \mathcal{N}$ is compact and perfect.*

PROOF. (1) Suppose $K = [T]$ where T is finitely branching and τ computes f as in (10-15), and let

$$S = T^{f[K]} = \{v \mid (\exists x \in K)[v \sqsubseteq f(x)]\}$$

be the tree of initial segments of the image $f[K]$. It is enough to prove that S is a finitely branching tree and $f[K] = [S]$.

To see first that S is finitely branching, suppose $v \in S$, let

$$B =_{\mathrm{df}} \{u \in T \mid v \sqsubseteq \tau(u) \vee v \mid \tau(u)\},$$

and suppose that $x \in [T]$. If $v \mid f(x)$, then for some n, $v \mid \tau(\overline{x}(n))$, so $\overline{x}(n) \in B$; and if $v \sqsubseteq f(x)$, then for some n, $v \sqsubseteq \tau(\overline{x}(n))$, and again $\overline{x}(n) \in B$. Thus, B is a bar for T, and by the Fan Theorem **9.9** it must have a finite subset

$$B_0 = \{u_0, \dots, u_n\} \subseteq B$$

which is also a bar. Thus, for every $x \in K$ such that $v \sqsubseteq f(x)$, there exists some u_i such that $v \sqsubseteq \tau(u_i) \sqsubseteq f(x)$, so that every child of v in S is an initial segment of some $\tau(u_i)$, and there are only finitely many of those.

Clearly, $f[K] \subseteq [S]$. To prove $[S] \subseteq f[K]$, suppose towards a contradiction that $y \in [S] \setminus f[K]$ and let

$$B =_{\mathrm{df}} \{u \in T \mid \tau(u) \mid y\}.$$

Now B is a bar for T, because the only way that $\tau(\overline{x}(n))$ can be compatible with y for every n is if $f(x) = y$. By the Fan Theorem again, there is a finite subset

$$B_0 = \{u_0, \dots, u_n\} \subseteq B$$

which is also a bar for T. Let

$$k = \max\{\mathrm{lh}(\tau(u_i)) \mid i \leq n\} + 1,$$

and choose some $x \in [T]$ such that $\overline{y}(k) \subseteq f(x)$, which exists because $y \in [S]$, so that y can be approximated arbitrarily well by points in the image of f. On the other hand, $u_i \sqsubseteq x$ for some i since B_0 is a bar; hence $\tau(u_i) \sqsubseteq f(x)$ because τ computes f; so both $\tau(u_i)$ and $\overline{y}(k)$ are initial segments of $f(x)$, and hence compatible; and since $\tau(u_i)$ has smaller length than $\overline{y}(k)$, this means that $\tau(u_i) \sqsubseteq y$, which contradicts the definition of B.

(2) With the same notation as in (1) and the additional hypothesis, let $v \in S$, so that for some $u \in T$, $v \sqsubseteq \tau(u)$. Since T is splitting, there exist distinct points

$$x_1, x_2 \in K \cap \mathcal{N}_u,$$

and since τ computes f,

$$\tau(u) \sqsubseteq f(x_1), \quad \tau(u) \sqsubseteq f(x_2). \tag{10-17}$$

But $f(x_1) \neq f(x_2)$, because f is an injection, so there exist incompatible $v_1 \sqsubseteq f(x_1)$, $v_2 \sqsubseteq f(x_2)$, which extend $\tau(u)$ by (10-17), so they split $\tau(u)$ and hence the smaller $v \sqsubseteq \tau(u)$ in S.   ⊣

10.20. Perfect Set Theorem (Suslin, 1916). *Every uncountable, analytic set has a non-empty, perfect subset.*

PROOF. Assume that $A = f[\mathbb{N}^*]$ is uncountable, suppose τ computes f, and let

$$T =_{\mathrm{df}} \{u \in \mathbb{N}^* \mid |f[\mathcal{N}_u]| >_c \aleph_0\}. \tag{10-18}$$

Clearly, T is a non-empty tree.

Lemma. *The tree T is τ-splitting, i.e., for each $u \in T$, there exist $u_1, u_2 \in T$ such that*

$$u \sqsubseteq u_1, \quad u \sqsubseteq u_2, \quad \tau(u_1) \mid \tau(u_2).$$

Proof. For any $u \in T$ and any fixed $x \in \mathcal{N}_u$,

$$f[\mathcal{N}_u] = \{f(x)\} \cup \bigcup \{f[\mathcal{N}_{u'}] \mid \tau(u') \mid f(x)\} \tag{10-19}$$

since $f(y) \neq f(x) \Longrightarrow \tau(u') \sqsubseteq f(y)$ for some u' such that $\tau(u')$ is incompatible with $f(x)$. If the Lemma fails at u, then

$$u \sqsubseteq u' \in T \Longrightarrow \tau(u') \sqsubseteq f(x);$$

thus each image $f[\mathcal{N}_{u'}]$ with $\tau(u') \mid f(x)$ in (10-19) involves some $u' \notin T$ and is countable, and there are only countably many choices for u'. Thus, $f[\mathcal{N}_u]$ is the union of a singleton and a countable family of countable sets, hence countable, contrary to hypothesis.   ⊣ (Lemma)

As in **10.10**, we choose functions

$$l : T \to T, \quad r : T \to T$$

which witness the τ-splitting property for T, i.e., for each $u \in T$,

$$u \sqsubseteq l(u), \quad u \sqsubseteq r(u), \quad \tau(l(u)) \mid \tau(r(u)),$$

and we define by the String Recursion Theorem **5.33** a function

$$\sigma : \{0, 1\}^* \rightharpoonup T$$

from the tree of all binary strings into T which satisfies the identities

$$\sigma(\emptyset) = \emptyset, \quad \sigma(u \star \langle 0 \rangle) = l(\sigma(u)), \quad \sigma(u \star \langle 1 \rangle) = r(\sigma(u)),$$

and which, as a consequence, is monotone. The key property of this σ is that it takes incompatible binary strings to τ-incompatible strings,

$$u \mid v \Longrightarrow \tau(\sigma(u)) \mid \tau(\sigma(v)), \tag{10-20}$$

and it is verified exactly as (10-11) was verified in the proof of **10.10**. In addition, σ computes a continuous $g : \mathcal{C} \to \mathcal{N}$,

$$g(x) = \sup \{\sigma(u) \mid u \sqsubseteq x\},$$

and evidently

$$g[\mathcal{C}] \subseteq [T]. \tag{10-21}$$

Consider now the composition $h = fg$ of the given f and this g, which is computed by the composition of τ and σ:

$$h(x) = \sup \{\tau(\sigma(u)) \mid u \sqsubseteq x\}.$$

This is continuous and injective by (10-20), so its image $h[\mathcal{C}] = fg[\mathcal{C}]$ is compact and perfect by **10.19**, and it is included in $f[T] \subseteq A$ by (10-21). $\quad\dashv$

The result means nothing, of course, until we prove that there are lots and lots of analytic sets.

10.21. Lemma. *Every closed pointset is analytic.*

PROOF. Let $T = T^F$ as in (10-8) for the given closed set $F \neq \emptyset$, so in addition to $F = [T]$ we also know that every string in T has an extension, i.e., there are no terminal nodes. Thus, we can fix a function $l : T \to T$ such that

$$u \in T \Longrightarrow u \sqsubseteq l(u) \ \& \ \mathrm{lh}(l(u)) = \mathrm{lh}(u) + 1.$$

Let also

$$\mathrm{rtail}(u) =_{\mathrm{df}} u \upharpoonright [0, \mathrm{lh}(u) - 1) \quad (\mathrm{lh}(u) > 0) \tag{10-22}$$

be the partial function which strips each non-empty string of its last element. By the String Recursion Theorem **5.33**, there is a function $\tau : \mathbb{N}^* \rightharpoonup T$ such that

$$\tau(u) = \begin{cases} u, & \text{if } u \in T, \\ l(\tau(\mathrm{rtail}(u))), & \text{if } u \notin T, \end{cases}$$

which is (easily) a *projection* of $\mathbb{N}^*$ onto T, i.e., it is total, length preserving and the identity on T, and which (as a consequence) computes a function $f : \mathcal{N} \twoheadrightarrow [T]$. $\quad\dashv$

10.22. Lemma. *Every continuous image of an analytic pointset is analytic.*

PROOF. If $A = f[B]$ and $B = g[\mathcal{N}]$, then $A = fg[\mathcal{N}]$, and the composition fg is continuous. $\quad\dashv$

10.23. Lemma. *If $f, g : \mathcal{N} \to \mathcal{N}$ are continuous functions, then the set*

$$E = \{x \mid f(x) = g(x)\}$$

of points on which they agree is closed.

PROOF. Because distinct points can be approximated by incompatible initial segments,

$$x \notin E \iff f(x) \neq g(x)$$
$$\iff (\exists u, v)[f(x) \in \mathcal{N}_u \;\&\; g(x) \in \mathcal{N}_v \;\&\; u \mid v],$$

which means that

$$cE = \bigcup \{ f^{-1}[\mathcal{N}_u] \cap g^{-1}[\mathcal{N}_v] \mid u \mid v \},$$

so that cE is the union of open sets and hence open.   ⊣

10.24. Theorem. *Countable unions and countable intersections of analytic pointsets are analytic.*

PROOF. Suppose that $A_n = f_n[\mathcal{N}]$ with each f_n continuous and define first $f : \mathcal{N} \to \mathcal{N}$ by the formula

$$f(z) = f_{z(0)}(\mathrm{tail}(z)),$$

where

$$\mathrm{tail}(z) =_{\mathrm{df}} (i \mapsto z(i+1)) = (z(1), z(2), \dots)$$

is the function which decapitates points. Evidently f is continuous: because each $f(z)(i)$ can be computed from finitely many values of z, first setting $n = z(0)$ and then using the finitely many values of $\mathrm{tail}(z)$ needed to compute $f_n(\mathrm{tail}(z))$. Moreover:

$$y \in \bigcup\nolimits_n f_n[\mathcal{N}] \iff (\exists n \in \mathbb{N}, x \in \mathcal{N})[y = f_n(x)]$$
$$\iff (\exists z \in \mathcal{N})[y = f_{z(0)}(\mathrm{tail}(z))]$$
$$\text{taking } z(0) = n, \mathrm{tail}(z) = x$$
$$\iff (\exists z \in \mathcal{N})[y = f(z)],$$

so $\bigcup\nolimits_n A_n = f[\mathcal{N}]$ and the union of the A_n's is analytic.

The key fact for this argument was that the mapping

$$z \mapsto (z(0), \mathrm{tail}(z))$$

is a surjection of $\mathcal{N}$ onto $\mathbb{N} \times \mathcal{N}$—actually a bijection—with continuous components. To prove that the intersection $\bigcap\nolimits_n A_n$ is analytic, we need a similar surjection

$$\pi : \mathcal{N} \twoheadrightarrow (\mathbb{N} \to \mathcal{N})$$

of $\mathcal{N}$ onto the set of infinite sequences of points. To construct such a π, fix some bijection $\rho : \mathbb{N} \times \mathbb{N} \rightarrowtail\!\!\!\to \mathbb{N}$ and set

$$\rho_n(z) = (i \mapsto z(\rho(n, i))). \tag{10-23}$$

Each $\rho_n : \mathcal{N} \to \mathcal{N}$ is clearly continuous and for each infinite sequence $\{x_n\}_{n \in \mathbb{N}}$ of points we can define z such that

$$z(\rho(n, i)) = x_n(i) \quad (n, i \in \mathbb{N}),$$

so that

$$\rho_n(z) = x_n \quad (n \in \mathbb{N});$$

in other words, the mapping

$$\pi(z) = (n \mapsto \rho_n(z))$$

is a surjection. Using $\mathbf{AC}_{\mathbb{N}}$ now,

$$
\begin{aligned}
y \in \bigcap_n A_n &\iff (\forall n)(\exists x)[y = f_n(x)] \\
&\iff (\exists \{x_n\}_{n \in \mathbb{N}})(\forall n)[y = f_n(x_n)] \\
&\iff (\exists z \in \mathcal{N})(\forall n)[y = f_n(\rho_n(z))] \\
&\iff (\exists z \in \mathcal{N})[(\forall n)[f_n(\rho_n(z)) = f_0(\rho_0(z))] \\
&\qquad\qquad\qquad\qquad\qquad \& \; y = f_0(\rho_0(z))].
\end{aligned}
\tag{10-24}
$$

For each n, the set

$$B_n = \{z \in \mathcal{N} \mid f_n(\rho_n(z)) = f_0(\rho_0(z))\}$$

is closed by **10.23**, hence the intersection

$$B = \bigcap_n B_n$$

is also closed. From (10-24), however,

$$\bigcap_n A_n = f_0 \rho_0[B],$$

which means that the intersection of the A_n's is analytic.   ⊣

10.25. Definition. The family $\mathcal{B}(X)$ of the **Borel** subsets of a topological space X is the smallest family of subsets of X which includes the open sets and is a σ-**field**, i.e., it is closed under countable unions and complementation:

$$
\begin{aligned}
(\forall n)[A_n \in \mathcal{B}(X)] &\implies \textstyle\bigcup_n A_n \in \mathcal{B}(X), \\
A \in \mathcal{B}(X) &\implies cA \in \mathcal{B}(X).
\end{aligned}
$$

We are mostly interested in Baire space of course,

$$\mathcal{B} =_{\mathrm{df}} \mathcal{B}(\mathcal{N}) = \text{the family of Borel pointsets}.$$

10.26. Exercise. *Prove that the definition makes sense, i.e., the intersection*

$$
\begin{aligned}
\mathcal{B}(X) = \bigcap \{\mathcal{E} \mid \;& \mathcal{G} \subseteq \mathcal{E} \\
& \& \, (\forall \{A_n\}_n \subseteq \mathcal{E})[\textstyle\bigcup_n A_n \in \mathcal{E}] \\
& \& \, (\forall A \in \mathcal{E})[cA \in \mathcal{E}]\}
\end{aligned}
$$

is a σ-field which contains the open sets, and hence the least such.

10.27. Exercise. *The intersection $\bigcap_n A_n$ of every sequence of Borel sets $\{A_n\}$ is a Borel set.*

10.28. Corollary. *Every Borel pointset is analytic* (Suslin) *and hence has property* **P** (Alexandroff, Hausdorff).

PROOF. Let

$$C\mathcal{A} = \{A \subseteq \mathcal{N} \mid cA \in \mathcal{A}\} \tag{10-25}$$

be the family of **co-analytic pointsets**, those with analytic complements. The family $\mathcal{A} \cap C\mathcal{A}$ of pointsets which are both analytic and co-analytic is a
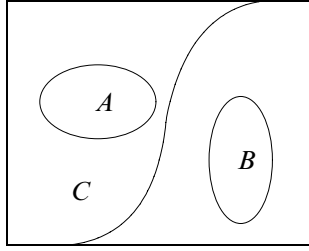
FIGURE **10.2**. C separates A from B.

σ-field, since it is closed under complementation by definition, and if each $A_n \in \mathcal{A} \cap C\mathcal{A}$, then $\bigcup_n A_n$ and

$$c(\textstyle\bigcup_n A_n) = \bigcap_n cA_n$$

are both analytic by the theorem. Since every open set

$$G = \textstyle\bigcup_n \{\mathcal{N}_u \mid \mathcal{N}_u \subseteq G\}$$

is a countable union of neighborhoods, hence analytic, and also co-analytic by **10.21**, $\mathcal{A} \cap C\mathcal{A}$ is a σ-field which contains all the open pointsets and hence includes every Borel set. ⊣

In the next two theorems we clarify somewhat the relation between analytic and Borel pointsets.

A pointset $C \subseteq \mathcal{N}$ *separates* a pointset A from another pointset B, if

$$A \subseteq C, \quad C \cap B = \emptyset.$$

Notice that if some C separates A from B, then $A \cap B = \emptyset$.

10.29. Lemma. (1) *If* $\{A_i\}$, $\{B_j\}$ *are two sequences of pointsets and for all* i, j, C_{ij} *separates* A_i *from* B_j, *then the set* $C = \bigcup_i \bigcap_j C_{ij}$ *separates* $\bigcup_i A_i$ *from* $\bigcup_j B_j$, *i.e.*,

$$\textstyle\bigcup_i A_i \subseteq \bigcup_i \bigcap_j C_{ij}, \quad \left(\bigcup_i \bigcap_j C_{ij}\right) \cap \bigcup_j B_j = \emptyset. \tag{10-26}$$

(2) *If* $\{A_i\}$, $\{B_j\}$ *are two sequences of pointsets and no Borel set separates* $A = \bigcup_i A_i$ *from* $B = \bigcup_j B_j$, *then there exist two numbers* i_0 *and* j_0 *such that no Borel set separates* A_{i_0} *from* B_{j_0}.

PROOF. (1) For any fixed i and all j, by hypothesis, $A_i \subseteq C_{ij}$, and so

$$A_i \subseteq \textstyle\bigcap_j C_{ij};$$

so taking the unions of both sides, we get

$$A = \textstyle\bigcup_i A_i \subseteq \bigcup_i \bigcap_j C_{ij},$$

which is the first inclusion claimed by the Lemma. For the second, we notice that the hypothesis $B_j \cap C_{ij} = \emptyset$ means exactly that

$$B_j \subseteq cC_{ij} \quad (cC_{ij} = \mathcal{N} \setminus C_{ij});$$

and so, fixing i and taking unions again, we get

$$B = \bigcup_j B_j \subseteq \bigcup_j C_{ij},$$

which, since i was arbitrary, yields

$$B \subseteq \bigcap_i \bigcup_j cC_{ij}.$$

Now Problem **x1.3** (De Morgan's laws) yields

$$\bigcap_i \bigcup_j cC_{ij} = c\left(\bigcup_i \bigcap_j C_{ij}\right),$$

so that $B \cap \left(\bigcup_i \bigcap_j C_{ij}\right) = \emptyset$, which was what we needed to prove.

(2) follows easily, by contradiction: because if some Borel set C_{ij} separated each A_i from each B_j, then $\bigcup_i \bigcap_j C_{ij}$ separates A from B—and it is a Borel set. ⊣

10.30. The Separation Theorem (Lusin) *If $A, B \subseteq \mathcal{N}$ are analytic pointsets and $A \cap B = \emptyset$, then there exists a Borel pointset C which separates A from B.*

PROOF. Suppose $A = f[\mathcal{N}], B = g[\mathcal{N}]$, where f, g are continuous functions which by Theorem **10.15** are computed by given, monotone string functions $\sigma, \tau : \mathbb{N}^* \to \mathbb{N}^*$,

$$f(x) = \lim_n \sigma(\overline{x}(n)), \quad g(y) = \lim_n \tau(\overline{y}(n)).$$

For any two strings u, v, put

$$A_u = f[\mathcal{N}_u] = \{f(x) \mid u \sqsubseteq x\}, \quad B_v = g[\mathcal{N}_v] = \{g(y) \mid v \sqsubseteq y\},$$

and record the fact that

$$A_u \subseteq \mathcal{N}_{\sigma(u)}, \quad B_v \subseteq \mathcal{N}_{\tau(v)}, \tag{10-27}$$

which follows directly from what it means for σ to compute f and τ to compute g. Notice also that $A_\emptyset = A, B_\emptyset = B$, and, easily, for all u, v,

$$A_u = \bigcup_i A_{u \star \langle i \rangle}, \quad B_v = \bigcup_j B_{v \star \langle j \rangle}. \tag{10-28}$$

We now assume towards a contradiction that no Borel set separates A from B and we apply repeatedly Lemma **10.29**, using (10-28): since no Borel set separates $A = A_\emptyset$ from $B = B_\emptyset$, there exist numbers i_0, j_0 such that no Borel set separates $A_{\langle i_0 \rangle}$ from $B_{\langle j_0 \rangle}$; and so, there exist i_1, j_1 such that no Borel set separates $A_{\langle i_0, i_1 \rangle}$ from $B_{\langle j_0, j_1 \rangle}$; and so, etc. Formally, we define recursively two sequences of numbers

$$x = (i_0, i_1, \dots), \ y = (j_0, j_1, \dots) \in \mathcal{N}$$

such that for all n, *no Borel set separates $A_{\overline{x}(n)}$ from $B_{\overline{y}(n)}$.* Now this means that,

$$\text{for all } n, \mathcal{N}_{\sigma(\overline{x}(n))} \cap \mathcal{N}_{\tau(\overline{y}(n))} \neq \emptyset, \tag{10-29}$$

since otherwise $\mathcal{N}_{\sigma(\overline{x}(n))}$ would separate $A_{\overline{x}(n)}$ from $B_{\overline{y}(n)}$, by (10-27); and, finally, (10-29) says directly that

$$f(x) = g(y) \in \bigcap_n \left(\mathcal{N}_{\sigma(\overline{x}(n))} \cap \mathcal{N}_{\tau(\overline{y}(n))}\right),$$

which contradicts the hypothesis, that $A \cap B = \emptyset$.   ⊣

10.31. Suslin's Theorem. *A pointset $A \subseteq \mathcal{N}$ is Borel if and only if both A and its complement cA are analytic.*

PROOF of the non-trivial direction is immediate, by applying the Separation Theorem to A and cA.   ⊣

Suslin introduced the family of analytic pointsets in 1917 and proved a slew of theorems about it, including the Perfect Set Theorem **10.20**, his famous characterization **10.31**, and that *not every analytic pointset is Borel*, which we will not prove here.[26] The Borel sets had been introduced more than a decade earlier by Borel and Lebesgue and they were the key to the successful development of the theory of *Lebesgue integration*, one of the chief achievements of 19th century analysis. For most purposes of integration theory, including its later, fundamental applications to *probability*, every pointset of interest is *almost equal* to a Borel set, in a precise sense which basically allows us to study the subject as if every pointset were Borel. Because of this, the Continuum Problem for Borel sets was considered very important, and its simultaneous, independent solutions published by Alexandroff and Haussdorff in 1916 (just before Suslin established the more general Theorem **10.20**) were celebrated as a major achievement.

The family of analytic sets falls far short from exhausting the powerset of $\mathcal{N}$, Problem **x10.9**. Still, one might hope that the method we used to solve the Continuum Problem for them might be extended to prove the full Continuum Hypothesis, but this too is far from the mark.

10.32. Theorem. (**AC**) *There exists a pointset $A \subset \mathcal{N}$ which is uncountable but contains no non-empty perfect set.*

PROOF. The key fact is that there are exactly as many non-empty, perfect sets as there are points in $\mathcal{N}$:

Lemma 1. *If $\mathscr{P} = \{P \subseteq \mathcal{N} \mid P \neq \emptyset, P \text{ perfect}\}$, then $|\mathscr{P}| =_c \mathfrak{c}$.*

Proof. For each $y \in \mathcal{N}$, the pointset

$$A_y = \{x \mid (\forall n)[y(n) \leq x(n)]\}$$

is easily perfect, equally easily $y \neq z \Longrightarrow A_y \neq A_z$, hence $\mathfrak{c} =_c |\mathcal{N}| \leq_c |\mathscr{P}|$. On the other hand, each perfect set $P = [T^P]$ is the body of a tree on $\mathbb{N}$ which determines it, so

$$|\mathscr{P}| \leq_c |\mathcal{P}(\mathbb{N}^*)| =_c |\mathcal{P}(\mathbb{N})| =_c \mathfrak{c}. \qquad \dashv \text{(Lemma 1)}$$

Fix a set

$$I =_c \mathfrak{c} =_c \mathscr{P}, \tag{10-30}$$

---

[26]The study of analytic and Borel pointsets is the core of *Descriptive Set Theory*, one of the most beautiful parts of our subject, but (unfortunately) beyond the scope of these Notes.

for example $I = \mathfrak{c}$, and bijections

$$\alpha \mapsto x_\alpha \in \mathcal{N}, \quad \alpha \mapsto P_\alpha \in \mathscr{P} \quad (\alpha \in I)$$

which witness the equinumerosities (10-30). Fix also a best wellordering $\leq$ of I. We will define by transfinite recursion on $(I, \leq)$ injections

$$f_\alpha : \mathbf{seg}(\alpha) \rightarrowtail A_\alpha \subset \mathcal{N}, \; g_\alpha : \mathbf{seg}(\alpha) \twoheadrightarrow B_\alpha \subset \mathcal{N} \quad (\alpha \in I),$$

so that the following conditions hold.

(1) If $\alpha \leq \beta$, then $f_\alpha \subseteq f_\beta, g_\alpha \subseteq g_\beta$, so that $A_\alpha \subseteq A_\beta$ and $B_\alpha \subseteq B_\beta$.
(2) For each $\alpha \in I$, $A_\alpha \cap B_\alpha = \emptyset$.
(3) For each $\alpha \in I$, $B_{S\alpha} \cap P_\alpha \neq \emptyset$, where S is the successor function in the well ordered set $(I, \leq)$.

Lemma 2. *If f_α, g_α ($\alpha \in I$) satisfy* (1) $-$ (3), *then*

$$A = \bigcup_{\alpha \in I} A_\alpha =_c \mathfrak{c},$$

but A has no non-empty, perfect subset.

Proof. That $A =_c I =_c \mathfrak{c}$ follows immediately from (1), since

$$\bigcup_\alpha f_\alpha : I \rightarrowtail\!\!\!\twoheadrightarrow A \text{ and } I =_c \mathfrak{c}.$$

For the second claim, the key observation is that

$$A \cap B_\beta = \emptyset \quad (\beta \in I);$$

this is because if $x \in A_\alpha \cap B_\beta$, then with $\gamma = \max\{\alpha, \beta\}$, by (1), $x \in B_\gamma \cap B_\gamma$, contradicting (2). Now if $P \neq \emptyset$ is perfect, then $P = P_\alpha$ for some $\alpha \in I$, hence there exists some $x \in P_\alpha \cap B_{S\alpha}$ and then $x \notin A$, so $P_\alpha \not\subseteq A$. $\dashv$ (Lemma 2)

The definitions of f_α, g_α are practically forced on us by conditions (1) $-$ (3). We outline the proofs of (1) $-$ (3) by transfinite induction together with the clauses of the transfinite recursion definition; pedantically the proof should be separated out and explained after the definition is completed.

(a) At the minimum 0 of I, set $f_0 = g_0 = \emptyset$.
(b) If λ is a limit point of I, set

$$f_\lambda = \bigcup_{\alpha < \lambda} f_\alpha, \quad g_\lambda = \bigcup_{\alpha < \lambda} g_\alpha.$$

Conditions (1) and (2) follow immediately from the induction hypothesis, and (3) is not relevant to this case.

(c) Suppose $\beta = S\alpha$ is a successor point in I. By the induction hypothesis, A_α and B_α are equinumerous with $\mathbf{seg}(\alpha)$ and $\mathbf{seg}(\alpha) <_c I$, because $\leq$ is a best wellordering. Thus, $|A_\alpha|, |B_\alpha|$ are both smaller than $\mathfrak{c}$, hence $|A_\alpha \cup B_\alpha| <_c \mathfrak{c}$, and we can find in the non-empty, perfect set $P_\alpha =_c \mathcal{N}$ distinct points

$$x, y \in P_\alpha \setminus (A_\alpha \cup B_\alpha);$$

we set

$$f_\beta = f_\alpha \cup \{(\alpha, x)\}, \quad g_\beta = g_\alpha \cup \{(\alpha, y)\},$$

and (1) $-$ (3) follow easily. $\dashv$

The construction obviously proves more than what is claimed in the theorem: $|A| =_c \mathfrak{c}$ and both A and its complement cA intersect every non-empty, perfect set. We leave for the problems some additional variations which make it even more obvious that the program of proving the Continuum Hypothesis by using the Cantor-Bendixson Theorem is hopeless.

Actually, it is not only this program for settling the Continuum Problem which fails: every attempt to prove or disprove **CH** from the axioms of **ZDC + AC** is doomed, by the following two central independence results.

10.33. Consistency of GCH, the Generalized Continuum Hypothesis (Gödel, 1939). The model L of constructible sets satisfies the Generalized Continuum Hypothesis **GCH**, (9-7), so, in particular, the Continuum Hypothesis cannot be refuted in **ZDC+AC**.

10.34. Independence of the Continuum Hypothesis CH (Cohen, 1963). There is a model of **ZDC+AC** in which the Continuum Hypothesis fails, hence **CH** cannot be proved in **ZDC+AC**. Cohen's forcing model can be modified in many ways to manipulate the cardinalities of pointsets and subsets of larger powersets.

10.35. What does the independence of CH mean? Both the Gödel and the Cohen methods of proof are very robust and they have been adapted to show that the Continuum Problem cannot be settled on the basis of many reasonable and plausible strengthenings of **ZDC+AC** by additional axioms. The same is true of the Axiom of Choice, of course, or the Axiom of Infinity for that matter, but it is clear that these propositions express new, fundamental set theoretic principles which are most likely true but cannot (and, in fact, cannot be expected to) be proved from simpler axioms by logic alone. The Continuum Hypothesis has the look of a technical, mathematical problem which should be settled definitively by a proof, but we seem to lack the insight needed to divine the necessary axioms.

Much has been made of this independence of **CH** (and many more assertions about sets) from variants of the known axioms of set theory, and some have used it to argue against any objective reality behind the "formal", axiomatic results of the subject. Using the method of *arithmetization* introduced by Gödel, however, we can translate questions about the *existence of proofs* into precise, technical conjectures about integers: since there exist such conjectures[27] which (like **CH**) can be shown to be undecidable in the known, plausible axiomatic theories, are we then forced to deny objective reality to

---

[27] The type of statements we have in mind here are of the form "*if* **ZDC+AC** *is consistent, then so is T*", where T is some strong extension of **ZDC+AC** which, in fact, implies the consistency of **ZDC+AC**. Gödel's *Second Incompleteness Theorem* implies that statements of this type are independent of **ZDC+AC** (unless **ZDC+AC** is inconsistent), and there are many of them about whose truth there is genuine controversy.

the natural numbers? In fact, it is not possible to discuss such problems intelligently without reference to notions and results of Mathematical Logic which are beyond the scope of these Notes, and we will resist the temptation.

Incidentally, there are scores of interesting propositions about sets which cannot be settled on the basis of **ZDC** or **ZDC+AC**; **CH** is only the most interesting of them. We mention here just three more independence results of this type, because they are relevant to the Perfect Set Theorem **10.20**.

10.36. (Gödel, 1939) In the model L of constructible sets, there exists an uncountable, co-analytic set which has no proper perfect subset. This means that we cannot improve the Perfect Set Theorem **10.20** in **ZDC+AC** to show that every co-analytic pointset has property **P**.

10.37. (Solovay, 1970) There is a model of **ZDC+AC** in which every "definable" pointset has property **P**. We will not attempt to define "definable", but analytic complements are definable.

10.38. (Solovay, 1970) There is a model of **ZDC** in which every pointset has property **P**.

Solovay's models are constructed by Cohen's forcing method, but like Gödel's L, they have many more canonical properties which yield numerous unprovability results. The first Solovay model witnesses (with **10.36**) that the property **P** for analytic complements cannot be proved or refuted in **ZDC+AC**. The second Solovay model shows that **ZDC** cannot prove the existence of an uncountable pointset with no non-empty, perfect subset; **DC** is not a sufficiently strong choice principle to effect the construction.

Problems for Chapter 10

x10.1. Prove that if $F \subseteq \mathcal{N}$ is closed, then there is a unique tree T on $\mathbb{N}$ without terminal nodes such that $F = [T]$, namely the tree T^F defined in (10-8).

x10.2. Prove that the decomposition (10-12) of a closed pointset F into a perfect set P and a countable set S determines uniquely P and S.

x10.3. Give an example of a closed pointset $F \subseteq \mathcal{N}$ and a continuous $f : \mathcal{N} \to \mathcal{N}$, such that the image $f[F]$ is not closed.

x10.4. Prove that every open pointset is an $\mathcal{F}_\sigma$ and every closed pointset is a $\mathcal{G}_\delta$. The definitions are reviewed in Footnote 25.

****x10.5.** Prove that the inverse image $g^{-1}[A]$ of an analytic pointset A by a continuous function $g : \mathcal{N} \to \mathcal{N}$ is analytic. HINT. Aim for an equivalence of the form

$$y \in g^{-1}[A] \iff (\exists x)[y = f(\rho_1(x)) = g(\rho_2(x))]$$

where f is continuous and ρ_n are defined by (10-23), and then use **10.23**.

***x10.6.** Prove that

$$\mathcal{N} = \bigcup_{i \in \mathbb{N}} A_i \Longrightarrow (\exists i)[A_i =_c \mathcal{N}],$$

i.e., $\mathcal{N}$ is not the union of a countable sequence of pointsets of smaller cardinality. HINT. This follows easily from König's Theorem **9.21**, but the relevant special case does not need the full Axiom of Choice.

x10.7. (**AC**) Prove that for every $\kappa \leq_c \mathfrak{c}$, there exists a pointset A with $|A| =_c \kappa$ which contains no non-empty, perfect subset.

x10.8. (**AC**) Prove that there exists an uncountable pointset A such that neither A nor its complement contain an uncountable Borel set.

x10.9. Prove that there are $\mathfrak{c}$-many analytic and Borel pointsets,

$$|\mathcal{A}| =_c |\mathcal{B}| =_c \mathfrak{c}.$$

10.39. Definition. A function $f : X \to Y$ from one topological space into another is **Borel measurable** if the inverse image $f^{-1}[G]$ of every open subset of Y is a Borel subset of X.

x10.10. The composition $gf : X \to Z$ of two Borel measurable functions $f : X \to Y$ and $g : Y \to Z$ is Borel measurable.

x10.11. The inverse image $f^{-1}[A]$ of a Borel set $A \subseteq Y$ by a Borel measurable function $f : X \to Y$ is a Borel subset of X.

10.40. Definition. Two topological spaces X, Y are **Borel isomorphic** if there exists a Borel measurable bijection $f : X \rightarrowtail\!\!\!\rightarrow Y$, whose inverse $f^{-1} : Y \rightarrowtail\!\!\!\rightarrow X$ is also Borel measurable. Borel isomorphic spaces have the same measure-theoretic structure and for all practical purposes can be "identified" in measure theory.

***x10.12.** Suppose $f : X \rightarrowtail Y$ and $g : Y \rightarrowtail X$ are Borel measurable injections between topological spaces, with the following additional property:[28] there exists Borel measurable functions $f_1 : Y \to X$ and $g_1 : X \to Y$ which are inverses of f and g in the sense that

$$f_1 f(x) = x \quad (x \in X),$$
$$g_1 g(y) = y \quad (y \in Y).$$

Prove that X and Y are Borel isomorphic. HINT. Use the proof of the Schröder-Bernstein Theorem **2.26**.

x10.13. Consider the Cantor set $\mathcal{C}$ as a topological subspace of $\mathcal{N}$ in the obvious way, the open sets being unions of neighborhoods of the form

$$\mathcal{N}_u = \{x \in \mathcal{C} \mid u \sqsubseteq x\} \quad (u \in \{0, 1\}^*).$$

Prove that $\mathcal{C}$ and $\mathcal{N}$ are Borel isomorphic.

---

[28] In fact, every Borel injection has this property, but the proof of this requires some work.

In the remaining problems we explore the connection of the specific, combinatorial notions we studied in Baire space with their general, topological versions.

10.41. Definition. A point x is a **limit point** of a set A in a topological space X if every open set G which contains x contains also some point of A other than x,

$$(\forall G)[G \text{ open and } x \in G \Longrightarrow (\exists y \in A \cap G)[x \neq y]].$$

A limit point of A may or may not be a member of A. *A point of A which is not a limit point of A is* **isolated in** A.

x10.14. Determine the limit points and the isolated points of the pointset

$$B = \{x \in \mathcal{N} \mid x(0) = 1 \vee (\forall n)[x(n) = 2] \vee (\exists n)[x(n) = 3]\}.$$

x10.15. Prove that x is a limit point of A if and only if every open set containing x contains infinitely many points of A.

x10.16. Prove that a set is closed in a topological space X if and only if it contains all its limit points.

x10.17. Prove that a pointset P is perfect if and only if it is closed and has no isolated points, i.e., every point of P is a limit point of P. This equivalence identifies the specific definition of perfect pointsets we adopted with the classical, topological definition.

10.42. Definition. A sequence $(n \mapsto x_n)$ of points in a topological space X **converges** to a point x or has x as its **limit** if every open set containing x contains all but finitely many of the terms of the sequence,

$$\lim_n x_n = x \iff_{\mathrm{df}} (\forall G \text{ open}, x \in G)(\exists n \in \mathbb{N})(\forall i \geq n)[x_i \in G].$$

x10.18. Prove that a point x is a limit point of a pointset A if and only if $x = \lim_n x_n$ is the limit of some sequence $(n \mapsto x_n \in A)$ of points in A. Which choice principle did you use, if any?

x10.19. Prove that a function $f : \mathcal{N} \to \mathcal{N}$ is continuous if and only if

$$f(\lim_n x_n) = \lim_n f(x_n),$$

whenever $\lim_n x_n$ exists. Which choice principle did you use, if any?

x10.20. *A topological space X is* **Hausdorff** *if for any two points $x \neq y$, there exist disjoint open sets $G \cap H = \emptyset$ such that $x \in G$ and $y \in H$.* Prove that if $f, g : X \to Y$ are continuous functions and Y is Hausdorff, then the set $\{x \in X \mid f(x) = g(x)\}$ is closed in X.

10.43. Definition. An **open covering** of a set K in a topological space X is any family $\mathcal{G}$ of open sets whose union includes K, $K \subseteq \bigcup \mathcal{G}$. A set K is **compact** in X if every open covering of K includes a finite subcovering, i.e., for every family $\mathcal{G}$ of open sets,

$$K \subseteq \bigcup \mathcal{G} \Longrightarrow (\exists G_0, \dots, G_n \in \mathcal{G})[K \subseteq \bigcup_{n \leq i} G_i].$$

*$\mathbf{x10.21.}$ Prove that a pointset is compact by Definition $\mathbf{10.18}$ if and only if it is compact by Definition $\mathbf{10.43}$. HINT. You will need König's Lemma $\mathbf{9.7}$.

$\mathbf{x10.22.}$ Prove that for any two topological spaces X, Y, any continuous function $f : X \rightarrow Y$ and any compact set $K \subseteq X$, the image $f[K]$ is compact in Y.

REPLACEMENT AND OTHER AXIOMS

We have just about reached one of the goals we set in Chapter **4**, which was to prove all the "naive" results of Chapter **2** from the axioms of Zermelo. Only a couple of minor points remain, but they are significant: they will reveal that Zermelo's axioms are not sufficient and must be supplemented by stronger principles of set construction. Here we will formulate and add to the axiomatic theory **ZDC** the AXIOM OF REPLACEMENT discovered in the early 1920's, a principle of set construction no less plausible than any of the constructive axioms $(\mathbf{I}) - (\mathbf{VI})$ but powerful in its consequences. We will also introduce and discuss some additional principles which are often included in axiomatizations of set theory. Using only a weak consequence of Replacement, we will construct the LEAST ZERMELO UNIVERSE $\mathcal{Z}$, a remarkably simple set which contains the natural numbers, Baire space, the real numbers and all the significant objects of study of classical mathematics. Everything we have proved so far can be interpreted as if $\mathcal{Z}$ comprised the entire universe of mathematical objects, yet $\mathcal{Z}$ is just a set—and a fairly small, easy to comprehend set, at that! Our main purpose in this chapter is to understand the Axiom of Replacement by investigating its simplest and most direct consequences. The real power of this remarkable proposition will become apparent in the next chapter.

According to (2) of **2.16**, if A is a countable set and for each $n \geq 2$,

$$A^n = \underbrace{A \times \cdots \times A}_{n \text{ times}},$$

then the union $\bigcup_{n=2}^{\infty} A^n$ is also countable. The obvious way to prove this from the axioms is to define first the sets A^n by the recursion

$$\begin{aligned} f(0) &= A \times A, \\ f(n+1) &= f(n) \times A, \end{aligned} \tag{11-1}$$

so that $f(n) = A^{n+2}$ and

$$\bigcup_{n=0}^{\infty} A^{n+2} = \bigcup f[\mathbb{N}]. \tag{11-2}$$

Cantor's basic **2.10** implies first (by induction) that each $f(n) = A^{n+2}$ is countable, and then that their union $\bigcup_{n=0}^{\infty} A^{n+2}$ must also be countable. Is there an error? Certainly not in the proof by induction, which is no different

than many others like it. There is a problem, however, with the recursive definition (11-1) which cannot be justified by the Recursion Theorem **5.6** as it stands. To apply **5.6** we need a set E, a function $h : E \to E$ on E and some $a \in E$, which then determine a unique $f : \mathbb{N} \to E$ satisfying

$$\begin{aligned} f(0) &= a, \\ f(n+1) &= h(f(n)). \end{aligned} \qquad (11\text{-}3)$$

In the case at hand there is no obvious E which contains A and all its products A^n, and instead of a function h, we have the operation

$$h(X) =_{\mathrm{df}} X \times A, \qquad (11\text{-}4)$$

which associates with each set X its product $X \times A$ with the given set A. To justify definition (11-1), we need a recursion theorem which validates recursive definitions of the form (11-3), for every object a and (unary) definite operation h. It looks quite innocuous, only a mild generalization of the Recursion Theorem—and it is just that—but in fact such a result cannot be established rigorously on the basis of the Zermelo axioms.

11.1. (VIII) Replacement Axiom. *For each set A and each unary definite operation H, the image*

$$H[A] =_{\mathrm{df}} \{H(x) \mid x \in A\}$$

of A by H is a set.

As a construction principle for sets, the Replacement Axiom is almost obvious, as plausible on intuitive grounds as the Separation Axiom. If we already understand A as a completed totality and H associates in a definite and unambiguous manner an object with each $x \in A$, then we can "construct" the image $H[A]$ by "replacing" each $x \in A$ by the corresponding $H(x)$.

11.2. Axiomatics. The axiomatic system **ZFDC** of **Zermelo-Fraenkel set theory** with **Dependent Choices** comprises the axioms of **ZDC** and the Replacement Axiom (**VIII**), symbolically

$$\textbf{ZFDC} = \textbf{ZDC} + \text{Replacement} = (\textbf{I}) - (\textbf{VIII}).$$

From now on we will use all the axioms of **ZFDC** without explicit mention and we will continue to annotate by the mark (**AC**) the results whose proof requires the full Axiom of Choice.[29]

Mostly we have used simple definite operations up until now, those directly supplied by the axioms like $\mathcal{P}(A)$ and $\bigcup \mathscr{E}$ and explicit combinations of them, e.g., the Kuratowski pair $(x, y) =_{\mathrm{df}} \{\{x\}, \{x, y\}\}$. Once we assume the Axiom of Replacement, however, definite operations come into center stage and we will need to deal with some which are not so simply defined. We describe in the next, trivial Proposition the basic method of definition we will use, primarily to point attention to it.

---

[29] The Axiom of Replacement was introduced independently by Thoralf Skolem and Abraham Fraenkel in the early 1920s.

11.3. Proposition. *Suppose C and P are definite conditions of n and $n + 1$ arguments, respectively, assume that*

$$(\forall \vec{x})[C(\vec{x}) \implies (\exists! w) P(\vec{x}, w)], \tag{11-5}$$

and let

$$F(\vec{x}) =_{df} \begin{cases} \text{the unique } w \text{ such that } P(\vec{x}, w), \text{ if } C(\vec{x}), \\ \emptyset, \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{otherwise.} \end{cases} \tag{11-6}$$

The n-ary operation F is definite.

In practice, we will appeal to this observation by setting

$$F(\vec{x}) =_{df} \text{ the unique } w \text{ such that } P(\vec{x}, w) \quad (C(\vec{x})), \tag{11-7}$$

after we verify (11-5), without specifying the irrelevant value of F outside the domain we care about. The Axiom of Replacement often comes into the proof of (11-5).

11.4. Exercise. *For each unary definite operation F, the operation*

$$G(X) =_{df} F[X] = \{F(x) \mid x \in X\} \quad (\text{Set}(X))$$

is also definite.

The next fundamental consequence of the Replacement Axiom generalizes the Transfinite Recursion Theorem in two ways: by allowing a definite operation instead of just a function in the statement, and by replacing the given well ordered set by an arbitrary grounded graph. The second generalization does not require the Replacement Axiom, Problem **x8.11**.

11.5. Grounded Recursion Theorem. *For each grounded graph G with edge relation $\rightarrow$ and each binary definite operation H, there exists exactly one function $f : G \rightarrow f[G]$ which satisfies the identity*

$$f(x) = H(f \restriction \{y \in G \mid x \rightarrow y\}, x).$$

PROOF. As in the proof of **7.24**, we first show a lemma which gives us a set of approximations of the required function. Instead of the initial segments of G (which do not make much sense for an arbitrary graph), these approximations are defined here on downward closed subsets of G. Recall the definition of the transitive closure of a graph $\Rightarrow_G$ given in **6.34**; we will skip all the subscripts in what follows, since only the single graph G is involved in the argument, and we will also use the inverse arrows,

$$u \leftarrow t \iff_{df} t \rightarrow u \iff u \text{ is immediately below } t, \tag{11-8}$$
$$x \Leftarrow t \iff_{df} t \Rightarrow x \iff x \text{ is (on some path) below } t. \tag{11-9}$$

Lemma. *For each node $t \in G$, there exists exactly one function σ with domain the set $\{x \in G \mid x \Leftarrow t\}$ which satisfies the identity*

$$\sigma(x) = H(\sigma \restriction \{y \in G \mid y \leftarrow x\}, x) \quad (x \Leftarrow t). \tag{11-10}$$

Proof. Suppose, towards a contradiction, that t is a minimal node of G where the Lemma fails. Thus, for each $u \leftarrow t$, there is exactly one function σ_u such that

$$\sigma_u(x) = H(\sigma_u \restriction \{y \in G \mid y \leftarrow x\}, x) \quad (x \Leftarrow u). \qquad (11\text{-}11)$$

First we notice that

$$[x \Leftarrow u \leftarrow t \,\&\, x \Leftarrow v \leftarrow t] \Longrightarrow \sigma_u(x) = \sigma_v(x); \qquad (11\text{-}12)$$

because if x were minimal in G where (11-12) failed, then

$$\begin{aligned}
\sigma_u(x) &= H(\sigma_u \restriction \{y \in G \mid y \leftarrow x\}, x) && \text{by (11-11)},\\
&= H(\sigma_v \restriction \{y \in G \mid y \leftarrow x\}, x) && \text{by the choice of } x,\\
&= \sigma_v(x) && \text{by (11-11) for } \sigma_v.
\end{aligned}$$

The operation $u \mapsto \sigma_u$ which assigns this σ_u to each $u \leftarrow t$ is definite, so by the Axiom of Replacement its image is a set and we can set

$$\sigma_1 =_{\mathrm{df}} \bigcup \{\sigma_u \mid u \leftarrow t\};$$

this σ_1 is a function by (11-12), and by the definition,

$$\sigma_1(x) \downarrow \iff (\exists u)[x \Leftarrow u \leftarrow t].$$

By another application of the Replacement Axiom,

$$\sigma_2 =_{\mathrm{df}} \{(v, H(\sigma_v \restriction \{x \mid x \leftarrow v\}), x) \mid v \leftarrow t \,\&\, \neg(\exists u)[v \Leftarrow u \leftarrow t]\}$$

is also a set, and by its definition it is a function with domain disjoint from that of σ_1. Thus

$$\sigma =_{\mathrm{df}} \sigma_1 \cup \sigma_2$$

is a function, and

$$\begin{aligned}
\sigma(x) \downarrow &\iff (\exists u)[t \to u \,\&\, u \Rightarrow x] \vee t \to u\\
&\iff t \Rightarrow x \quad (\text{by } \mathbf{6.35}).
\end{aligned}$$

Moreover, σ satisfies (11-10) because σ_1 and σ_2 do. Finally, the same argument by which we proved (11-12) shows that no more than one σ with domain $\{x \in G \mid x \Leftarrow t\}$ can satisfy (11-10), and that completes the proof of the Lemma. $\dashv$ (Lemma)

To prove the Theorem, we apply the Lemma as in **7.24** to "the successor graph"

$$\begin{aligned}
\mathrm{Succ}(G) \;\; &=_{\mathrm{df}} \;\; G \cup \{t^*\},\\
x \to_{\mathrm{Succ}(G)} y \;\; &\Longleftrightarrow_{\mathrm{df}} \;\; x \to y \vee [x = t^* \,\&\, y \in G],
\end{aligned}$$

which has just one more node than G, at the top. $\dashv$

11.6. Corollary. (1) *For each well ordered set U and each binary definite operation H, there exists exactly one function $f : U \to f[U]$ which satisfies the identity*

$$f(x) = H(f \restriction \mathbf{seg}(x), x) \quad (x \in U). \qquad (11\text{-}13)$$

(2) *For each object a and each definite unary operation F, there exists a unique sequence $(n \mapsto a_n)$ which satisfies the identities*

$$a_0 = a, \quad a_{n+1} = F(a_n) \quad (n \in \mathbb{N}). \tag{11-14}$$

We call $(n \mapsto a_n)$ the **orbit** of a under F.

PROOF. For (1) we apply **11.5** to the graph $(\mathrm{Field}(U), >_U)$, and for (2) to the graph $(\mathbb{N}, \to)$, where

$$n \to m \iff_{\mathrm{df}} n = m + 1. \qquad \dashv$$

11.7. Exercise. *Which definite operation H do we use to prove (2) of the Corollary?*

The orbit of a set A by the unionset operation reveals the hidden structure of A under the membership relation by exposing the members of A, the members of the members of A, the members of those, etc. ad infinitum.

11.8. Definition. A class or set M is **transitive** if $\bigcup M \subseteq M$, equivalently

$$(\forall x \in M)(\forall t \in x)[t \in M],$$

or just $x \in M \Longrightarrow x \subseteq M$.

11.9. Exercise. *The sets $\emptyset$, $\{\emptyset, \{\emptyset\}\}$, $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$, the set*

$$\mathbb{N}_0 = \{\emptyset, \ \{\emptyset\}, \ \{\{\emptyset\}\}, \ \dots \} \tag{11-15}$$

postulated by the Axiom of Infinity and every class whose members are all atoms are transitive.

11.10. Transitive Closure Theorem. *Every set A is a member of some transitive set M, in fact, there is a least (under $\subseteq$) transitive set $M = \mathrm{TC}(A)$ such that $A \in \mathrm{TC}(A)$. We call $\mathrm{TC}(A)$ the* **transitive closure** *of A.*

PROOF. By (2) of **11.6**, there is a unique sequence $n \mapsto \mathrm{TC}_n(A)$ which satisfies the identities

$$\begin{aligned} \mathrm{TC}_0(A) &= \{A\}, \\ \mathrm{TC}_{n+1}(A) &= \bigcup \mathrm{TC}_n(A), \end{aligned} \tag{11-16}$$

and we set

$$\mathrm{TC}(A) =_{\mathrm{df}} \bigcup\nolimits_n \mathrm{TC}_n(A). \tag{11-17}$$

Clearly $A \in \mathrm{TC}(A)$ and $\mathrm{TC}(A)$ is transitive, because

$$u \in \mathrm{TC}_n(A) \Longrightarrow u \subseteq \bigcup \mathrm{TC}_n(A) = \mathrm{TC}_{n+1}(A).$$

If M is transitive and $A \in M$, then $\mathrm{TC}_0(A) = \{A\} \subseteq M$, and by induction

$$\mathrm{TC}_n(A) \subseteq M \Longrightarrow \mathrm{TC}_{n+1}(A) = \bigcup \mathrm{TC}_n(A) \subseteq \bigcup M \subseteq M,$$

so that in the end $\mathrm{TC}(A) = \bigcup\nolimits_n \mathrm{TC}_n(A) \subseteq M$. $\qquad \dashv$

11.11. Exercise. *If A is transitive, then $\mathrm{TC}(A) = A \cup \{A\}$.*

To understand better the remark about "revealing the hidden $\in$-structure" of A, consider the following natural concepts.

11.12. Definition. A set A is *hereditarily free of atoms* or **pure** if it belongs to some transitive set which contains no atoms; equivalently, if $TC(A)$ contains no atoms. A set A is **hereditarily finite** if it belongs to some transitive, finite set; equivalently, if $TC(A)$ is finite. A set A is **hereditarily countable** if it belongs to some transitive, countable set; equivalently, if $TC(A)$ is countable.

The point of the definitions is that $\{\{a\}\}$ is a set but not a pure set if a is an atom, because we need a to construct it; $\{\mathbb{N}\}$ is finite but not hereditarily finite because we need all the natural numbers to construct it; $\{\mathcal{N}\}$ is countable but not hereditarily countable because we need to "collect into a whole" an uncountable collection of objects in $\mathcal{N}$ before we can construct the singleton $\{\mathcal{N}\}$ by one final, trivial act of collection. Put another way, $\{\mathcal{N}\}$ is not hereditarily countable because "its concept involves" an uncountable infinity of objects, the members of its sole member $\mathcal{N}$.

11.13. Exercise. *The Principle of Purity* **3.25** *is equivalent to the assertion that every set is pure.*

11.14. Exercise. *A transitive set is hereditarily finite if it is finite, and hereditarily countable if it is countable.*

Next we consider the closure of a set under both the unionset and powerset operations.

11.15. Theorem (Basic Closure Lemma). *For each set I and each natural number n, let $M_n = M_n(I)$ be the set defined by the recursion*

$$M_0 = I, \quad M_{n+1} = M_n \cup \bigcup M_n \cup \mathcal{P}(M_n). \tag{11-18}$$

The **basic closure** *of I is the union*

$$M = M(I) =_{\text{df}} \bigcup\nolimits_{n=0}^{\infty} M_n(I), \tag{11-19}$$

and it has the following properties.

(1) M is a transitive set which contains $\emptyset$ and I, it is closed under the pairing $\{x, y\}$, unionset $\bigcup \mathscr{E}$ and powerset $\mathcal{P}(A)$ operations and it contains every subset of each of its elements.

(2) M is the least (under $\subseteq$) transitive set which contains I and is closed under $\{x, y\}$, $\bigcup \mathscr{E}$ and $\mathcal{P}(X)$.

(3) If I is pure and transitive, then each M_n is a pure, transitive set and satisfies

$$M_{n+1} = \mathcal{P}(M_n). \tag{11-20}$$

As a consequence, M is a pure, transitive set.

PROOF. (1) By the definition, $\emptyset, I \in M_1 \subseteq M$. If $x, y \in M$, then from the obvious $M_n \subseteq M_{n+1}$, there exists some m such that $\{x, y\} \subseteq M_m$, so $\{x, y\} \in M_{m+1}$. The key inclusion for the remaining claims is

$$x \in M_n \Longrightarrow x \subseteq \bigcup M_n \subseteq M_{n+1} \subseteq M,$$

which implies immediately that M is transitive. It also implies

$$x \in M_n \Longrightarrow \bigcup x \subseteq \bigcup M_{n+1} \subseteq M_{n+2},$$
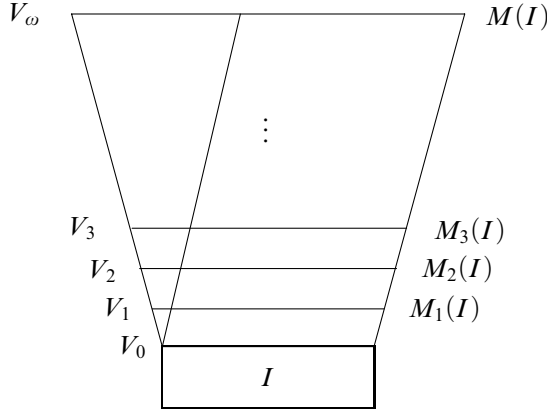
FIGURE **11.1**. Logarithmic[30] renditions of $M(I)$ and V_ω.

so $x \in M_n \Longrightarrow \bigcup x \in M_{n+3} \subseteq M$ and M is closed under $\bigcup x$. The same argument shows that M is closed under $\mathcal{P}(X)$ and the last assertion follows by this closure and transitivity.

(2) If M' is closed under $\{x, y\}$ and $\bigcup \mathcal{E}$, then it is also closed under $A \cup B = \bigcup \{A, B\}$; and if M' is also closed under $\mathcal{P}(X)$, then a simple induction shows that $M_n \in M'$ for each n, and so $M \subseteq M'$ by the transitivity of M'.

(3) If I is transitive with no atoms, then every M_n is transitive and has no atoms by a trivial induction on n. This implies that M is a transitive set with no atoms and hence pure, but also that $M_n \cup \bigcup M_n \subseteq \mathcal{P}(M_n)$, so that $M_{n+1} = \mathcal{P}(M_n)$.   ⊣

11.16. Exercise. *True or false*: *for every transitive set* X, $X \subseteq \mathcal{P}(X)$.

11.17. Exercise. *If* $I \subseteq J$, *then for each* n, $M_n(I) \subseteq M_n(J)$, *and hence*

$$I \subseteq J \Longrightarrow M(I) \subseteq M(J).$$

11.18. The grounded, pure, hereditarily finite sets. The least basic closure is that of the empty set, $M(\emptyset) \subseteq M(I)$, for every I. In the classical notation (which we will explain in the next chapter), $M_n(\emptyset) = V_n$, so that each V_n and their union are determined by the identities

$$V_0 = \emptyset, \quad V_{n+1} = \mathcal{P}(V_n), \quad V_\omega =_{\mathrm{df}} \bigcup_{n=0}^\infty V_n = M(\emptyset). \qquad (11\text{-}21)$$

For example, $\emptyset \in V_1$, $\{\emptyset\} \in V_2$ and $\{\{\emptyset\}, \{\{\emptyset\}\}\} \in V_4$! These sets are pictured on the left in Figure **11.1**. The set V_ω is grounded, pure and transitive, each V_n is finite by an easy induction, so every set in V_ω is grounded, pure and hereditarily finite, and V_ω itself is countable. These are the sets which can

---

[30]Picturing universes of sets by cones like this is traditional but misleading; the successive powersets grow hyperexponentially in size, so it would be more accurate to draw a cone with curved, hyperexponential sides.

be constructed "from nothing" (literally, the empty set) by iterating any finite number of times the operation of collecting into a whole (putting between braces) some of the objects already constructed.

The closure properties of $M(I)$ itemized in (1) of **11.15** are precisely those required of the universe $\mathcal{W}$ by axioms $(\mathbf{II}) - (\mathbf{V})$, as we discussed them in **3.26**, and if the set $\mathbb{N}_0$ of (11-15) demanded by the Axiom of Infinity $(\mathbf{VI})$ is a subset of I, we also have $\mathbb{N}_0 \in M(I)$. Notice also that since $M(I)$ is transitive, for $A, B \in M(I)$,

$$A \neq B \Longrightarrow (\exists t \in M(I))[t \in (A \setminus B) \cup (B \setminus A)], \qquad (11\text{-}22)$$

which says of $M(I)$ what the Axiom of Extensionality demands of $\mathcal{W}$ by (3-12). This suggests that if we take "object" to mean "member of $M(I)$", for any $I \supseteq \mathbb{N}_0$, then we can reinterpret every proof from the axioms $(\mathbf{I}) - (\mathbf{VI})$ as an argument about the members of $M(I)$ instead of all objects, in the end proving a theorem about $M(I)$ instead of $\mathcal{W}$. It is an important idea, worth abstraction and a name.

11.19. Definition. A transitive class M is a **Zermelo universe** if it is closed under the pairing $\{x, y\}$, unionset $\bigcup \mathcal{E}$ and powerset $\mathcal{P}(X)$ operations, and contains the set $\mathbb{N}_0$ defined by (11-15). The least Zermelo universe is $\mathcal{Z} = M(\mathbb{N}_0)$, determined by the identities

$$\mathcal{Z}_0 = \mathbb{N}_0, \quad \mathcal{Z}_{n+1} = \mathcal{P}(\mathcal{Z}_n), \quad \mathcal{Z} = \bigcup_{n=0}^{\infty} \mathcal{Z}_n. \qquad (11\text{-}23)$$

11.20. Exercise. *The class $\mathcal{W}$ of all objects is a Zermelo universe. Every Zermelo universe contains the empty set as well as every subset of each of its members.*

A Zermelo universe M is a *model* of the axioms $(\mathbf{I}) - (\mathbf{VI})$, and a very special model at that, since it interprets standardly the basic relations of membership and sethood—it only restricts the domain of objects in which we interpret propositions. The claim that logical consequences of $(\mathbf{I}) - (\mathbf{VI})$ are true in every Zermelo universe is called a *metatheorem*, a theorem about theorems. To make general results of this type completely precise and prove them rigorously requires concepts from *Mathematical Logic*. In specific instances, however, lemma by lemma and proposition by proposition, it is quite simple to see what the specific consequence of the axioms means for an arbitrary domain of objects and to verify it in every Zermelo universe: this is because, in fact, we have been using the axioms as closure properties of the universe, about which we have assumed nothing more but that it satisfies them.

11.21. Proposition. *Every Zermelo universe M is closed under the Kuratowski pair operation (x, y) defined in (4-1), as well as the Cartesian product $A \times B$, function space $(A \to B)$, and partial function space $(A \rightharpoonup B)$ operations, provided these are defined using the Kuratowski pair. In addition, if $A \in M$ and $\sim$ is an equivalence relation on A, then $\sim$ and the quotient $[\![A/\sim]\!]$ defined in* **4.12** *are also in M.*

PROOF. The Kuratowski pair $(x, y) = \{\{x\}, \{x, y\}\}$ of any two members of M is obtained by taking unordered pairs twice, so it is certainly in M. If $A, B \in M$, then $A \cup B = \bigcup\{A, B\}$, and by the proof of **4.2**,

$$A \times B \subseteq \mathcal{P}(\mathcal{P}(A \cup B)) \in M,$$

so $A \times B \in M$. The remaining claims are proved similarly. $\dashv$

11.22. Exercise. *Every Zermelo universe M contains a Peano system as defined in **5.1**.*

11.23. Proposition. (1) *The Axiom of Dependent Choices is true in every Zermelo universe M, in the following sense: if $a \in A \in M$, $P \subseteq A \times A$, $P \in M$ and $\mathbb{N} \in M$ is a system of natural numbers in M, then*

$$a \in A \,\&\, (\forall x \in A)(\exists y \in A)P(x, y)$$
$$\Longrightarrow (\exists f : \mathbb{N} \to A)[f \in M \,\&\, f(0) = a \,\&\, (\forall n \in \mathbb{N})P(f(n), f(n+1))].$$

(2) (**AC**) *The Axiom of Choice is true in every Zermelo universe M, in the following sense, following **8.4**: for every family $\mathscr{E} \in M$ of non-empty and pairwise disjoint sets, there exists some set $S \in M$ which is a choice set for $\mathscr{E}$, i.e.,*

$$S \subseteq \bigcup \mathscr{E}, \quad (\forall X \in \mathscr{E})(\exists u)[S \cap X = \{u\}].$$

PROOF. (1) The hypothesis of the implication to be proved implies by **DC** that there exists some function $f : \mathbb{N} \to A$ such that $f(0) = a$ and for every $n \in \mathbb{N}$, $P(f(n), f(n+1))$. Since $(\mathbb{N} \to A) \in M$, we also have $f \in M$ by transitivity.

Part (2) is proved similarly. $\dashv$

Although we chose specific versions of the choice principles to simplify these arguments, their numerous equivalents are also true in every Zermelo universe M. This can be verified directly, or by observing that the equivalence proofs we have given can be "carried out within M".

11.24. Exercise. (**AC**) *If M is any Zermelo universe, $A, B \in M$, and P is any binary definite condition, then*

$$(\forall x \in A)(\exists y \in B)P(x, y)$$
$$\Longrightarrow (\exists f \in M)[f : A \to B \,\&\, (\forall x \in A)P(x, f(x))]. \quad (11\text{-}24)$$

11.25. The least Zermelo universe $\mathcal{Z}$. Let us concentrate on the least Zermelo universe $\mathcal{Z}$, to focus the argument. It is a pure, transitive set, constructed by starting with the simple set $\mathbb{N}_0$ and iterating the powerset operation infinitely many times, much as we construct the natural numbers starting with 0 and iterating infinitely many times the successor operation. We can think of the sets in $\mathcal{Z}$ as precisely those objects whose existence is guaranteed by the axioms (**I**) – (**VI**). The natural interpretations of **DC** and **AC** are also true in $\mathcal{Z}$, the latter under the assumption that **AC** is true in $\mathcal{W}$. Using the closure properties of Zermelo universes already established and looking back at Chapters **5**, **6** and **10** and ahead at Appendix **A**, we can verify that $\mathcal{Z}$ contains not only the

specific system of natural numbers $\mathbb{N}$ we constructed in Chapter **5**, but also the Baire space $\mathcal{N}$ defined from this $\mathbb{N}$ and the specific systems of rational and real numbers constructed in Appendix **A**. By the uniqueness results, any one of these systems is as good as any other, so we can say that $\mathcal{Z}$ contains the integers, Baire space, the rationals and the reals.

Combining these remarks with some knowledge of classical mathematics, it is not hard to give a convincing argument that *all the objects studied in classical algebra, analysis, functional analysis, topology, probability, differential equations, etc. can be found (to within isomorphism) in $\mathcal{Z}$.* Many fundamental objects of abstract set theory are also in $\mathcal{Z}$, all we have constructed before this chapter in developing the theory of inductive posets, well ordered sets, etc. In slogan form: **we can develop classical mathematics and all the set theory needed for it as if all mathematical objects were members of $\mathcal{Z}$.**

The same can be said of every Zermelo universe, of course, but the concrete, simple definition of $\mathcal{Z}$ makes it possible to analyze its structure and investigate the special properties of its members. For example, no set which is a member of itself belongs to $\mathcal{Z}$: because no $X \in \mathbb{N}_0$ satisfies $X \in X$ (easily), and if n were least such that some $X \in X \in \mathcal{Z}_{n+1}$, then $X \in X \subseteq \mathcal{Z}_n$ by the definition, so $X \in \mathcal{Z}_n$, contradicting the choice of n. This looks good, we had some trouble with sets which belong to themselves. Actually, the iterative construction of $\mathcal{Z}$ ensures a much stronger regularity property for its members, discovered by von Neumann.

11.26. Definition. An object x is **ill founded** if it is the beginning of a descending $\in$-chain, i.e., if there exists a function $f : \mathbb{N} \to E$ such that

$$x = f(0) \ni f(1) \ni f(2) \ni \cdots .$$

Objects which are not ill founded are **well founded** or **grounded**. If $X \in X$, then $X \ni X \ni X \ni \cdots$, so X is ill founded. Problem **x11.14** gives a simple characterization of ill founded sets directly in terms of the $\in$ relation, which suggests that ill foundedness is a generalization of self-membership.

11.27. Exercise. *Atoms are grounded, as is $\emptyset$ and $\mathbb{N}_0$. A set is grounded if and only if all its members are grounded, if and only if its powerset is grounded. The class of all grounded sets is transitive.*

11.28. Proposition. *If I is grounded, then so is its basic closure $M(I)$. In particular, the least Zermelo universe $\mathcal{Z}$ and all its members are grounded.*

Proof. Assume that I is grounded, let (towards a contradiction) n be least such that M_n is ill founded and suppose that $M_n \ni x_1 \ni \cdots$ is a descending $\in$-chain. By hypothesis $n > 0$. Since $x_1 \in M_{n-1}$ and $x_1 \in y \in M_{n-1}$ contradict the choice of n, we must have $x_1 \subseteq M_{n-1}$, so $x_2 \in M_{n-1}$ and the descending $\in$-chain $M_{n-1} \ni x_2 \ni \cdots$ contradicts again the choice of n. It follows that M is also grounded, since any descending chain $M \ni x_1 \ni \cdots$ would also witness that M_n is ill founded for whatever M_n contains x_1. The consequence about $\mathcal{Z}$ follows because $\mathbb{N}_0$ is grounded. $\dashv$
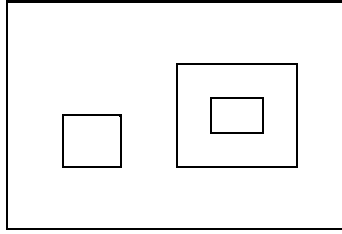
Figure **11.2**. $\{\emptyset, \{\emptyset\}\}$ as a disappointing gift.

In the old gag, the excited birthday boy opens up the huge box with his present, only to find inside it another box, and inside that another, and so on, until the last, tiny box is empty: his present is just the boxes. We can think of a pure, grounded set as a disappointing gift of this sort, except that each box may contain several boxes, not just one; no matter which one the birthday boy chooses to open up each time, eventually he finds the empty box, $\emptyset$. Most axiomatizations of set theory ban ill founded sets from the start by adopting the following principle proposed by von Neumann.

11.29. Principle of Foundation. *Every set is grounded.* This is also called the principle (or axiom) of *Regularity* in the literature.

It is worth putting down here an equivalent version of this Principle, which is somewhat opaque but useful.

11.30. Proposition. *The Principle of Foundation is true if and only if for every non-empty set X, there is some $m \in X$ such that*

$$m \cap X = \emptyset. \tag{11-25}$$

Proof. Assume first the Principle of Foundation and suppose, towards a contradiction, that $X \neq \emptyset$ but no $m \in X$ satisfies (11-25). This means that for some a,

$$a \in X \,\&\, (\forall m \in X)(\exists t \in X)[t \in m],$$

and then **DC** gives us an infinite descending $\in$-chain beginning with $X \ni a$ which contradicts the hypothesis. Conversely, if the Principle of Foundation fails and some infinite descending $\in$-chain starts with some set

$$X = f(0) \ni f(1) \ni f(2) \ni \cdots,$$

then the set $f[\mathbb{N}] = \{f(0), f(1), \dots\}$ is not empty and intersects each of its members, so none of them satisfies (11-25). ⊣

11.31. Zermelo-Fraenkel set theory (with choice), ZFC. By far the most widely used—the "official"—system of axioms for sets is the **Zermelo-Fraenkel Theory** (with choice), which accepts the Principles of Purity **3.25** and Foundation in addition to those of **ZFDC** and the Axiom of Choice, symbolically,

$$\textbf{ZFC} = \textbf{ZFDC} + \textbf{AC} + \text{Purity} + \text{Foundation.}$$
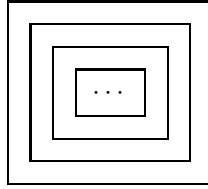
Figure **11.3**. Ω as the ultimately frustrating gift.

There are many and convincing arguments in favor of this industry standard, some of which we discuss immediately below. We will come back to the question in Chapter **12** and Appendix **B**, where we will also explain a few good foundational reasons for sticking with the weaker **ZFDC** in these Notes. As a practical matter, the principles of Purity and Foundation do not come up in the part of the subject we are covering, and the full **AC** is only needed rarely, so we can easily keep track of it.

11.32. Are all sets grounded? The most blatant exception to the Principle of Foundation would be a set which is its own singleton,

$$\Omega = \{\Omega\}. \tag{11-26}$$

We can think of Ω as the ultimately frustrating gift: each box has exactly one box inside it, identical with the one you just opened, and you can keep opening them forever without ever finding anything. How about sets Ω^1 and Ω^2 such that

$$\Omega^1 = \{\emptyset, \Omega^2\}, \quad \Omega^2 = \{\Omega^1\}? \tag{11-27}$$

These equations look unlikely, even bizarre, but it is not clear that our axioms rule them out. As a matter of fact they do not: we will construct in Appendix **B** some quite reasonable models of **ZFDC+AC** which contain some $\Omega = \{\Omega\}$ and many other sets with similar properties.

Recall the discussion about the *large* and the *small* heuristic views of the universe of objects $\mathcal{W}$ in **3.26**. The large view conceives $\mathcal{W}$ as the largest possible collection of objects which satisfies the axioms, while the small view takes it to comprise just the objects guaranteed by them.

On the large view, we have no more evidence in favor or against the Principle of Foundation now than we did back in Chapter **3**, except that we have proved all these things about sets without ever using it. But then again, we never saw a need for ill founded sets either.

On the small view, we have amassed some considerable evidence, at least for **ZDC+AC**, and it is all in favor of the Principle of Foundation: we now have a precise idea of what sets are "guaranteed" by the axioms of **ZDC+AC**, they are the members of $\mathcal{Z}$ and they are all pure and grounded. It may be argued that we did not build $\mathcal{Z}$ out of whole cloth, we worked within a "given" universe $\mathcal{W}$ of objects: in fact, we needed to assume that $\mathcal{W}$ satisfies

the Axiom of Replacement in addition to the axioms of **ZDC**+**AC**. This is certainly true, but so is the obvious response to it: aside from any rigorous axiomatization, the definition of $\mathcal{Z}$ and the proofs of its basic properties can be understood intuitively, naively, and they carry considerable force of persuasion. An informal description of $\mathcal{Z}$ would have made perfect sense in Chapter 3, as an intuitive conception of "restricted set" which justifies the axioms of **ZDC**+**AC** and the principles of Purity and Foundation. We have not been able to produce any such plausible, intuitive model of **ZDC** which contains ill founded sets from any hypotheses which do not beg the question.[31]

Could we construct simple models like $\mathcal{Z}$ for the theories **ZFDC** and **ZFC**? Let us first give them a name.

11.33. Definition. A **ZFDC-universe** is any Zermelo universe M which further satisfies the Axiom of Replacement in the following sense: for each $A \in M$ and each unary definite operation H,

$$(\forall x \in M)[H(x) \in M] \Longrightarrow H[A] = \{H(x) \mid x \in A\} \in M.$$

A **ZFC-universe** is any **ZFDC**-universe M which contains no atoms and such that every set $A \in M$ admits a well ordering in M.

The axioms of **ZFDC** assert precisely that the class $\mathcal{W}$ of all objects is a **ZFDC**-universe, and, correspondingly, the axioms of **ZFC** claim that $\mathcal{W}$ is a **ZFC**-universe.

11.34. Theorem. *The* **von Neumann class**

$$\mathcal{V} =_{\mathrm{df}} \{X \mid X \text{ is a pure, grounded set}\} \tag{11-28}$$

is a **ZFDC***-universe; and if* **AC** *holds, then* $\mathcal{V}$ *is a* **ZFC***-universe.*

PROOF. The fact that $\mathcal{V}$ is a Zermelo universe is quite trivial, most of it following from Exercise **11.27**. To verify that $\mathcal{V}$ also satisfies the Axiom of Replacement, notice that (whether $A \in \mathcal{V}$ or not), if H is unary, definite and such that for every $x \in A$, the value $H(x)$ is a pure, grounded set, then the image $H[A]$ has only pure and grounded members, so it is (easily) pure and grounded.

The second claim follows immediately, because if **AC** holds, then every $A \in \mathcal{V}$ admits a well ordering $\leq_A \in \mathcal{P}(A \times A)$ and $\leq_A \in \mathcal{V}$. $\dashv$

There is another, elegant and useful characterization of the pure grounded sets which follows easily from the Grounded Recursion Theorem **11.5**.

11.35. Definition. A **Mostowski surjection** or **decoration** of a graph G with edge relation $\rightarrow$ is a surjection $d : G \twoheadrightarrow d[G]$ which assigns a set to each node of G such that

$$d(x) = \{d(y) \mid y \leftarrow x\} \quad (x \in G). \tag{11-29}$$

---

[31]Models like $M(\mathbb{N}_0 \cup \Omega)$ beg the question, because they need some Ω with the requisite self-membership property to get started.
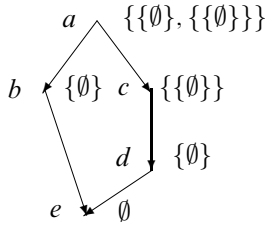
FIGURE **11.4**. Mostowski collapsing; $d_G[G] = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\{\emptyset\}, \{\{\emptyset\}\}\}\}$.

11.36. Theorem (Mostowski Collapsing Lemma). (1) *Every grounded graph G admits a unique decoration d_G, and its image $d_G[G]$ is a transitive, pure, grounded set.*

(2) *A set A is pure and grounded if and only if there exists a grounded graph G and a node $x \in G$, such that $A = d_G(x)$ for the unique decoration d_G of G.*

PROOF. (1) The existence of a unique decoration of a grounded G follows immediately from the Grounded Recursion Theorem **11.5** applied to G, with the definite operation

$$H(f) = \text{Image}(f) = \{f(x) \mid f(x)\downarrow\}.$$

The image $d_G[G]$ is transitive, since if $s \in t \in d_G[G]$, then $s \in t = d_G(y)$ for some $y \in G$, and then $s = d_G(x)$ for some $x \leftarrow y$, so $s \in d_G[G]$. Since each $d_G(x)$ is a set, by (11-29), $d_G[G]$ is a transitive set with no atoms and hence pure. Finally, $d_G[G]$ is grounded, because if $x_0 \ni x_1 \ni \cdots$ were an infinite, descending $\in$-chain in it and $s_0, s_1, \ldots$ were chosen so that $d_G(s_i) = x_i$, then $s_0 \to s_1 \to \cdots$ would be an infinite descending chain in the grounded graph G.

(2) If $A = d_G(x)$ with x a node in some grounded graph, then A is a member of a transitive, pure, grounded set by (1) and hence pure and grounded. For the converse, let $G = \text{TC}(A)$ be the transitive closure of A and define on it

$$x \to y \iff_{\text{df}} y \in x.$$

The graph G is grounded, because $G = \text{TC}(A)$ is a grounded set, Problem **x11.16**. In addition,

$$d_G(x) = x \quad (x \in G); \tag{11-30}$$

because if x were a G-minimal counterexample to (11-30), then

$$
\begin{aligned}
d_G(x) &= \{d_G(y) \mid y \ {}_G{\leftarrow}\ x\}, \\
&= \{y \mid y \leftarrow x\} \quad \text{by the choice of } x, \\
&= \{y \mid y \in x\} \quad \text{by the def. of } \to, \\
&= x \quad\quad\quad\quad\ \text{because } x \text{ is a set.}
\end{aligned}
$$

In particular, $A = d_G(A)$, which proves (2).   ⊣

The class $\mathcal{V}$ is not a set (Problem **x11.20**) and it is quite hard to find **ZFDC**-universes which are sets. See Problems **x12.43**, **x12.44** and **xB.12**.

11.37. Consistency and independence results. All the consistency and independence results we have discussed in **8.24**, **8.25**, **10.33**, **10.34**, **10.36** and **10.37** can be strengthened by adding the Axiom of Replacement to the relevant theories. This is as good a place as any to collect the most general versions of these fundamental results, which are outside the scope of these Notes.

(1) (Gödel, 1939) *The universe L of constructible sets is a model of* **ZFC**, *which further satisfies the Generalized Continuum Hypothesis* **GCH**. It follows that the Axiom of Choice **AC** cannot be refuted from the other axioms of **ZFC**, and that **GCH** cannot be refuted in **ZFC**.

(2) (Cohen, 1963) *None of the choice principles* $\mathbf{AC}_{\mathbb{N}}$, **DC** *and* **AC** *can be proved from a weaker one using the constructive axioms of Zermelo* (**I**) – (**VI**) *and the Axiom of Replacement* (**VIII**).

(3) (Cohen, 1963) *There is a model of* **ZFC** *in which the Continuum Hypothesis* **CH** *is false,* so **CH** *cannot be proved in* **ZFC**.

(4) (Solovay, 1970) *There is a model of* **ZFC** *in which every "definable", uncountable pointset has a perfect subset, and hence has cardinality* $\mathfrak{c}$. This means in particular, that we cannot define a specific pointset A and then prove in **ZFC** that it has cardinality intermediate between $\aleph_0$ and $\mathfrak{c}$.

(5) (Solovay, 1970) *There is a model of* **ZFDC** *in which every uncountable pointset has a perfect subset*, so *we cannot prove in* **ZFDC** *the existence of uncountable pointsets without perfect subsets*. Solovay's model also satisfies the Principles of Purity and Foundation.

Problems for Chapter 11

x11.1. Prove the Separation Axiom (**III**) from the remaining axioms in the group (**I**) – (**V**) and the Axiom of Replacement (**VIII**).

x11.2. For each set A, each unary, definite operation F and each binary, definite operation G, there exists a least under $\subseteq$ set $\overline{A}$ which contains A as a subset and is closed under F and G, i.e.,

$$A \subseteq \overline{A}, \quad x \in \overline{A} \Longrightarrow F(x) \in \overline{A}, \quad x, y \in \overline{A} \Longrightarrow G(x, y) \in \overline{A}.$$

(The same is true for any number of operations, of any number of arguments.)

x11.3. The Axiom of Replacement is constructively equivalent with the following proposition: for every set A and every unary definite operation F, there exists a set B which contains A and is closed under F, i.e.,

$$A \subseteq B \,\&\, F[B] \subseteq B.$$

x11.4. The Axiom of Replacement is constructively equivalent with the following proposition: for every set A and every unary definite operation F, the restriction

$$F \restriction A =_{\mathrm{df}} \{(x, F(x)) \mid x \in A\}$$

of F to A is a function, i.e., a set of pairs.

x11.5. If (x, y) is a definite, binary operation which satisfies the first property of ordered pairs (OP1) in **4.1**, then it also satisfies the second, (OP2). (This cannot be proved in **ZDC+AC**, see Problem **xB.4**.)

x11.6. If $|A|$ is a definite operation which satisfies the first condition on weak cardinal assignments (C1), then it automatically also satisfies the third one, (C3). (This cannot be proved in **ZDC+AC**, see Problem **xB.8**.)

x11.7. The definite condition of functionhood defined in (4-14) satisfies the equivalence

$$\mathrm{Function}(f) \iff \mathrm{Set}(f) \,\&\, (\forall w \in f)(\exists x, y)[w = (x, y)]$$
$$\&\, (\forall x, y, y')[[(x, y) \in f \,\&\, (x, y') \in f] \implies y = y'],$$

i.e., f is a function exactly when it is a single-valued set of pairs. (See also Problem **xB.9**.)

x11.8. There exists a sequence $(n \mapsto \aleph_n)$ which satisfies the identities

$$\aleph_0 = |\mathbb{N}|, \quad \aleph_{n+1} = \aleph_n^+.$$

We introduced these names for the first few infinite cardinals in (9-6), but this is not the same as proving the existence of *the sequence* $(n \mapsto \aleph_n)$. (See also Problem **xB.10**.)

x11.9. Extended recursion with parameters. For every unary definite operation G and every ternary definite operation H, there exists a unary definite operation F which satisfies the identities

$$F(0, y) = G(y),$$
$$F(n + 1, y) = H(F(n, y), n, y).$$

x11.10. If $\mathscr{E}$ is a non-empty family of transitive sets, then the union $\bigcup \mathscr{E}$ and the intersection $\bigcap \mathscr{E}$ are also transitive.

x11.11. For every class A there exists a least, transitive class $\overline{A}$ which contains A, that is, such that $A \subseteq \overline{A}$ and for every transitive class B, $A \subseteq B \implies \overline{A} \subseteq B$.

x11.12. The class of all pure sets is transitive, as are the classes of hereditarily finite and hereditarily countable sets.

x11.13. If $x_1 \in x_2 \in \cdots \in x_n = x_1$, then x_1 is ill founded.

x11.14. An object x is ill founded if and only if there exists some set A such that

$$x \in A \,\&\, (\forall s \in A)(\exists t \in A)[s \ni t].$$

x11.15. If Ω^1 and Ω^2 exist which satisfy (11-27), then they are distinct, hereditarily finite, pure sets.

x11.16. A set A is grounded if and only if its transitive closure $\mathrm{TC}(A)$ is grounded.

***x11.17.** A set is in V_ω if and only if it is pure, grounded and hereditarily finite. HINT. Show first that every finite, transitive, pure grounded set is in V_ω.

x11.18. For each transitive set I, let

$$J = \{x \in I \mid x \text{ is pure and grounded}\}$$

and prove that

$$M(J) = \{x \in M(I) \mid x \text{ is pure and grounded}\}.$$

x11.19. If a set $\Omega = \{\Omega\}$ exists as in (11-26), then

$$\{x \in M(\Omega) \mid x \text{ is grounded}\} = V_\omega.$$

x11.20. Prove that the class $\mathcal{V}$ of all pure, grounded sets is not a set.

11.38. Definition. A class K of atoms **supports** a set A if

$$x \in \mathrm{TC}(A) \,\&\, \mathrm{Atom}(x) \Longrightarrow x \in K,$$

and we let

$$\mathcal{W}[K] = \{x \mid x \text{ is supported by } K\}, \qquad\qquad (11\text{-}31)$$
$$\mathcal{V}[K] = \{x \mid x \text{ is grounded and supported by } K\}. \qquad (11\text{-}32)$$

x11.21. The class $\mathcal{W}[K]$ of sets supported by a class of atoms K is a **ZFDC**-universe.

x11.22. The class $\mathcal{V}[K]$ of grounded sets supported by a class K of atoms is a **ZFDC**-universe.

x11.23. Let $G = (\mathbb{N}, \to)$ where $m \to n \iff n < m$. Show that G is grounded and compute the image $d_G(m)$ of each m under the unique decoration of G.

x11.24. Give an example of an infinite, grounded graph G with $d_G[G] = \{\emptyset\}$.

***x11.25.** Let $G = (\mathbb{N} \setminus \{0\}, \to)$, where

$$m \to n \iff m \neq n \,\&\, n \text{ divides } m.$$

Show that G is grounded and compute the image $d_G(m)$ of each m under the unique decoration of G.

x11.26. An *extended decoration* of a graph G is any surjection $d : G \twoheadrightarrow d[G]$ such that for all $x \in G$,

$$d(x) = \begin{cases} x, & \text{if } x \text{ is an atom,} \\ \{d(y) \mid y \leftarrow x\}, & \text{if } x \text{ is a set.} \end{cases}$$

Prove that every grounded graph admits a unique extended decoration and that a (not necessarily pure) set A is grounded if and only if there exists a grounded graph G and some $x \in G$, such that $A = d(x)$.

*__x11.27. Grounded $\in$-recursion__. For each binary definite operation H, there exists a definite operation $F(t)$, such that for every grounded set x,

$$F(x) = H(F \restriction x, x),$$

where the function $F \restriction x = \{(t, F(t)) \mid t \in x\}$ is the restriction of the operation F to the set x.

This is a special case of the next, slightly more complex generalization of the Grounded Recursion Theorem __11.5__.

*__x11.28.__ Suppose $t < x$ is a binary definite condition which satisfies (1) for every x, the class $\{t \mid t < x\}$ is a set, and (2) there does not exist a sequence $(n \mapsto x_n)$ such that for all n, $x_{n+1} < x_n$. Prove that for every definite, binary operation H there exists another F, such that for every x,

$$F(x) = H(F \restriction \{t \mid t < x\}, x),$$

where the function $F \restriction \{t \mid t < x\} = \{(t, F(t)) \mid t < x\}$ is the restriction of F to the set $\{t \mid t < x\}$.

CHAPTER 12

ORDINAL NUMBERS

The Axiom of Replacement finds its most important applications in von Neumann's beautiful theory of ORDINAL NUMBERS, and in the construction of the CUMULATIVE HIERARCHY of pure, grounded sets. One can live without knowing the ordinals, to be sure, but not as well: they bring many gifts, among them true cardinal numbers which give substance to the "virtual" theory of equinumerosities with which we have been making do. The Cumulative Hierarchy extends the iteration of the power operation we have used to construct V_ω "as far as it will go" and presents the pure, grounded sets as the most compelling intuitive understanding of what sets really are. It is not so clear one can live without knowing that, not among set theorists, at any rate.

Cantor describes his conception of "ordinal types" just a few pages after the definition of cardinals quoted in **4.19**, and in a very similar vein.

> Every ordered set U has a definite 'ordinal type', . . . which we will denote by $\overline{U}$. By this we understand the general concept which results from U if we only abstract from the nature of the elements u, and retain the order or precedence among them. Thus the ordinal type $\overline{U}$ is itself an ordered set whose elements are units which have the same order of precedence amongst one another as the corresponding elements of U, from which they are derived by abstraction. . . . A simple consideration shows that two ordered sets have the same ordinal type if, and only if, they are similar, so that of the two formulas $U =_o V$, $\overline{U} = \overline{V}$, one is always a consequence of the other.

Cantor is speaking about arbitrary linearly ordered sets, but we will consider here only the problem of defining "ordinal types" for well ordered sets. He states explicitly the first key property

$$U =_o \overline{U} \tag{12-1}$$

of the ordinal assignment operation, and argues for

$$U =_o V \Longrightarrow \overline{U} = \overline{V}. \tag{12-2}$$

Cantor's implied "simple consideration" for (12-2) should also justify (for well ordered sets) the stronger implication

$$U \leq_o V \Longrightarrow \overline{U} \sqsubseteq \overline{V}; \tag{12-3}$$
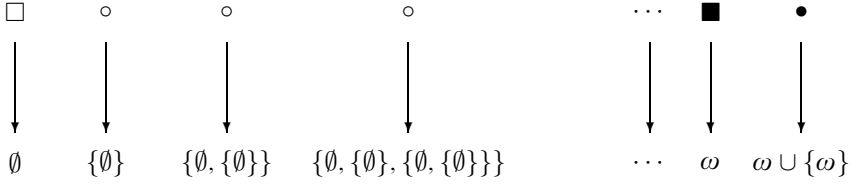
FIGURE **12.1**. The von Neumann map of $U : 0_U, 1_U, 2_U, \ldots, \omega_U, S_U(\omega_U)$.

because the position of a point x in a well ordered set depends only on the points preceding it, so the "unit" $\overline{x}$ abstracted from x and coding its place in U should depend only on the initial segment $\mathbf{seg}_U(x)$. Thus, the problem of representing Cantor's conception of ordinals in axiomatic set theory comes down to the following: *can we assign a well ordered set $\overline{U}$ to each well ordered set U, so that* (12-1) *and* (12-3) *hold?* Von Neumann's ingenious idea is to define $\overline{U}$ by replacing *recursively* each member of U by the set of its predecessors. The construction is captured exactly by the Mostowski Collapsing Lemma **11.36**.

12.1. Ordinal numbers. The **von Neumann map** of a well ordered set U is the unique decoration of the associated grounded graph $(\text{Field}(U), >_U)$, so that by (11-29),

$$\mathbf{v}_U(x) = \{\mathbf{v}_U(y) \mid y <_U x\}, \quad (x \in \text{Field}(U)). \tag{12-4}$$

We define the **ordinal number** of U to be the image

$$\text{ord}(U) =_{\text{df}} \mathbf{v}_U[\text{Field}(U)] \tag{12-5}$$

of its von Neumann map, and we set

$$\text{ON}(\alpha) \iff \alpha \in \text{ON} \iff_{\text{df}} (\exists \text{ well ordered } U)[\alpha = \text{ord}(U)]. \tag{12-6}$$

Suppose for example that

$$U : 0_U, 1_U, 2_U, \ldots, \omega_U, S_U(\omega_U)$$

is a well ordered set with least element 0_U, next 1_U, ..., first limit point ω_U, followed by the last (largest) point $S_U(\omega_U)$. We compute the values of its von Neumann surjection by repeated applications of (12-4) (skipping the subscript):

$$\begin{aligned}
\mathbf{v}(0_U) &= \{\mathbf{v}(x) \mid x < 0_U\} &&= \emptyset &&= 0, \\
\mathbf{v}(1_U) &= \{\mathbf{v}(x) \mid x < 1_U\} &&= \{\emptyset\} &&= 1, \\
\mathbf{v}(2_U) &= \{\mathbf{v}(x) \mid x < 2_U\} &&= \{\emptyset, \{\emptyset\}\} &&= 2, \\
\mathbf{v}(3_U) &= \{\mathbf{v}(x) \mid x < 3_U\} &&= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} &&= 3, \\
&\qquad \vdots \\
\mathbf{v}(\omega_U) &= \{\mathbf{v}(x) \mid x < \omega_U\} &&= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots\} = \omega \\
\mathbf{v}(S_U(\omega_U)) &= \{\mathbf{v}(x) \mid x < S_U(\omega_U)\} &&= \omega \cup \{\omega\}
\end{aligned}$$

$$\mathbf{v}[U] = \{0, 1, 2, \ldots, \omega, \omega \cup \{\omega\}\}.$$

These computations are special cases of the following, general facts:
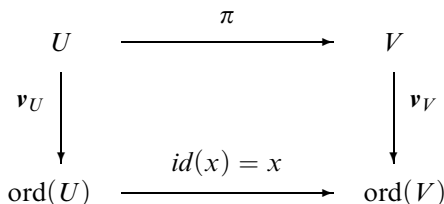
FIGURE **12.2**. The von Neumann map under initial similarities.

12.2. Exercise. *If 0_U is the least point in a well ordered set U, then $\mathbf{v}_U(0_U) = \emptyset$; and if $S(x)$ is the successor of x in U, then*

$$\mathbf{v}_U(S(x)) = \mathbf{v}_U(x) \cup \{\mathbf{v}_U(x)\}.$$

12.3. Exercise. *If x is a limit point in a well ordered set U, then $\emptyset \in \mathbf{v}_U(x)$ and*

$$\alpha \in \mathbf{v}_U(x) \Longrightarrow \alpha \cup \{\alpha\} \in \mathbf{v}_U(x).$$

12.4. Exercise. *If ω_U is the first limit point in a well ordered set U, then*

$$\omega = \mathbf{v}_U(\omega_U) = \bigcap\{X \mid \emptyset \in X \,\&\, (\forall \alpha \in X)[\alpha \cup \{\alpha\} \in X]\}, \qquad (12\text{-}7)$$

and, in particular, $\omega = \mathbf{v}_U(\omega_U)$ is independent of the particular well ordered set U used to compute it.

This exercise makes clear about ω_U what is evident about $0_U, 1_U, \ldots$ in Figure **12.1**; the value $\mathbf{v}_U(x)$ is independent of the particular element $x \in U$, and depends only on the place of x in U, whether it is the first element, the fifth, the first limit point or whatever. This is a general fact about the von Neumann map, which we can make precise as follows.

12.5. Lemma (First Ordinal Property). *If $\pi : U \rightarrowtail\!\!\!\rightarrow \pi[U] \sqsubseteq V$ is an initial similarity from U into another well ordered set V, then the diagram in Figure **12.2** commutes, i.e.,*

$$\mathbf{v}_V(\pi(x)) = \mathbf{v}_U(x) \quad (x \in U). \qquad (12\text{-}8)$$

PROOF. Towards a contradiction, let x be the least element of U such that $\mathbf{v}_V(\pi(x)) \neq \mathbf{v}_U(x)$, and compute:

$$
\begin{aligned}
\mathbf{v}_V(\pi(x)) &= \{\mathbf{v}_V(y) \mid y <_V \pi(x)\} \quad \text{by definition,} \\
&= \{\mathbf{v}_V(\pi(t)) \mid t <_U x\} \quad \text{because } \pi \text{ is initial,} \\
&= \{\mathbf{v}_U(t) \mid t <_U x\} \quad\ \text{by the choice of } x, \\
&= \mathbf{v}_U(x),
\end{aligned}
$$

contradicting the choice of x. The key step here is the second one, where we used the fact that an initial similarity "has no gaps" in its image, so that each $y <_V \pi(x)$ is $\pi(t)$ for some $t <_U x$. $\dashv$

12.6. Exercise. *For any two well ordered sets U, V,*

$$U \leq_o V \Longrightarrow \operatorname{ord}(U) \subseteq \operatorname{ord}(V);$$

and so, if $U =_o V$, then $\operatorname{ord}(U) = \operatorname{ord}(V)$.

12.7. Lemma (Second ordinal property). *For each well ordered set U and each $x \in U$,*

$$\boldsymbol{v}_U(x) = \mathrm{ord}(\mathbf{seg}_U(x)). \qquad (12\text{-}9)$$

As a consequence, each von Neumann value $\boldsymbol{v}_U(x)$ is an ordinal number, and conversely, each ordinal number α is the von Neumann value $\boldsymbol{v}_U(x)$ of some point in a well ordered set.

Put another way: each member of an ordinal is an ordinal and every ordinal is a member of an ordinal.

PROOF. If we apply Lemma **12.5** with $\mathbf{seg}_U(x)$ for U, U for V and the identity initial similarity $\pi : \mathbf{seg}_U(x) \rightarrowtail U$, we get:

$$\boldsymbol{v}_{\mathbf{seg}_U(x)}(y) = \boldsymbol{v}_{\mathbf{seg}_U(x)}(\pi(y)) = \boldsymbol{v}_U(y) \qquad (y <_U x),$$

so that, in the end

$$\boldsymbol{v}_U(x) = \{\boldsymbol{v}_U(y) \mid y <_U x\} = \{\boldsymbol{v}_{\mathbf{seg}_U(x)}(y) \mid y <_U x\} = \mathrm{ord}(\mathbf{seg}_U(x)).$$

For the second claim, let $V = \mathrm{Succ}(U)$ be the next well ordered set to U, with t added on top, as in **7.16**: now $U = \mathbf{seg}_V(t)$, and so $\mathrm{ord}(U) = \boldsymbol{v}_V(t)$. ⊣

Thus, we can think of ordinal numbers as standing either for lengths of well ordered sets, or for places of points in a well ordered set. The latter agrees more with the use of ordinals in ordinary language, where "first", "second", ... customarily describe the place of objects in a sequence.

12.8. Exercise (The finite von Neumann ordinals). *If $\leq_\mathbb{N}$ is the usual ordering of the set $\mathbb{N}$ of natural numbers, then*

$$\mathrm{ord}(\mathbb{N}, \leq_\mathbb{N}) = \omega,$$

as this is defined in (12-7). *Moreover, if we set*

$$S_\omega(n) = n \cup \{n\} \quad (n \in \omega),$$

then $(\omega, \emptyset, S_\omega)$ is a Peano system, and $\boldsymbol{v}_\mathbb{N} : \mathbb{N} \twoheadrightarrow \omega$ is the unique (by Theorem **5.4**) *isomorphism of the natural numbers with ω.*

It is usual in advanced set theory to take $(\omega, \emptyset, S)$ as *the* Peano system we fixed in **5.9**, i.e., to identify $\mathbb{N}$ with ω. This is sometimes convenient, but neither necessary, nor especially natural—it is hard to argue that $\{\emptyset, \{\emptyset\}\}$ is a better representation of the number 2 than Zermelo's $\{\{\emptyset\}\}$ which comes from the proof of Theorem **5.3**, or the third member of any Peano system.

Next comes the basic fact about ordinal numbers.

12.9. Lemma (Third ordinal property). *Each ordinal number α is well ordered by the relation*

$$u \leq_\alpha v \iff_{\mathrm{df}} u = v \lor u \in v \quad (u, v \in \alpha); \qquad (12\text{-}10)$$

and if $\alpha = \mathrm{ord}(U)$ for a well ordered set U, then the von Neumann map $\boldsymbol{v}_U : U \twoheadrightarrow \alpha$ is a similarity.

It follows that every well ordered set is similar with an ordinal number, and every well orderable set is equinumerous with an ordinal.

PROOF. Notice first that by Lemma **11.36**, each $\alpha = \nu_U[U]$ is a transitive, pure and grounded set, and (writing ν for ν_U),

$$x <_U y \Longrightarrow \nu(x) \in \nu(y). \tag{12-11}$$

Moreover,

$$x \neq y \Longrightarrow \nu(x) \neq \nu(y), \quad (x, y \in U);$$

because if $x \neq y$, then either $x <_U y$ or $y <_U$, so that by (12-11), either $\nu(x) \in \nu(y)$ or $\nu(y) \in \nu(x)$; and, in either case, we can't have $\nu(x) = \nu(y)$ since α is grounded. Thus $\nu : U \rightarrowtail\!\!\!\rightarrow \alpha$ is an injection, hence a bijection, and the relation $\leq_\alpha$ is the image of $\leq_U$ by ν, i.e.,

$$x \leq_U y \iff \nu(x) \leq_\alpha \nu(y) \quad (x, y \in U);$$

so $\leq_\alpha$ well orders α, and ν is a similarity.

The last claim follows because similarities are bijections. $\dashv$

This remarkable result says, in effect, that there exist sufficiently long $\in$-chains to mirror every wellordering, and it is a characteristic consequence of the Replacement Axiom. As we have been doing with structured sets throughout, by "the ordinal α" we will mean ambiguously the set α or the well ordered set $(\alpha, \leq_\alpha)$, so that, for example, Lemma **12.9** is expressed simply by

$$U =_o \mathrm{ord}(U).$$

12.10. Exercise. *For every ordinal number α, $\mathrm{ord}(\alpha) = \alpha$.*

12.11. Corollary (Characterization of ordinals). *A set α is an ordinal if and only if it is transitive, pure, grounded and $\in$-connected, i.e.,*

$$\mathrm{ON}(\alpha) \iff [x = y \vee x \in y \vee y \in x] \qquad (x, y \in \alpha).$$

PROOF. Every ordinal has these properties, by Lemmas **11.36** and **12.9**. For the converse, suppose A is transitive, pure, grounded and $\in$-connected, and set (as if A were an ordinal)

$$x \leq_A y \iff x = y \vee x \in y, \quad (x, y \in A).$$

This relation (easily) well orders A, so let $\nu_A : A \rightarrowtail\!\!\!\rightarrow \mathrm{ord}(A, \leq_A)$ be its von Neumann map. We now claim that ν_A is the identity map; if not, let x be $\leq_A$-least such that $\nu_A(x) \neq x$, and compute:

$$\begin{aligned}
\nu_A(x) &= \{\nu(y) \mid y <_A x\} \\
&= \{y \mid y <_A x\} && \text{(by the choice of } x) \\
&= \{y \mid y \in x\} && \text{(by the definition of } \leq_A) \\
&= x && \text{(because } x \text{ has no atoms),}
\end{aligned}$$

contradicting the choice of x. Thus $\mathrm{ord}(A, \leq_A) = \nu_A[A] = A$, and A is an ordinal. $\dashv$

   This characterization of ordinal numbers is especially simple when we assume that all sets are grounded and pure, as we do in **ZFC**: a *set then is an ordinal exactly when it is transitive and $\in$-connected.*

   The three basic ordinal properties also give a strong solution to Cantor's problem of defining ordinal types for well ordered sets, as we formulated it in (12-1) – (12-3).

12.12. Theorem. *The definite operation* $U \mapsto \mathrm{ord}(U)$ *on well ordered sets satisfies the following conditions*:

$$U =_o \mathrm{ord}(U), \qquad\qquad\qquad (12\text{-}12)$$
$$U \leq_o V \Longrightarrow \mathrm{ord}(U) \sqsubseteq \mathrm{ord}(V), \qquad\qquad (12\text{-}13)$$
$$\mathrm{ON}(\alpha) \Longrightarrow \alpha = \{\beta \in \mathrm{ON} \mid \beta <_o \alpha\}. \qquad (12\text{-}14)$$

   PROOF. The first property (12-12) is a restatement of Lemma **12.9**.

   To prove (12-13), suppose $\pi : U \rightarrowtail \pi[U] \sqsubseteq V$ is an initial similarity. By Lemma **12.7**, taking images,

$$\boldsymbol{v}_V[\pi[U]] = \boldsymbol{v}_U[U] = \mathrm{ord}(U);$$

and since $\boldsymbol{v}_V$ is a similarity of V with $\mathrm{ord}(V)$, it carries initial segments onto initial segments, so that

$$\mathrm{ord}(U) = \boldsymbol{v}_U[U] = \boldsymbol{v}_V[\pi[U]] \sqsubseteq \mathrm{ord}(V).$$

   Finally, for (12-14), if $\alpha = \mathrm{ord}(U)$, then:

$$\begin{aligned}
\alpha &= \{\boldsymbol{v}_U(y) \mid y \in U\} \\
&= \{\mathrm{ord}(\mathbf{seg}_U(y)) \mid y \in U\} \quad \text{(by Lemma \textbf{12.7})} \\
&= \{\beta \in \mathrm{ON} \mid \beta <_o \alpha\},
\end{aligned}$$

the last because the well ordered sets which are $<_o U$ are exactly those similar with the proper initial segments of U. ⊣

12.13. Exercise. *For any two well ordered sets* U, V,

$$U =_o V \iff \mathrm{ord}(U) = \mathrm{ord}(V),$$

and so for each well ordered set U, *there is exactly one ordinal number* α *such that* $U =_o \alpha$.

   Conditions (12-12) and (12-13) are precisely Cantor's (12-1) and (12-3). The key, last condition (12-14) is characteristic of the von Neumann ordinal assignment, Problem **x12.4**. This is an interesting result; we formulated it as a problem only because it makes for a good one, and we will not need to appeal to it. However, it is easy to get lost in proving scores of elementary properties of ordinals, some useful, others just challenging, and the proofs from the definition are a bit confusing: it is not entirely natural to think of the membership relation as an ordering. It is good practice, at least in the beginning, to prove properties of ordinals directly from their three, basic properties isolated above, which summarize their most basic features.

12.14. Lemma (Ordinal comparison). *For any two ordinals α, β,*

$$\alpha \leq_o \beta \iff \alpha = \beta \vee \alpha \in \beta \iff \alpha \sqsubseteq \beta \iff \alpha \subseteq \beta.$$

PROOF. We give a round-robin argument of the strict versions of the claimed equivalences.

(1) $\alpha <_o \beta \implies \alpha \in \beta$ follows immediately from (12-12) and (12-14), since the hypothesis means that $\alpha = \mathrm{ord}(U)$ and $\beta = \mathrm{ord}(V)$ with $U <_o V$.

(2) $\alpha \in \beta \implies \alpha \subsetneqq \beta$ and (3) $\alpha \subsetneqq \beta \implies \alpha \subsetneq \beta$ follow equally easily from (12-12) and (12-14) and we will skip them.

(4) $\alpha \subsetneq \beta \implies \alpha <_o \beta$. The hypothesis gives us an injection from α to β (the identity!), and so $\alpha \leq_o \beta$ by Corollary **7.32**; but $\alpha =_o \beta$ implies $\alpha = \beta$ by Exercise **12.6** which contradicts the assumed, proper inclusion $\alpha \subsetneq \beta$, and so $\alpha <_o \beta$. ⊣

It is traditional to use for the ordering on ordinals the simplest notation,

$$\alpha \leq \beta \iff_{\mathrm{df}} \alpha \leq_o \beta \quad (\alpha, \beta \in \mathrm{ON}), \tag{12-15}$$

keeping in mind its equivalent characterizations in Lemma **12.14**. We summarize its properties in one, now simple result.

12.15. Theorem (The ordering of ON). (1) *The class* ON *of ordinal numbers is well ordered by the condition $\alpha \leq \beta$, in the following precise sense*:

$$\alpha \leq \alpha, \quad \alpha \leq \beta \,\&\, \beta \leq \gamma \implies \alpha \leq \gamma, \quad \alpha \leq \beta \,\&\, \beta \leq \alpha \implies \alpha = \beta,$$
$$\alpha < \beta \vee \alpha = \beta \vee \beta < \alpha,$$

and for every definite condition P,

$$(\exists \alpha \in \mathrm{ON})P(\alpha) \implies (\exists \alpha \in \mathrm{ON})[P(\alpha) \,\&\, (\forall \beta < \alpha)\neg P(\beta)].$$

In particular, there is no infinite descending chain of ordinals,

$$\alpha_0 \geq \alpha_1 \geq \alpha_2 \geq \cdots \implies (\exists n)[\alpha_n = \alpha_{n+1}]. \tag{12-16}$$

When $P(\alpha)$ holds for some α, we set

$$(\mu\alpha \in \mathrm{ON})P(\alpha) = \min\{\alpha \in \mathrm{ON} \mid P(\alpha)\}. \tag{12-17}$$

(2) *For each ordinal number there is a next one,*

$$S(\alpha) =_{\mathrm{df}} (\mu\beta \in \mathrm{ON})[\alpha < \beta] = \alpha \cup \{\alpha\}. \tag{12-18}$$

(3) *Each set A of ordinal numbers has a least upper bound,*

$$\sup A =_{\mathrm{df}} (\mu\beta \in \mathrm{ON})(\forall \alpha \in A)[\alpha \leq \beta] = \bigcup A, \tag{12-19}$$

which is the maximum of A (if A has a maximum) and 0 if $A = \emptyset$.

PROOF is left for the Problems, **x12.1** – **x12.3**. ⊣

12.16. Exercise. *For each non-empty set of ordinals $\mathscr{E}$,*

$$(\mu\alpha \in \mathrm{ON})[\alpha \in \mathscr{E}] = \bigcap \mathscr{E}.$$

The **successor ordinals** are those of the form $S(\alpha)$ and the **limit ordinals** are those which are not successors or 0, so that $\alpha < \lambda \Longrightarrow S(\alpha) < \lambda$; these are also characterized by the property

$$\text{Limit}(\lambda) \iff \lambda \neq 0 \,\&\, \lambda = \sup\{\alpha \mid \alpha < \lambda\}. \qquad (12\text{-}20)$$

We can prove properties of ordinals by *transfinite induction* and define operations on them by *transfinite recursion*, as follows.

12.17. Theorem (Ordinal induction). *For every unary definite condition P,*

$$(\forall \alpha)[(\forall \xi < \alpha)P(\xi) \Longrightarrow P(\alpha)] \Longrightarrow (\forall \alpha)P(\alpha).$$

PROOF. Towards a contradiction, let α be least such that $\neg P(\alpha)$; now $P(\xi)$ holds for all $\xi < \alpha$, and so $P(\alpha)$ holds by the hypothesis, which we assumed it does not. ⊣

12.18. Theorem (Ordinal recursion). *For every binary definite operation H, there exists a unary definite operation F, which satisfies the identity*

$$F(\alpha) = H(F \restriction \alpha, \alpha) \quad (\alpha \in \text{ON}). \qquad (12\text{-}21)$$

Here $F \restriction \alpha$ is the function $\{(\xi, F(\xi)) \mid \xi \in \alpha\}$ obtained by restricting $F(\xi)$ to $\alpha = \{\xi \mid \xi < \alpha\}$.

Similarly, with parameters, given $H(w, \alpha, x)$, there exists $F(\alpha, x)$ such that for all α, x,

$$F(\alpha, x) = H(\{(\xi, F(\xi, x)) \mid \xi < \alpha\}, \alpha, x) \quad (\alpha \in \text{ON}).$$

PROOF of the simpler, parameter-free version.

For each β, by Corollary **11.6** on the well ordered set $(\beta, \leq_\beta)$, there exists exactly one function $f_\beta : \beta \twoheadrightarrow E_\beta$ which satisfies the identity

$$\begin{aligned} f_\beta(\alpha) &= H(f_\beta \restriction \{x \in \beta \mid x <_\beta \alpha\}, \alpha) \quad (\alpha < \beta), \\ &= H(f_\beta \restriction \alpha, \alpha), \end{aligned} \qquad (12\text{-}22)$$

using the fact that $<_\beta$ coincides with $\in$ and the members of β are ordinals. We claim that

$$\text{if } \alpha < \beta \text{ and } \alpha < \gamma, \text{ then } f_\beta(\alpha) = f_\gamma(\alpha);$$

if not, then there would exist a least α for which this fails for some β and γ, and then (12-22) yields a contradiction immediately. Thus, we can set

$$F(\alpha) = f_{S(\alpha)}(\alpha),$$

so $F(\alpha) = f_\beta(\alpha)$ for any $\beta > \alpha$ and (12-22) implies the required identity for the operation F.

The proof for the version with parameters is only notationally more complex. ⊣

Using this theorem, we can define arithmetical operations on ON and study their structure. We will leave most of this for the problems, but it is worth recording here the two most basic definitions, as examples of Theorem **12.18**, and in order to have some notation available to name specific ordinals.

12.19. Theorem (Ordinal addition and multiplication). *There exist binary, definite operations* $\alpha + \beta$ *and* $\alpha \cdot \beta$ *on the ordinals which satisfy the following identities*:

$$\alpha + 0 = \alpha,$$
$$\alpha + S(\beta) = S(\alpha + \beta),$$
$$\alpha + \lambda = \sup\{\alpha + \beta \mid \beta < \lambda\}, \text{ if } \mathrm{Limit}(\lambda). \tag{12-23}$$

$$\alpha \cdot 0 = 0,$$
$$\alpha \cdot S(\beta) = (\alpha \cdot \beta) + \alpha,$$
$$\alpha \cdot \lambda = \sup\{\alpha \cdot \beta \mid \beta < \lambda\}, \text{ if } \mathrm{Limit}(\lambda). \tag{12-24}$$

PROOF. We set $\alpha + \beta = F(\beta, \alpha)$, where $F(\beta, \alpha)$ is defined by the following recursion on $\beta \in \mathrm{ON}$, with α as parameter:

$$F(\beta, \alpha) = \begin{cases} \alpha, & \text{if } \beta = 0, \\ S(F(\gamma, \alpha)), & \text{if } \beta = S(\gamma), \text{ for some } \gamma, \\ \sup\{F(\xi, \alpha) \mid \xi < \beta\}, & \text{if } \mathrm{Limit}(\beta). \end{cases}$$

We leave for Problem **x12.6** the (similar) argument for multiplication. $\dashv$

We have already introduced the ordinal ω in (12-7), and proved in Exercise **12.8** that $\omega = \mathrm{ord}(\mathbb{N}, \leq_{\mathbb{N}})$. The ordinals following it immediately are obviously

$$\omega + 1 = S(\omega), \ \omega + 2 = S(\omega + 1), \ \omega + 3 = S(\omega + 2), \ldots$$

and right above these comes

$$\omega + \omega = \sup\{\omega + n \mid n \in \omega\} = \omega \cdot 2. \tag{12-25}$$

This is the second limit ordinal, the first one above ω. Each $\omega \cdot n$ can be obtained by adding ω to itself n times, directly from the definition. Next comes

$$\omega^2 = \sup\{\omega \cdot n \mid n < \omega\},$$

after a while $\omega^3 = \omega^2 \cdot \omega$, etc.

Many of the properties of ordinal addition and multiplication are most easily derived from the properties of these operations on (well ordered) sets, as we defined them in **7.37** and **7.38**, using the following three exercises.

12.20. Exercise. *For every ordinal* α,

$$\alpha + 1 = \mathrm{ord}(\mathrm{Succ}(\alpha)),$$

by the definition of the successor poset $\mathrm{Succ}(P)$ *in* **7.16**.

12.21. Exercise. *For all ordinals* α, β,

$$\alpha + \beta = \mathrm{ord}(\alpha +_o \beta),$$

by the definition of addition of posets in **7.37**. *It follows that addition of ordinals is associative but not commutative, and that*

$$\beta < \gamma \implies \alpha + \beta < \alpha + \gamma. \tag{12-26}$$

12.22. Exercise. *For all ordinals α, β,*

$$\alpha \cdot \beta = \mathrm{ord}(\alpha \cdot_o \beta),$$

by the definition of multiplication of posets in **7.38**. *It follows that multiplication of ordinals is associative but not commutative, and that*

$$0 < \alpha \,\&\, \beta < \gamma \Longrightarrow \alpha \cdot \beta < \alpha \cdot \gamma. \tag{12-27}$$

On the other hand, some of the properties of ordinal arithmetic are more easily shown by ordinal induction, directly from their recursive definitions:

12.23. Exercise (Distribution of $\cdot$ over $+$ to the right). *For all ordinals α, β, γ,*

$$\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma.$$

The solutions of these exercises suggest methods for establishing several additional results of ordinal arithmetic, which we will leave for the problems. Note that in addition to commutativity, many other properties of natural number arithmetic fail for ordinals, including the distribution of $\cdot$ over $+$ to the left, Problem **x12.13**.

12.24. Limits of ordinal sequences. If $\alpha_0 \leq \alpha_1 \leq \cdots$ is a non-decreasing sequence of ordinal numbers, we will write

$$\lim_n \alpha_n = \sup\{\alpha_n \mid n = 0, 1, \dots \}.$$

The notation is useful, but we must be careful when we use it because these limits do not satisfy the usual "limit theorems" of calculus, cf. Problem **x12.9**.

Next we describe von Neumann's elegant solution of the problem of *cardinal assignment* **4.20** for well orderable sets, which is based on the fact that every one of them is equinumerous with an ordinal, Lemma **12.9**.

12.25. Definition (Von Neumann cardinals). We set

$$|A| = \begin{cases} (\mu\xi \in \mathrm{ON})[A =_c \xi], & \text{if } A \text{ is well orderable,} \\ A, & \text{otherwise,} \end{cases} \tag{12-28}$$

and we assume from now on that the cardinal assignment we fixed back in **4.21** of Chapter **3** is, in fact, this one. The values of $|A|$ for well orderable A are the **von Neumann cardinals**,

$$\mathrm{Card}_v(\kappa) \iff \text{for some well orderable } A, \kappa = |A|, \tag{12-29}$$

and they are easy to characterize as the **initial ordinals**:

12.26. Exercise. $\mathrm{Card}_v(\kappa) \iff \mathrm{ON}(\kappa) \,\&\, (\forall \alpha < \kappa)[\kappa \neq_c \alpha]$, *and for every $\kappa \in \mathrm{Card}_v$, $|\kappa| = \kappa$.*

By Lemma **9.11**,

$$\mathrm{Card}_v(\kappa) \Longrightarrow \mathrm{Card}_v(\kappa^+),$$

and by Lemmas **9.18** and **9.20**, if $\mathscr{E}$ is a non-empty set of cardinals, then

$$(\forall \kappa \in \mathscr{E})\mathrm{Card}_v(\kappa) \Longrightarrow \mathrm{Card}_v(\inf_c(\mathscr{E})) \text{ and } \mathrm{Card}_v(\sup_c(\mathscr{E}));$$

in fact immediately from (12-19) and **12.16**, if $\mathscr{E}$ is a non-empty set of von Neumann cardinals, then

$$\inf_c(\mathscr{E}) = \min(\mathscr{E}) = \bigcap \mathscr{E}, \quad \sup_c(\mathscr{E}) = \sup \mathscr{E} = \bigcup \mathscr{E}.$$

The von Neumann construction provides a strong cardinal assignment on the class of well orderable sets, in the sense of **4.21**:

12.27. Exercise. *If A and B are well orderable, then*

$$A =_c |A|, \text{ and } A =_c B \iff |A| = |B|. \tag{12-30}$$

Moreover, for each set $\mathscr{E}$ of sets, the class $\{|X| \mid X \in \mathscr{E}\}$ is a set.

The most useful property of von Neumann cardinals is the simplest:

$$\text{if } \kappa, \lambda \in \text{Card}_v, \text{ then } \kappa =_c \lambda \iff \kappa = \lambda;$$

this follows immediately from the last two exercises and transforms all the equinumerosities between von Neumann cardinals into equations. Finally, for well orderable cardinals, we can write

$$\kappa \cdot (\lambda + \mu) = \kappa \cdot \lambda + \kappa \cdot \mu, \quad \left(\kappa^\lambda\right)^\mu = \kappa^{\lambda \cdot \mu},$$

etc., without the annoying subscript $_c$.

12.28. Cardinals, Choice and Replacement. One can make a good case that Cantor's *units* in the intuitive description of cardinals quoted in **4.19** are modeled faithfully by the von Neumann ordinals, and the quotation

> $\overline{\overline{A}}$ grows, so to speak, out of A in such a way that from every element
> x of A a special unit of A arises

describes precisely the construction of $|A| = \text{ord}(A) = \nu[A]$ relative to some best wellordering of A, Problem **x12.28**. Whatever the value of the imagery, von Neumann's construction is certainly very useful, if only because it produces a genuine cardinal arithmetic from the calculus of equinumerosities that we have developed, at least when we assume the Axiom of Choice, so that all sets have von Neumann cardinals.

About **AC**, we have been careful to state results about cardinality without assuming it, whenever this was possible, but of course the main effect has been to make clear just how poor cardinal arithmetic is without it. The main problem is the equivalence of cardinal comparability with **AC**: we do not have much of an arithmetic if we cannot compare numbers for size, and we cannot assume comparability without (necessarily) conceding the truth of the full Axiom of Choice.

Granting **AC**, how important is the existence of "true cardinals" which satisfy (12-30) and whose construction requires not only **AC** but also the Axiom of Replacement? Not much, by any account, unless you are allergic to subscripts. Thus, it might appear that von Neumann's solution of the problem of Cardinal Assignment is primarily an exercise in mathematical elegance. There is some truth to this, but one must not draw the further conclusion that the Axiom of Replacement is unimportant for cardinal arithmetic, just because

its basic identities can be established in **ZDC+AC**, as equinumerosities. The problem is that **ZDC+AC** cannot prove the existence of any cardinals above the first infinite sequence

$$\aleph_0, \aleph_1, \aleph_2, \dots,$$

and in fact it cannot even show that the sequence $(n \mapsto \aleph_n)$ exists, Problem **xB.10**. In particular, the existence of singular cardinals cannot be shown in **ZDC+AC**, so that the whole theory of cofinality remains possibly vacuous without the Axiom of Replacement.

The upshot is that to have a decent cardinal arithmetic, you must assume both the Axioms of full Choice and Replacement, i.e., to work in a theory at least as strong as **ZFDC+AC**. It is sometimes claimed that the Principle of Foundation is also necessary for cardinal arithmetic, but this is not true—although some of the most important applications of cardinals are to the structure of von Neumann's universe $\mathcal{V}$ of pure, grounded sets.

12.29. Proposition (The alephs). *By recursion on $\alpha \in \mathrm{ON}$, we set*

$$\begin{aligned} \aleph_0 &= |\mathbb{N}| = \omega, \\ \aleph_{\beta+1} &= \aleph_\beta^+, \\ \aleph_\lambda &= \sup\{\aleph_\beta \mid \beta < \lambda\}, \text{ if } \mathrm{Limit}(\lambda). \end{aligned} \tag{12-31}$$

Each $\aleph_\alpha$ is a von Neumann cardinal,

$$\alpha < \beta \Longrightarrow \aleph_\alpha <_c \aleph_\beta \quad (\alpha, \beta \in \mathrm{ON}),$$

and every infinite von Neumann cardinal is $\aleph_\alpha$ for some α.

PROOF. To check first that $\alpha \mapsto \aleph_\alpha$ is strictly increasing in cardinality, fix α and (towards a contradiction) choose β least such that $\alpha < \beta$ and $\aleph_\alpha =_c \aleph_\beta$: now $\beta \neq \alpha + 1$, since $\aleph_\alpha <_c \aleph_\alpha^+ = \aleph_{\alpha+1}$; $\beta = \gamma + 1$ for some $\gamma > \alpha$ implies that $\aleph_\alpha <_c \aleph_\gamma <_c \aleph_\beta$ (by the choice of β), which is a contradiction; and if β is limit, then $\alpha + 1 < \beta$, and so $\aleph_\alpha <_c \aleph_{\alpha+1} <_c \aleph_\beta$, which is also a contradiction.

It follows that each $\aleph_\alpha$ is a von Neumann cardinal: this is immediate for 0 and successor ordinals, and for limit λ, if $\aleph_\lambda$ is not a cardinal, then $\aleph_\lambda \leq_c \aleph_\beta$ for some $\beta < \lambda$, which contradicts $\aleph_\beta <_c \aleph_\lambda$.

Finally, to show that every von Neumann cardinal is an aleph, let (towards a contradiction) κ be the least counterexample. Directly from the definition, $\kappa \neq \omega$ and $\kappa \neq_c \lambda^+$ for any ordinal $\lambda < \kappa$, since that implies $\kappa = |\lambda|^+$, $|\lambda|$ is an aleph by the choice of κ, and then κ is also an aleph. Let

$$\beta = \{\alpha \leq \kappa \mid \aleph_\alpha <_c \kappa\}.$$

This is an ordinal, since it is closed under $<$; it is a limit ordinal, because $\aleph_\alpha < \kappa \Longrightarrow \aleph_{\alpha+1} < \kappa$; and by its definition,

$$\alpha < \beta \Longrightarrow \aleph_\beta < \kappa.$$

On the other hand, if λ is a cardinal and $\lambda < \kappa$, then $\lambda = \aleph_\alpha$ for some $\alpha < \beta$, by the choice of κ, so that finally

$$\kappa = \sup\{\aleph_\alpha \mid \alpha < \beta\} = \aleph_\beta. \qquad \dashv$$
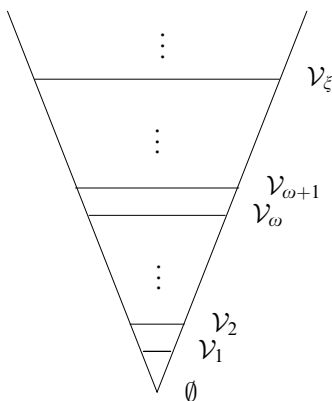
FIGURE **12.3**. Logarithmic rendition of the pure, grounded sets.

The operation $\alpha \mapsto \aleph_\alpha$ supplies a useful notation for cardinal arithmetic, especially when we accept the Axiom of Choice.

12.30. Exercise. *The Axiom of Choice* **AC** *is equivalent to the proposition that "every infinite cardinal is an aleph",*

$$\mathbf{AC} \iff (\forall \text{ infinite } A)(\exists \alpha \in \mathrm{ON})[A =_c |A| = \aleph_\alpha].$$

12.31. Exercise. (**AC**) *The Generalized Continuum Hypothesis is equivalent to the cardinal identity*

$$\mathbf{GCH} \iff 2^{\aleph_\alpha} = \aleph_{\alpha+1} \quad (\alpha \in \mathrm{ON}).$$

12.32. The Cumulative Hierarchy of Pure, Grounded Sets. For each ordinal α we define the set $\mathcal{V}_\alpha$ by the following recursion on ON:

$$\mathcal{V}_0 = \emptyset,$$
$$\mathcal{V}_{\alpha+1} = \mathcal{P}(\mathcal{V}_\alpha),$$
$$\mathcal{V}_\lambda = \bigcup_{\alpha < \lambda} \mathcal{V}_\alpha, \quad \text{if Limit}(\lambda).$$

The **von Neumann universe** is the union of all the $\mathcal{V}_\alpha$'s

$$\mathcal{V} =_{\mathrm{df}} \bigcup_{\alpha \in \mathrm{ON}} \mathcal{V}_\alpha = \{x \mid \text{for some } \alpha \in \mathrm{ON}, x \in \mathcal{V}_\alpha\}, \qquad (12\text{-}32)$$

and on it we define the **rank** operation by

$$\mathrm{Rank}(x) = (\mu\alpha \in \mathrm{ON})[x \in \mathcal{V}_{\alpha+1}] \quad (x \in \mathcal{V}). \qquad (12\text{-}33)$$

We have already used the symbol $\mathcal{V}$ to denote the class of pure grounded sets, because of the next result.

12.33. Theorem. (1) *Each $\mathcal{V}_\alpha$ is a pure, transitive, grounded set, and*

$$\alpha \le \beta \Longrightarrow \mathcal{V}_\alpha \subseteq \mathcal{V}_\beta.$$

(2) *If λ is a limit ordinal, $\lambda \ge \omega \cdot 2$, then $\mathcal{V}_\lambda$ is a Zermelo universe.*

(3) *For each pure set A, $A \subseteq \mathcal{V} \Longrightarrow A \in \mathcal{V}$.*

(4) *The von Neumann universe $\mathcal{V}$ comprises the pure, grounded sets.*

Proof. The arguments for Parts (1) and (2) are those we used to prove the corresponding properties of the basic closure sets $M(I)$ with transitive I in Lemma **11.15**, and we will not repeat them.

(3) Assume that $A \subseteq \mathcal{V}$ and let $\mathrm{Rank}[A] = \{\mathrm{Rank}(x) \mid x \in A\}$ be the image of the rank operation on A. This is a set of ordinals, so there exists some ordinal κ strictly above its members,

$$x \in A \Longrightarrow \mathrm{Rank}(x) < \kappa \Longrightarrow x \in \mathcal{V}_\kappa,$$

and (using the purity of A),

$$A = \{x \in \mathcal{V}_\kappa \mid x \in A\} \in \mathcal{V}_{\kappa+1}.$$

(4) We use the contrapositive of (3), which says that for pure A,

$$A \notin \mathcal{V} \Longrightarrow (\exists x \in A)[x \notin \mathcal{V}]. \tag{12-34}$$

Suppose first that M is pure, transitive and grounded but $M \notin \mathcal{V}$. The transitivity of M and (12-34) yield

$$(\forall x \in M \setminus \mathcal{V})(\exists y \in M \setminus \mathcal{V})[y \in x],$$

and then **DC** gives us a descending $\in$-chain which proves M ill founded, contradicting the hypothesis. Thus, every pure, transitive, grounded set is in $\mathcal{V}$, so for every pure, grounded set A, $\mathrm{TC}(A) \in \mathcal{V}$, and then $A \in \mathcal{V}$, since $A \in \mathrm{TC}(A)$ and $\mathcal{V}$ is transitive.   ⊣

Recall that from Theorem **11.34**, $\mathcal{V}$ is a **ZFDC**-universe, in fact a **ZFC**-universe if we assume **AC**.

12.34. The naive notion of pure, grounded set. In discussing the Principle of Foundation in **11.32**, we argued that the definition of the least Zermelo universe $\mathcal{Z}$ and the proofs of its basic properties can be understood directly and naively, as we usually understand mathematics, and that they carry considerable force of persuasion as an intuitive conception of "set" which justifies the axioms of **ZDC**+**AC** and the Principle of Foundation. In the same vein, we can argue that the "construction" of von Neumann's universe $\mathcal{V}$ in **12.32** and Theorem **12.33** can be understood directly and naively outside the details of any specific axiomatization, and that it puts forward a natural, intuitive conception of "pure, grounded set" which justifies the axioms of **ZFC**. It is worth looking into these arguments a little closer.

The construction of $\mathcal{Z}$ begins with the infinite set

$$\mathbb{N}_0 = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \dots\}$$

and iterates the powerset operation $\mathcal{P}(A)$ infinitely many times. Since the "infinity" of iterations involved is no more and no less than that embodied by $\mathbb{N}_0$, we can say that *to understand $\mathcal{Z}$ we must understand two infinitary things: the set $\mathbb{N}_0$* (basically the natural numbers) *and the powerset operation*.

The construction of $\mathcal{V}$ starts (literally) with nothing, the empty set, but it proceeds to iterate the powerset operation $\mathcal{P}(A)$ through all the ordinals. On the same analysis, it is fair to say that *to understand $\mathcal{V}$ we must understand the class of ordinals* ON *and the powerset operation*. One may attempt to speak eloquently about the ordinals and justify them, as one might try to justify the natural numbers or the powerset operation. It should be clear, however, that the ordinals represent a separate and different new ingredient in our intuitive understanding of $\mathcal{V}$, they cannot be reduced to $\mathbb{N}_0$ and the taking of powersets. From this point of view, the justification of the axioms of **ZFC** which we find in this intuitive construction is considerably weaker than the justification of **ZDC**+**AC** we get from contemplating $\mathcal{Z}$.

In Appendix **B** we will consider alternative set universes, including some which contain both atoms and ill founded sets, and in more advanced textbooks one can find a multitude of fascinating models of set theory constructed (primarily) by extensions and combinations of Gödel's constructibility and Cohen's forcing. Part of the reason we have worked here in the weak systems of **ZDC** and **ZFDC** is to ensure that the elementary results of axiomatic set theory we have established apply directly to (essentially) all these models. These models, however, are all constructed starting with some given model of **ZFC**, and it is not clear how to produce for any of them independent, intuitive notions of **what sets are**, which justify directly the axioms they satisfy without also justifying the axioms of **ZFC**.

It appears that (as of now), the intuitive conception of **pure, grounded set**, which is gleaned from an informal analysis of **12.32** and **12.33**, is by far the best replacement we have for Cantor's unfettered (and self-contradictory) notion of "collection into a whole of definite and separate objects".

12.35. About atoms and applications. In **3.25** we argued (with Zermelo) that it is useful to allow atoms in axiomatic set theory, so that our theorems apply directly to sets of planets or frogs, as well as to the pure, grounded sets made up out of nothing. Can the universally accepted, standard theory **ZFC** which does not allow atoms justify the applications of set theory? There are two good responses to this.

First, we can *model* physical objects and relations among them by *structures* made up of pure sets, much as we model (up to isomorphism) ordered pairs, functions, the natural numbers, etc. For example, to study the behavior of a system of heavenly bodies $P_1, \ldots, P_k$ interacting with each other under gravity, we might represent each of them by a function $\overline{P}_i : \mathbb{R} \to \mathbb{R}^7$, which assigns to each real number $t \in \mathbb{R}$ the mass, position and velocity of P_i, relative to some fixed coordinate system and units. The laws of gravity and motion will then determine these functions; and the physicist does not care whether the set theoretic objects which model the functions $\overline{P}_1, \ldots, \overline{P}_k$ are pure or not—what matters are the relations among these functions, which can then be interpreted as relations among the planets and checked against reality by observation.
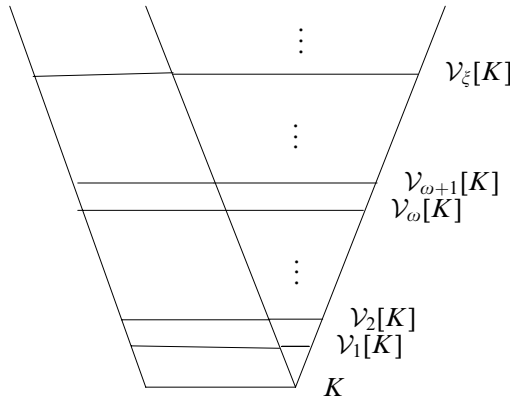
FIGURE **12.4**. Rendition of the grounded sets over K.

Second, if we value the ability to talk directly about planets within set theory, we can allow a class of atoms K and replace $\mathcal{V}$ by the class $\mathcal{V}[K]$ of grounded sets supported by K defined in **11.38**. This is a **ZFDC**-universe by Problem **x11.22**; it satisfies the Axiom of Choice;, and, in the interesting case, when K is a set, looks very much like and has essentially all the properties of $\mathcal{V}$, and by the same proofs, cf. Problem **x12.32**. It can be constructed in **ZFDC** by the ordinal recursion

$$
\begin{aligned}
\mathcal{V}_0[K] &= K, \\
\mathcal{V}_{\alpha+1}[K] &= \mathcal{P}(\mathcal{V}_\alpha[K]), \\
\mathcal{V}_\lambda[K] &= \textstyle\bigcup_{\alpha<\lambda} \mathcal{V}_\alpha[K], \quad \text{if } \mathrm{Limit}(\lambda), \\
\mathcal{V}[K] &= \textstyle\bigcup_\alpha \mathcal{V}_\alpha[K].
\end{aligned}
$$

The main point is that the physicist does not care about the difference between these two approaches, and probably cannot even see it: because all that matters for the application of mathematics (in the example) are the functions $\overline{P}_i$ which codify the properties of the planets, just as all that matters about the natural numbers is that they form a Peano system—what they actually *are* is of no consequence. And so, ultimately, it is most useful to "ban" atoms and accept the simpler **ZFC** as the "standard" axiomatic set theory, which is what is done without exception in all advanced work on our topic.

Problems for Chapter 12

x12.1. Prove (1) of Theorem **12.15**.

x12.2. Prove (2) of Theorem **12.15**.

x12.3. Prove (3) of Theorem **12.15**.

*__x12.4. Characterization of von Neumann ordinals.__ Suppose $\phi(V)$ is a definite operation which assigns well ordered sets to well ordered sets and which satisfies the following three conditions:

$$V =_o \phi(V),$$
$$U \leq_o V \Longrightarrow \phi(U) \sqsubseteq \phi(V),$$
$$\text{Field}(\phi(V)) = \{\text{Field}(\phi(U)) \mid U <_o V\}.$$

Prove that $\text{ord}(V) = \alpha \Longrightarrow \phi(V) = (\alpha, \leq_\alpha)$.

x12.5. The class ON is not a set.

x12.6. Justify the definition of ordinal multiplication in Theorem **12.19**.

x12.7. For all ordinals $\alpha, \beta, \gamma, \delta$:

$$0 + \alpha = \alpha, \text{ and } \omega \leq \alpha \Longrightarrow 1 + \alpha = \alpha,$$
$$0 < \beta \Longrightarrow \alpha < \alpha + \beta,$$
$$\alpha \leq \beta \,\&\, \gamma \leq \delta \Longrightarrow \alpha + \gamma \leq \beta + \delta,$$
$$\alpha \leq \beta \,\&\, \gamma < \delta \Longrightarrow \alpha + \gamma < \beta + \delta.$$

Show also that, in general,

$$\alpha < \beta \text{ does not imply } \alpha + \gamma < \beta + \gamma.$$

x12.8. If $\alpha_0 < \alpha_1 < \cdots$ is a strictly increasing sequence of ordinals, then the limit $\lim_n \alpha_n$ is a limit ordinal.

x12.9. Give examples of strictly increasing sequences of ordinals such that

$$\lim_n (\alpha_n + \beta) \neq \lim_n \alpha_n + \beta,$$
$$\lim_n (\alpha_n + \beta_n) \neq \lim_n \alpha_n + \lim_n \beta_n.$$

x12.10. For all ordinals $\alpha, \beta, \gamma, \delta$:

$$0 \cdot \alpha = 0$$
$$0 < \alpha \,\&\, 1 < \beta \Longrightarrow \alpha < \alpha \cdot \beta$$
$$\alpha \leq \beta \,\&\, \gamma \leq \delta \Longrightarrow \alpha \cdot \gamma \leq \beta \cdot \delta$$
$$0 < \alpha \leq \beta \,\&\, \gamma < \delta \Longrightarrow \alpha \cdot \gamma < \beta \cdot \delta.$$

Show also that even when $\gamma > 0$, in general,

$$\alpha < \beta \text{ does not imply } \alpha \cdot \gamma < \beta \cdot \gamma.$$

x12.11 (Cancellation laws). For all ordinals α, β, γ,

$$\alpha + \beta < \alpha + \gamma \implies \beta < \gamma,$$
$$\alpha + \beta = \alpha + \gamma \implies \beta = \gamma,$$
$$\alpha \cdot \beta < \alpha \cdot \gamma \implies \beta < \gamma,$$
$$0 < \alpha \,\&\, \alpha \cdot \beta = \alpha \cdot \gamma \implies \beta = \gamma.$$

Show also that, in general,

$$0 < \alpha \;\&\; \beta \cdot \alpha = \gamma \cdot \alpha \text{ does not imply } \beta = \gamma.$$

x12.12. For all $\alpha \geq \omega$ and $n < \omega$:

$$n + \alpha = \alpha,$$
$$(\alpha + 1) \cdot n = \alpha \cdot n + 1 \quad (n > 1),$$
$$(\alpha + 1) \cdot \omega = \alpha \cdot \omega.$$

x12.13. Give an example of three ordinals α, β, γ for which

$$(\alpha + \beta) \cdot \gamma \neq \alpha \cdot \gamma + \beta \cdot \gamma.$$

HINT. Use the preceding problem.

x12.14. If $n < m$, then $\omega^n + \omega^m = \omega^m$.

x12.15 (Ordinal subtraction). If $\alpha \leq \gamma$, then there exists exactly one β such that $\gamma = \alpha + \beta$.

x12.16. Every ordinal $\alpha < \omega^2$ can be written uniquely in the form

$$\alpha = \omega \cdot x + y \qquad (x, y < \omega).$$

*__x12.17.__ For any ordinal $\alpha < \omega^N$, there are unique $n < N, x < \omega, \beta < \omega^n$ such that $\alpha = \omega^n \cdot x + \beta$. HINT. Choose n largest such that $\omega^n \leq \alpha$, and x largest such that $\omega^n \cdot x \leq \alpha$.

x12.18. For any $N > 0$, every ordinal $\alpha < \omega^N$ can be written uniquely in the form

$$\alpha = \omega^{n_1} \cdot x_1 + \omega^{n_2} \cdot x_2 + \cdots + \omega^{n_s} \cdot x_s + x_{s+1}, \qquad (12\text{-}35)$$

where $N > n_1 > n_2 > \cdots > n_s$ and $x_1, \ldots, x_{s+1} < \omega$. HINT. Use induction on N and the preceding problem.

12.36. Definition (Normal operations). A unary, definite operation F on the ordinals is **normal** if it is strictly increasing

$$\alpha < \beta \Longrightarrow F(\alpha) < F(\beta),$$

and continuous at limit ordinals, i.e.,

$$F(\lambda) = \sup\{F(\beta) \mid \beta < \lambda\}, \text{ if } \mathrm{Limit}(\lambda).$$

x12.19. If $F : \mathrm{ON} \to \mathrm{ON}$ is a normal operation, then for every $\alpha, \alpha \leq F(\alpha)$, and for every limit ordinal λ, $F(\lambda)$ is a limit ordinal.

x12.20. For any fixed α, the operations

$$S_\alpha(\beta) = \alpha + \beta, \quad P_\alpha(\beta) = \alpha \cdot \beta$$

of "adding to" and "multiplying on the left by" α are normal.

x12.21. The composition $F(\alpha) = G(H(\alpha))$ of normal operation is normal.

x12.22. Define *ordinal exponentiation* α^β (for $\alpha > 1$), so that

$$\alpha^0 = 1,$$
$$\alpha^{S(\beta)} = \alpha^\beta \cdot \alpha,$$
$$\alpha^\lambda = \sup\{\alpha^\beta \mid \beta < \lambda\}, \text{ if Limit}(\lambda),$$

and prove the following:

(1) If $\beta < \gamma$, then $\alpha^\beta < \alpha^\gamma$.

(2) For each $\alpha > 1$, the operation $E_\alpha(\beta) = \alpha^\beta$ is normal.

(3) $\alpha^{(\beta+\gamma)} = \alpha^\beta \cdot \alpha^\gamma$.

(4) $\left(\alpha^\beta\right)^\gamma = \alpha^{\beta \cdot \gamma}$.

HINT. Exercises **x12.20** and **x12.21** and (2) simplify considerably the proofs of (3) and (4).

*****x12.23.** If $\alpha > 0$, then there is a largest β such that $\omega^\beta \leq \alpha$, and for that β, $\alpha = \omega^\beta + \gamma$ with some $\gamma < \alpha$. HINT. Use Problem **x12.15** to get the γ, and prove that $\gamma < \omega^{\beta+1}$, which contradicts $\gamma \geq \alpha$.

*****x12.24.** (1) If $\beta < \gamma$, then $\omega^\beta + \omega^\gamma = \omega^\gamma$.

(2) Every ordinal $\alpha > 0$ can be written uniquely in the form $\alpha = \omega^\beta + \gamma$ with $\gamma < \alpha$.

*****x12.25** (Cantor normal form)**.** Every ordinal $\alpha > 0$ can be written uniquely in the form of a finite sum of non-increasing powers of ω,

$$\alpha = \omega^{\beta_1} + \omega^{\beta_2} + \cdots + \omega^{\beta_s} \quad (\beta_1 \geq \beta_2 \geq \cdots \geq \beta_s), \qquad (12\text{-}36)$$

or, equivalently,

$$\alpha = \omega^{\beta_1} \cdot n_1 + \omega^{\beta_2} \cdots n_2 + \cdot + \omega^{\beta_t} \cdot n_t$$
$$(\beta_1 > \beta_2 > \cdots > \beta_t, n_i < \omega, n_i \neq 0). \quad (12\text{-}37)$$

HINT. For the uniqueness, prove first by induction on s that

if $\beta \geq \beta_1 \geq \beta_2 \geq \cdots \geq \beta_s$ and $\gamma = \omega^{\beta_1} + \omega^{\beta_2} + \cdots + \omega^{\beta_s}$,

$$\text{then } \gamma < \omega^\beta + \gamma.$$

If ε_0 is the least fixed point of the normal operation $(\alpha \mapsto \omega^\alpha)$, then its Cantor canonical form is the useless

$$\varepsilon_0 = \omega^{\varepsilon_0} = \omega^{\omega^{\varepsilon_0}} = \cdots$$

But the ordinals less than ε_0 have non-trivial Cantor canonical forms which provide simple and (sometimes) useful representations of them.

x12.26. Find the Cantor canonical form of $\omega \cdot (\omega^\omega + 1) + (\omega^\omega + 1) \cdot \omega$.

*****x12.27.** The only ordinals which belong to the least Zermelo universe $\mathcal{Z}$ are the finite ones.

x12.28. If $\leq$ is a best wellordering of A, then $|A| = \boldsymbol{v}_U[A]$, i.e., $|A|$ is the ordinal assigned to the well ordered set $(A, \leq)$ by its von Neumann map.

x12.29. The class Card_v of von Neumann cardinal numbers is not a set.

x12.30. (**AC**) The definite operation $\beth_\alpha$ (read "Beth-alpha") is defined by the following recursion on ON:

$$\begin{aligned}
\beth_0 &= \aleph_0 = |\mathbb{N}| = \omega, \\
\beth_{\beta+1} &= 2^{\beth_\beta}, \\
\beth_\lambda &= \sup\{\beth_\beta \mid \beta < \lambda\}, \text{ if } \mathrm{Limit}(\lambda).
\end{aligned} \tag{12-38}$$

Prove that for every ordinal α,

$$|\mathcal{V}_{\omega+\alpha}| = \beth_\alpha.$$

x12.31. For every ordinal α, $\mathrm{Rank}(\alpha) = \alpha$.

x12.32. For each set of atoms K, the rank hierarchy $(\alpha \mapsto \mathcal{V}_\alpha[K])$ has the following properties.

(1) Each $\mathcal{V}_\alpha[K]$ is a transitive, grounded set, supported by K, and

$$\alpha \leq \beta \Longrightarrow \mathcal{V}_\alpha[K] \subseteq \mathcal{V}_\beta[K].$$

(2) If λ is a limit ordinal, $\lambda \geq \omega \cdot 2$, then $\mathcal{V}_\lambda[K]$ is a Zermelo universe.
(3) For each set A supported by K, $A \subseteq \mathcal{V}[K] \Longrightarrow A \in \mathcal{V}[K]$.
(4) The universe $\mathcal{V}[K]$ comprises the grounded sets supported by K.

x12.33. For each set of atoms K and each ordinal α,

$$\mathcal{V}_\alpha = \mathcal{V}_\alpha[K] \cap \{x \mid x \text{ is pure}\}.$$

*__x12.34.__ Suppose $\pi : K \rightarrowtail\!\!\!\rightarrow K$ is a permutation of a set of atoms K. Prove that there is a unique extension $\pi^* : \mathcal{V}[K] \rightarrowtail\!\!\!\rightarrow \mathcal{V}[K]$ of π, which is an **automorphism** of $\mathcal{V}[K]$, i.e.,

$$x \in y \iff \pi^*(x) \in \pi^*(y) \qquad (x, y \in \mathcal{V}[K]).$$

12.37. Definition (Closed unbounded classes)**.** A class of ordinals $M \subseteq \mathrm{ON}$ is **unbounded** if

$$(\forall \xi \in \mathrm{ON})(\exists \alpha \in \mathrm{ON})[\xi < \alpha \,\&\, \alpha \in M];$$

and it is **closed** if for every set of ordinals $A \neq \emptyset$,

$$A \subseteq M \Longrightarrow \sup A \in M.$$

12.38. Exercise. *If M is a closed class of ordinals and*

$$\alpha_0 < \alpha_1 < \cdots \qquad (\alpha_n \in M)$$

is an increasing sequence of ordinals in M, then $\lim_n \alpha_n \in M$.

*__x12.35.__ If M_1 and M_2 are closed, unbounded classes of ordinals, then their intersection $M_1 \cap M_2$ is closed and unbounded.

*$\textbf{x12.36.}$ Every normal operation on the ordinals has a fixed point, i.e., for some ordinal, α, $F(\alpha) = \alpha$; in fact, the class

$$\mathrm{fp}(M) = \{\alpha \in \mathrm{ON} \mid F(\alpha) = \alpha\}$$

of fixed points of F is closed and unbounded. $\textsc{Hint.}$ To get one fixed point, let $\alpha_0 = 0$, $\alpha_{n+1} = F(\alpha_n)$, and show that $F(\lim_n \alpha_n) = \lim_n \alpha_n$.

$\textbf{x12.37.}$ (1) There exists a von Neumann cardinal κ, such that

$$\kappa = \aleph_\kappa.$$

(2) ($\textbf{AC}$) There exists a von Neumann cardinal λ, such that

$$\lambda = \beth_\lambda.$$

$\textbf{12.39. Definition.}$ Suppose $\alpha \leq \beta$ are infinite limit ordinals. A function $f : \alpha \to \beta$ is $\textbf{cofinal}$ if

$$\sup\{f(\xi) \mid \xi < \alpha\} = \beta.$$

The identity $(\xi \mapsto \xi)$ is a cofinal function on every limit ordinal, for example, but $(n \mapsto \aleph_n)$ is also cofinal, from ω to $\aleph_\omega$.

The next problem establishes a useful characterization of the cofinality operation, defined in $\textbf{9.23}$, which is often taken as its definition.

$\textbf{x12.38.}$ For each von Neumann cardinal κ,

$$\mathrm{cf}(\kappa) = \min\{\alpha \mid \text{there exists a cofinal } f : \alpha \to \kappa\}.$$

$\textbf{x12.39.}$ Prove that for all von Neumann cardinals $\lambda \leq \kappa$, there exists a cofinal function $f : \lambda \to \kappa$ if and only if $\mathrm{cf}(\lambda) = \mathrm{cf}(\kappa)$.

$\textbf{x12.40.}$ For every regular λ, $\mathrm{cf}(\aleph_\lambda) = \lambda$, so there exists cardinals of every regular cofinality.

*$\textbf{x12.41.}$ There exist singular cardinals of every regular cofinality.

$\textbf{x12.42.}$ For every von Neumann cardinal λ with $\mathrm{cf}(\lambda) \geq \aleph_1$, the set $\mathcal{V}_\lambda$ is a Zermelo universe which further satisfies the following special case of the Replacement Axiom: if F is a definite operation and $x \in \mathcal{V}_\lambda \Longrightarrow F(x) \in \mathcal{V}_\lambda$, then the image $F[A]$ of every countable $A \in \mathcal{V}_\lambda$ is also in $\mathcal{V}_\lambda$.

$\textbf{12.40. Definition.}$ ($\textbf{AC}$) An uncountable cardinal number κ is $\textbf{strongly inaccessible}$ if it is regular and

$$\lambda < \kappa \Longrightarrow 2^\lambda < \kappa.$$

*$\textbf{x12.43.}$ ($\textbf{AC}$) If κ is strongly inaccessible, then $\mathcal{V}_\kappa$ is a $\textbf{ZFC}$-universe.

*$\textbf{x12.44.}$ ($\textbf{AC}$) If a pure and grounded set M is a $\textbf{ZFDC}$-universe, then $M = \mathcal{V}_\kappa$, for a strongly inaccessible cardinal κ.

12.41. Frege cardinals. We have followed Cantor in his approach to the theory of cardinal numbers, by which the property

$$A =_c |A| \tag{12-39}$$

is most fundamental. There is another approach due to Frege, which takes $|A|$ to be not a set of "units" equinumerous with A, but the abstract notion of "being equinumerous with A". Frege understands "1", for example, as the common property of all singletons. To model this idea in set theory, it is not important to define $|A|$ so that it is equinumerous with A, in fact, it is not even necessary for $|A|$ to be a set! The only important property of cardinals is the last one,

$$A =_c B \iff |A| = |B|, \tag{12-40}$$

which (in effect) makes the operation $|A|$ a "determining surjection" of the "equivalence condition" $=_c$, with the cardinal numbers as the "quotient class", in the natural extension to classes of the terminology in **x4.5**. Frege tried to capture this idea by setting

$$|A| = \{X \mid X =_c A\}, \tag{12-41}$$

but the class $\{X \mid X =_c A\}$ is not a set (when $A \neq \emptyset$, easily) and the (necessary for the theory) assumption that it is led Frege to a contradiction.

Von Neumann cardinals reconcile the Cantor and Frege approaches by satisfying both (12-39) and (12-40), but their definition depends on both the Axioms of Choice and Replacement. Problem **x12.46** describes another approach, due to Scott, which succeeds in defining *Frege cardinals* without the Axiom of Choice, but (essentially) only for pure, grounded sets. Scott's construction is important, not so much for rescuing Frege cardinals (since little cardinal arithmetic can be done without **AC** anyway), but for the simplicity and elegance of the method, which has many uses beyond the present one. First we describe Scott's general method, and then its application to Frege cardinals.

12.42. Definition. An **equivalence condition** on a class A is any binary, definite condition $\sim$ which has the properties of an equivalence relation, i.e., for all objects $x, y, z \in A$

$$x \sim x, \ x \sim y \Longrightarrow y \sim x, \ x \sim y \ \& \ y \sim z \Longrightarrow x \sim z.$$

A unary, definite operation F is **determining for** $\sim$ if

$$x \sim y \iff F(x) = F(y) \quad (x, y \in A);$$

the class of values of F for arguments in A is the **quotient** (class) of A by $\sim$ determined by F,

$$F[A] =_{df} \{F(x) \mid x \in A\}.$$

For example, the condition $=_o$ of similarity is an equivalence condition on the class of well ordered sets, and the von Neumann ordinal assignment $\mathrm{ord}(U)$ is a determining operation for it, with quotient the class ON of ordinals. The

condition $=_c$ of equinumerosity is an equivalence condition on the class of well orderable sets, and the von Neumann cardinal $|A|$ operation is determining for it, with quotient the class of von Neumann cardinals.

We can think of an equivalence condition $\sim$ on a class A very much as if A were a set and $\sim \subseteq A \times A$ an ordinary equivalence relation on it. There is no easy way to define a determining operation for $\sim$, however, because the classical construction of equivalence classes **4.12** truly leads to "classes" which need not be sets in this case: this is the problem with Frege's definition of the number 1 above.

x12.45. (Scott) Suppose $\sim$ is an equivalence condition on a class A of pure, grounded sets, and for each $x \in A$ let

$$\rho(x) =_{\mathrm{df}} (\mu\alpha \in \mathrm{ON})(\exists y \in \mathcal{V}_\alpha)[y \sim x],$$
$$F(x) =_{\mathrm{df}} \{y \in \mathcal{V}_{\rho(x)} \mid y \sim x\}.$$

Prove that F is a determining operation for $\sim$ on A.

x12.46. (Scott) Define the **Scott cardinal** $|A|_s$ of every set A which is equinumerous with a pure, grounded set, so that for all such sets A and B,

$$A =_c B \iff |A|_s = |B|_s.$$

THE REAL NUMBERS

In this Appendix we will show how the rational and the real numbers can be represented faithfully in set theory as the natural numbers are; that is, we will identify some characteristic, set theoretic properties of these systems and we will prove from the axioms of **ZDC** the existence and uniqueness (up to isomorphism) of structured sets with these properties. The proofs are quite simple as far as set theory goes, but they use ideas from algebra and analysis, which we will present in outline.

There are two, standard representations of the real numbers in set theory, Cantor's *equivalence classes of Cauchy sequences* and Dedekind's *Dedekind cuts*. Here we will employ a somewhat novel construction which interweaves between these two and (incidentally) establishes their equivalence. The basic notion that we need is that of a quotient of a set A by an equivalence relation on A, described in Problem **x4.5**.

Recall that a **determining surjection** for an equivalence relation $\sim$ on a set A is any surjection

$$\pi : A \twoheadrightarrow B$$

such that for all $x, y \in A$,

$$x \sim y \iff \pi(x) = \pi(y).$$

When this holds, we call B a **quotient of A by** $\sim$. The *canonical surjection* of $\sim$ is the mapping

$$x \mapsto [x/\sim] \quad (x \in A)$$

with quotient the set of equivalence classes $[\![A/\sim]\!]$, but there may be others, more illuminating of the situation, e.g., those described in Problems **x4.6**, **x4.7** and **x4.8**. Determining surjections are especially useful in the study of congruence relations.

A.1. Definition. Suppose $\sim$ is an equivalence relation on A and

$$f : A \times A \to A$$

The material in this Appendix can be read after Chapter **10**. It assumes some theoretical knowledge of the Calculus, and it is not a prerequisite for understanding the remainder of these Notes.
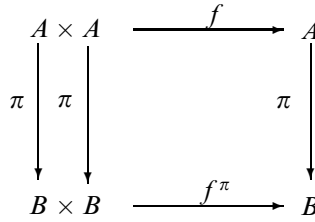
FIGURE **A.1**. Congruence for f.

is s binary function. We call $\sim$ a **congruence for** f if for all $x, x', y, y' \in A$,

$$x \sim x' \,\&\, y \sim y' \Longrightarrow f(x, y) \sim f(x', y').$$

Similarly, $\sim$ is a **congruence for** a binary relation $P \subseteq A \times A$, if for all x, x', y, y',

$$x \sim x' \,\&\, y \sim y' \Longrightarrow [x \, P \, y \iff x' \, P \, y].$$

We can obviously define the notion of congruence for functions and relations of any number of arguments, in the same way.

The next theorem deals with one of the simplest and most basic algebraic constructions.

A.2. Theorem. *Let $\pi : A \twoheadrightarrow B$ be a determining surjection of some equivalence relation $\sim$, so that for all x, y in A, $x \sim y \iff \pi(x) = \pi(y)$.*

(1) *If $\sim$ is a congruence for a function $f : A \times A \to A$, then there exists exactly one function $f^\pi : B \times B \to B$ on the quotient B such that the diagram in Figure **A.1** commutes, i.e.,*

$$f^\pi(\pi(x), \pi(y)) = \pi(f(x, y)) \quad (x, y \in A). \tag{A-1}$$

(2) *If $\sim$ is a congruence for a relation $P \subseteq A \times A$, then there exists exactly one relation $P^\pi \subseteq B \times B$ on the quotient B such that*

$$\pi(x) \, P^\pi \, \pi(y) \iff x \, P \, y \quad (x, y \in A).$$

PROOF. The form of (A-1) makes it clear that at most one function can satisfy it, so it is enough to show that at least one function does. Put

$$f^\pi =_{\mathrm{df}} \{((\pi(x), \pi(y)), \pi(z)) \mid x, y, z \in A \,\&\, f(x, y) = z\}.$$

To verify that the set of pairs f^π is a function, we must check that

$$((u, v), w), ((u, v), w') \in f^\pi \Longrightarrow w = w'. \tag{A-2}$$

From the hypothesis of (A-2) and the definition of f^π, there exist $x, y, z \in A$ such that

$$u = \pi(x), v = \pi(y), \quad w = \pi(z), \quad f(x, y) = z$$

and also $x', y', z' \in A$, so that

$$u = \pi(x'), \; v = \pi(y'), \quad w' = \pi(z'), \quad f(x', y') = z'.$$

It follows that

$$\pi(x) = \pi(x'), \ \pi(y) = \pi(y'), \ \pi(f(x, y)) = \pi(f(x', y'))$$

since $\sim$ is determined by π and it is a congruence for f, and the last of these equalities yields the desired $w = w'$.

The characteristic property (A-1) of f^π follows immediately from its definition and the proof of Part (2) is similar.   ⊣

A.3. Exercise. *Prove Part* (2) *of* **A.2.**

The axiomatic characterizations of the rationals and the reals are based on the notion of an *ordered field*, which codifies the basic properties of addition, multiplication and ordering in these number systems.

A.4. Definition. A field is a structured set $(F, 0, 1, +, \cdot)$ of objects with the following properties.

(F1) $0, 1 \in F$, $0 \neq 1$, and $+, \cdot$ are binary functions on F.

(F2) The addition function $+$ satisfies the identities

1. $(x + y) + z = x + (y + z)$,
2. $x + y = y + x$,
3. $x + 0 = x$,

and for every x, there exists some x', such that $x + x' = 0$.

(F3) The multiplication function $\cdot$ satisfies the identities

1. $(x \cdot y) \cdot z = x \cdot (y \cdot z)$,
2. $x \cdot y = y \cdot x$,
3. $x \cdot 1 = x$,

and for every $x \neq 0$, there exists some x'', such that $x \cdot x'' = 1$.

(F4) Addition and multiplication together satisfy the identity

$$x \cdot (y + z) = x \cdot y + x \cdot z.$$

A.5. Lemma. *Every field F has the following properties*:

(1) *For each x there exists exactly one x' such that $x + x' = 0$, and we denote it $-x$; for every $x \neq 0$ there exists exactly one x'' such that $x \cdot x'' = 1$ and we denote it by x^{-1}.*

(2) $x \cdot 0 = 0$.

(3) $x \cdot y = 0 \Longrightarrow x = 0 \lor y = 0$.

(4) $(-x) \cdot y = -(x \cdot y)$.

PROOF. (1) If $x + x' = 0$ and $x + y = 0$, then from the axioms

$$\begin{aligned}
y = y + 0 = 0 + y &= (x + x') + y = x + (x' + y) \\
&= x + (y + x') = (x + y) + x' \\
&= 0 + x' = x' + 0 = x'.
\end{aligned}$$

The proof about x^{-1} is similar.

(2) $x \cdot 0 = x \cdot (0 + 0) = x \cdot 0 + x \cdot 0$, and therefore

$$0 = x \cdot 0 + -(x \cdot 0) = (x \cdot 0 + x \cdot 0) + -(x \cdot 0)$$
$$= x \cdot 0 + ((x \cdot 0) + -(x \cdot 0)) = x \cdot 0 + 0 = x \cdot 0.$$

(3) If $y \neq 0$, then some y^{-1} exists such that $y \cdot y^{-1} = 1$, so that

$$x = x \cdot 1 = x \cdot (y \cdot y^{-1}) = (x \cdot y) \cdot y^{-1} = 0 \cdot y^{-1} = y^{-1} \cdot 0 = 0.$$

(4) $x \cdot y + (-x) \cdot y = y \cdot x + y \cdot (-x) = y \cdot (x + (-x)) = y \cdot 0 = 0$, and
(1) implies that $(-x) \cdot y = -(x \cdot y)$.   ⊣

We gave this proof in full as an example, justifying each step from the
field axioms. In the future we will cut corners, skip details or (more often)
use identities which obviously hold in every field without proof or explicit
mention.

A.6. Exercise. *Every field F satisfies the identity*

$$(x + y)^2 = x^2 + 2xy + y^2,$$

where $2 = 1 + 1$. (Give the proof in full detail.)

A.7. Exercise. *The doubleton* $\{0, 1\}$ *of the first two natural numbers is a field,
with the obvious operations, and in this field* $1 + 1 = 0$. It follows that the field
axioms do not imply $1 + 1 \neq 0$ and we must be a bit careful!

A.8. Definition. An **ordered field** is a structured set

$$(F, 0, 1, +, \cdot, \leq)$$

where $(F, 0, 1, +, \cdot)$ is a field, the binary relation $\leq$ is a linear ordering of F
and the following conditions hold for all $x, y, z \in F$:

$$x \leq y \implies x + z \leq y + z,$$
$$z > 0 \,\&\, x \leq y \implies z \cdot x \leq z \cdot y,$$

where $z > 0$ naturally abbreviates $0 \leq z \,\&\, z \neq 0$.

A.9. Exercise. *In every ordered field,*

$$z > 0 \,\&\, x < y \implies z \cdot x < z \cdot y.$$

A.10. Lemma. *Every element x in an ordered field F satisfies the inequality*
$x \cdot x = x^2 \geq 0$, *so that* $0 < 1$ *and for all* x, $x > 0 \implies x + 1 > 0$.

Proof. If $x = 0$, then $x^2 = 0 \geq 0$, and if $x > 0$, then $x \cdot x \geq x \cdot 0 = 0$, so
that the only interesting case is when $x < 0$. Adding $-x$ to both sides of this
inequality, we get $0 < -x$, so that we can multiply $x < 0$ by $-x$ and we get
$(-x) \cdot x < (-x) \cdot 0$, i.e., $-(x^2) < 0$ from the preceding Lemma, and adding
x^2 to this inequality we get $0 < x^2$. The conclusion $0 < 1$ follows because
$0 \neq 1$ and $1 = 1^2$, and the last claim holds because $0 < x \implies 1 < x + 1$, so
that $0 < x + 1$ by the transitivity of $\leq$.   ⊣

The Lemma makes it clear that we will not find in ordered fields the anomaly
$1 + 1 = 0$ of Exercise **A.7**. Something much stronger is true.

A.11. Lemma. *Suppose F is an ordered field and set*

$$\mathbb{N}_F = \bigcap \{X \subseteq F \mid 0 \in X \ \& \ (\forall x)[x \in X \Longrightarrow x + 1 \in X]\};$$

it follows that $(\mathbb{N}_F, 0, (x \mapsto x + 1))$ is a Peano system.

PROOF. By $(x \mapsto x + 1)$ we mean the function S which associates with each $x \in \mathbb{N}_F$ the element $x + 1$ of F, which is also a member of $\mathbb{N}_F$ by the definition. The first three axioms of Peano are obvious and the fourth $(x + 1 \neq 0)$ holds because by the definition,

$$\mathbb{N}_F \subseteq \{x \in F \mid 0 \leq x\},$$

and by the Lemma, $x \geq 0 \Longrightarrow x + 1 \geq 1 > 0$. The Induction Axiom follows immediately from the definition of $\mathbb{N}_F$ as an intersection.   ⊣

A.12. Exercise. *Suppose F is an ordered field, $\mathbb{N} = \mathbb{N}_F$ is the set of its natural numbers and $+_\mathbb{N}$, $\cdot_\mathbb{N}$, $\leq_\mathbb{N}$ the addition, multiplication and the wellordering of $\mathbb{N}_F$ as these are defined in Chapter **5**. Prove that these functions and the relation $\leq$ coincide with the respective objects in F, e.g.,*

$$(\forall x, y \in \mathbb{N})[x +_\mathbb{N} y = x + y].$$

The members of $\mathbb{N}_F$ are the **natural numbers of F**. The (rational) **integers** of F comprise the natural numbers and their negations,

$$\mathbb{Z}_F = \mathbb{N}_F \cup \{-x \mid x \in \mathbb{N}_F\},$$

and it is easy to check that they are exactly all differences $u - v$, with $u, v \in \mathbb{N}_F$.

The basic idea for the axiomatic characterization of the rationals is that they are an ordered field, and that every fraction is a quotient of integers, i.e.,

$$\frac{u - v}{m} = m^{-1} \cdot (u - v),$$

where m, u and v are natural numbers and $m \neq 0$. This simple observation yields not only the axioms for the rationals, but also proofs of their existence and uniqueness.

A.13. Definition. A **system of rational numbers** is any ordered field F which satisfies the condition

$$(\forall x \in F)(\exists m, u, v \in \mathbb{N}_F)[m \neq 0 \ \& \ x = m^{-1} \cdot (u - v)].$$

A.14. Theorem (Uniqueness of the rational numbers). *For any two systems of rational numbers F_1, F_2 there exists exactly one bijection*

$$\pi : F_1 \rightarrowtail\!\!\!\rightarrow F_2$$

which is an isomorphism, i.e.,

1. $\pi(0_1) = 0_2$, $\pi(1_1) = 1_2$.
2. $\pi(x +_1 y) = \pi(x) +_2 \pi(y)$, $\pi(x \cdot_1 y) = \pi(x) \cdot_2 \pi(y)$.
3. $x \leq_1 y \iff \pi(x) \leq_2 \pi(y)$.

In stating this theorem we decorated the various objects with the super-scripts 1 or 2 to clarify the field to which they belong, e.g., $+_1$ is addition in F_1 and 0_2 is the zero element of F_2. This is awkward and unnecessary, because it is always obvious which superscript is needed: e.g., the identity $\pi(0) = 0$ cannot mean anything else but $\pi(0_1) = 0_2$, since π is a function with domain F_1 and image F_2. In the proof and in the future we will follow the general algebraic practice by which all the zero elements are 0, all additions are $+$, etc. We will also begin to skip the $\cdot$ of multiplication,

$$xy =_{df} x \cdot y.$$

PROOF. By the uniqueness of the natural numbers and **A.11**, we know that there exists a "canonical" isomorphism

$$\rho : \mathbb{N}_1 \rightarrowtail\!\!\!\rightarrow \mathbb{N}_2,$$

where $\mathbb{N}_1$ and $\mathbb{N}_2$ are the sets of natural numbers in F_1 and F_2, respectively. We set

$$\pi = \{(m^{-1}(u - v), \rho(m)^{-1}(\rho(u) - \rho(v))) \mid m, u, v \in \mathbb{N}_1, m \neq 0\},$$

so that $\pi \subseteq F_1 \times F_2$ and it is enough to show first that π is a function, then that it is a bijection, and finally that it is an isomorphism, as we defined this in the formulation of the theorem.

To verify first that π is a function, we must show that if

$$m_1^{-1}(u_1 - v_1) = m_2^{-1}(u_2 - v_2), \tag{A-3}$$

then

$$\rho(m_1)^{-1}(\rho(u_1) - \rho(v_1)) = \rho(m_2)^{-1}(\rho(u_2) - \rho(v_2)). \tag{A-4}$$

The field axioms imply easily that (A-3) and (A-4) are respectively equivalent to

$$m_2 u_1 + m_1 v_2 = m_1 u_2 + m_2 v_1,$$
$$\rho(m_2)\rho(u_1) + \rho(m_1)\rho(v_2) = \rho(m_1)\rho(u_2) + \rho(m_2)\rho(v_1),$$

and the first of these yields immediately

$$\rho(m_2 u_1 + m_1 v_2) = \rho(m_1 u_2 + m_2 v_1)$$

which in turn implies the second, because ρ is an isomorphism of $\mathbb{N}_1$ with $\mathbb{N}_2$ and it respects addition and multiplication by Problem **x5.4**.

The same simple method can be used directly to prove the additional con-clusions, that π is one-to-one and finally an isomorphism. ⊣

A.15. Exercise. *Work out in detail the proofs of*

$$\pi(x + y) = \pi(x) + \pi(y),$$
$$x \leq y \iff \pi(x) \leq \pi(y).$$

A.16. Theorem (Existence of the rational numbers). *There exists a system of rational numbers.*

PROOF. If we have the rationals, we can define the set

$$A = \{(m, u, v) \mid m, u, v \in \mathbb{N} \,\&\, m \neq 0\}$$

of triples of natural numbers, and on it the relation

$$(m, u, v) \sim (m', u', v') \iff_{\mathrm{df}} m'u + mv' = mu' + m'v,$$

which (quite obviously) satisfies

$$(m, u, v) \sim (m', u', v') \iff \frac{u - v}{m} = \frac{u' - v'}{m'}.$$

This means that $\sim$ is an equivalence relation determined by the surjection

$$\pi : A \twoheadrightarrow \mathbb{Q}, \quad \pi(m, u, v) = \frac{u - v}{m}. \tag{A-5}$$

We do not have the rationals yet, but we have A and $\sim$: the idea for the proof is to define the rationals as a quotient of A by $\sim$ so that (A-5) holds. First we must show that

(1) *$\sim$ is an equivalence relation.* As an example, we verify that $\sim$ is transitive. From its definition, if

$$(m_1, u_1, v_1) \sim (m_2, u_2, v_2) \,\&\, (m_2, u_2, v_2) \sim (m_3, u_3, v_3),$$

then the identities

$$m_2 u_1 + m_1 v_2 = m_1 u_2 + m_2 v_1,$$
$$m_3 u_2 + m_2 v_3 = m_2 u_3 + m_3 v_2$$

hold in the natural numbers, and if we multiply the first of these by m_3 and the second by m_1 and then add them, we get

$$m_3 m_2 u_1 + m_3 m_1 v_2 + m_1 m_3 u_2 + m_1 m_2 v_3$$
$$= m_3 m_1 u_2 + m_3 m_2 v_1 + m_1 m_2 u_3 + m_1 m_3 v_2.$$

Subtract now $m_3 m_1 v_2$ and $m_1 m_3 u_2$ from the two sides and divide by m_2, which gives

$$m_3 u_1 + m_1 v_3 = m_1 u_3 + m_3 v_1,$$

i.e., $(m_1, u_1, v_1) \sim (m_3, u_3, v_3)$. Reflexivity and symmetry are proved in the same way.

(2) *Definition of the rationals.* Since $\sim$ is an equivalence relation, there exists a surjection

$$\pi : A \twoheadrightarrow \mathbb{Q}$$

onto some set $\mathbb{Q}$ which determines it, so that

$$(m_1, u_1, v_1) \sim (m_2, u_2, v_2) \iff \pi(m_1, u_1, v_1) = \pi(m_2, u_2, v_2).$$

This $\mathbb{Q}$ is the set of rationals in the system under construction, and it remains only to specify 0 and 1, to define addition, multiplication and the ordering, and finally to prove that the axioms for the rationals hold. To help follow the argument we will start right away using the notation

$$\frac{u - v}{m} =_{\mathrm{df}} \pi(m, u, v)$$

as an abbreviation, i.e., without defining separately "subtraction" or "division".

The zero and the one are defined in the obvious way,

$$0 = \frac{0-0}{1} =_{\mathrm{df}} \pi(1,0,0), \quad 1 = \frac{1-0}{1} =_{\mathrm{df}} \pi(1,1,0).$$

(3) *Addition of rationals.* With the representation of rationals as quotients of a difference of numbers by a number which we are using, the classical formula for addition of fractions takes the form

$$\frac{u_1 - v_1}{m_1} + \frac{u_2 - v_2}{m_2} = \frac{(m_2 u_1 + m_1 u_2) - (m_2 v_1 + m_1 v_2)}{m_1 m_2}.$$

So we define first on the set A the binary function f_+ which corresponds to this formula,

$$f_+((m_1, u_1, v_1), (m_2, u_2, v_2)) = (m_1 m_2, (m_2 u_1 + m_1 u_2), (m_2 v_1 + m_1 v_2)).$$

With a bit of arithmetic we can prove that for all $x, y, x', y' \in A$,

$$x \sim x' \,\&\, y \sim y' \Longrightarrow f_+(x, y) \sim f_+(x', y'),$$

i.e., $\sim$ *is a congruence for* f_+. It follows by **A.2** that there exists a (unique) function

$$+ : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$$

which satisfies the identity

$$\pi(x) + \pi(y) = \pi(f_+(x, y)) \quad (\pi(x), \pi(y) \in \mathbb{Q}).$$

Verification of the axioms (F2) for addition needs a bit more of arithmetic, but at least the condition for 0 is obvious:

$$\pi(m, u, v) + \pi(1, 0, 0) = \pi(m \cdot 1, 1 \cdot u + m \cdot 0, 1 \cdot v + m \cdot 0) = \pi(m, u, v).$$

(4) *Multiplication of rationals.* Following the same method, we define first the function $f. : A \times A \to A$ which corresponds to multiplication when we represent rationals by triples of natural numbers,

$$f.((m_1, u_1, v_1), (m_2, u_2, v_2)) =_{\mathrm{df}} (m_1 m_2, u_1 u_2 + v_1 v_2, u_1 v_2 + u_2 v_1),$$

we verify next that $\sim$ is a congruence for $f.$ and we define the multiplication operation on fractions $\cdot$ by **A.2** so that it satisfies the identity

$$\pi(x) \cdot \pi(y) = \pi(f.(x, y)) \quad (\pi(x), \pi(y) \in \mathbb{Q}).$$

Verification of axioms (F3) and (F4) requires just a few computations.

(5) *Ordering of the rationals.* The critical equivalence in this case is

$$\frac{u_1 - v_1}{m_1} \leq \frac{u_2 - v_2}{m_2} \iff m_1 v_2 + m_2 u_1 \leq m_2 v_1 + m_1 u_2.$$

We first define the relation $P \subseteq A \times A$ by

$$(m_1, u_1, v_1)\, P\, (m_2, u_2, v_2) \iff_{\mathrm{df}} m_1 v_2 + m_2 u_1 \leq m_2 v_1 + m_1 u_2,$$

we verify that $\sim$ is a congruence for P and using **A.2** we define $\leq$ on the quotient $\mathbb{Q}$ so that

$$\pi(x) \leq \pi(y) \iff x\,P\,y \quad (\pi(x), \pi(y) \in \mathbb{Q}).$$

It follows easily that $\leq$ is a linear ordering, and that the structured set

$$(\mathbb{Q}, 0, 1, +, \cdot, \leq)$$

is an ordered field.

It remains to verify that $\mathbb{Q}$ is a system of rational numbers.

Lemma 1. *For each natural number k, $\pi(1, k, 0) \in \mathbb{N}_{\mathbb{Q}}$, i.e., the rational $\pi(1, k, 0)$ belongs to the set of natural numbers of the ordered field $\mathbb{Q}$.*

Proof. By induction on k, $\pi(1, 0, 0) = 0$ (by definition) and (easily) by the definition of rational addition

$$\pi(1, Sk, 0) = \pi(1, k, 0) + 1,$$

so that $\pi(1, k, 0) \in \mathbb{N}_{\mathbb{Q}} \Longrightarrow \pi(1, Sk, 0) \in \mathbb{N}_{\mathbb{Q}}$. $\dashv$ (Lemma 1)

Lemma 2. *For all $(m, u, v) \in A$,*

$$\pi(m, u, v) = \pi(1, m, 0)^{-1}(\pi(1, u, 0) - \pi(1, v, 0)), \qquad \text{(A-6)}$$

where $^{-1}$ and $-$ are the multiplicative and additive inverse (partial) functions of the field $\mathbb{Q}$.

Proof. Having proved already that $\mathbb{Q}$ is a field, we know that (A-6) is equivalent to

$$\pi(1, m, 0)\pi(m, u, v) + \pi(1, v, 0) = \pi(1, u, 0),$$

and the latter identity is easy to verify with a direct computation.$\dashv$ (Lemma 2)

The two Lemmas together show that the structured set $(\mathbb{Q}, 0, 1, +, \cdot, \leq)$ satisfies the characteristic property of the rationals and this completes the proof. $\dashv$

As we did with the natural numbers, we now fix a specific system of rational numbers $(\mathbb{Q}, 0, 1, +, \cdot, \leq)$ whose elements we will henceforth call **rationals**. This is convenient, it helps avoid awkward expressions like "members of any system of rational numbers" and the like. However, it is important to emphasize (once more) that the significant mathematical fact is the existence and uniqueness up to isomorphism of one such system: it was precisely the corresponding mathematical facts about the natural numbers that we have used in the proofs of this Appendix, not the specific identity of "the natural numbers".

A.17. Exercise. *The set $\mathbb{Q}$ of rationals is countable.*

A.18. Exercise. *In the proof of Part (1) of the theorem we "subtracted" the same number from an identity and then "divided" an identity by the same number. Justify these steps by verifying the following two properties of the natural*

numbers:

$$x + y = x + z \implies y = z,$$
$$c \neq 0 \,\&\, c \cdot x = c \cdot y \implies x = y.$$

A.19. Exercise. *For every ordered field F, there exists exactly one* **imbedding** *of the rationals in F, i.e., an injection*

$$\pi : \mathbb{Q} \rightarrowtail F$$

which satisfies the identities

$$\pi(0) = 0, \ \pi(1) = 1,$$
$$\pi(x + y) = \pi(x) + \pi(y), \ \pi(xy) = \pi(x)\pi(y),$$
$$x \leq y \iff \pi(x) \leq \pi(y).$$

It follows that the image $\pi[\mathbb{Q}] \subseteq F$ of π is a system of rational numbers (with the 0 and 1 of F and the restrictions of the operations and the ordering of F).

There is a beautiful theorem of Cantor which characterizes the ordering of the rationals independently of their algebraic structure. For it, we need first some definitions.

A.20. Definition. Suppose $\leq$ is a linear ordering on a set A and $B \subseteq A$. We call B **dense in** A if

$$(\forall x, y \in A)[x < y \implies (\exists b \in B)[x < b \,\&\, b < y]].$$

A linear ordering $\leq$ is **dense in itself** if its field (A) is dense in A.

A.21. Exercise. *The ordering of every ordered field is dense in itself and has no minimum or maximum element.*

A.22. Theorem (Cantor). *Any two countable, linear, dense in themselves orderings $(A, \leq_A)$ and $(B, \leq_B)$ without minimum or maximum element are* **similar**, *i.e., there exists an order-preserving correspondence $f : A \rightarrowtail B$.*

In particular, every countable, linear, dense in itself ordering is similar with the ordering $\leq_{\mathbb{Q}}$ of the rational numbers.

PROOF. The hypothesis gives us enumerations without repetitions

$$A = \{a_0, a_1, \dots , \}, \quad B = \{b_0, b_1, \dots \}$$

of A and B. We will define by recursion a sequence

$$f_0, f_1, \dots ,$$

with the following properties, for every $n \in \mathbb{N}$.

(1) f_n is a *finite, partial function* from A to B, i.e.,

$$\text{Function}(f_n) \,\&\, f_n \subseteq A \times B,$$

and f_n is finite as a set of ordered pairs.

(2) f_n is monotone and one-to-one on its domain, i.e.,

$$x, y \in \text{Domain}(f_n) \,\&\, x <_A y \implies f_n(x) <_B f_n(y).$$

(3) $f_n \subseteq f_{n+1}$.

(4) $\{a_0, a_1, \ldots, a_n\} \subseteq \mathrm{Domain}(f_n)$.

(5) $\{b_0, b_1, \ldots, b_n\} \subseteq \mathrm{Image}(f_n)$.

If we can succeed in this, then the union $f =_{\mathrm{df}} \bigcup_{n=0}^{n=\infty} f_n$ is (easily) a function by (3), it is one-to-one and monotone by (2) and

$$\mathrm{Domain}(f) = A, \quad \mathrm{Image}(f) = B$$

by (4) and (5).

At the basis of the recursive definition we start with

$$p_0 = \{(a_0, b_0)\},$$

so that all the conditions of the result hold trivially.

Suppose now that we have already defined f_n and enumerate its finite domain of definition and image in increasing order:

$$\mathrm{Domain}(f_n) = D_n = \{x_0 <_A x_1 < \cdots <_A x_m\},$$
$$\mathrm{Image}(f_n) = I_n = \{y_0 <_B y_1 <_B \cdots <_B y_m\}.$$

Since f_n is monotone, we have

$$f_n(x_i) = y_i \quad (i = 0, \ldots, m).$$

We construct the next f_{n+1} in two steps, i.e., first we will define a partial function $f'_{n+1} \supseteq f_n$ which satisfies (1) – (4) and then define $f_{n+1} \supseteq f'_{n+1}$ which satisfies all (1) – (5).

STEP 1. If $a_{n+1} \in \mathrm{Domain}(f_n)$, set $f'_{n+1} = f_n$. Otherwise there are three cases.

CASE 1a. $a_{n+1} <_A x_0$. In this case we find some $y' \in B$ satisfying $y' <_B y_0$ (which exists because B has no minimum) and set

$$f'_{n+1} = f_n \cup \{(a_{n+1}, y')\}.$$

CASE 1b. $a_{n+1} >_A x_m$. In this case we find some $y' \in B$ satisfying $y' >_B y_m$ (which exists because B has no maximum) and we set

$$f'_{n+1} = f_n \cup \{(a_{n+1}, y')\}.$$

CASE 1c. For some i, $x_i <_A a_{n+1} <_A x_{i+1}$. In this case we find some $y' \in B$ satisfying $y_i <_B y' <_B y_{i+1}$ (which exists because B is dense in itself) and we set

$$f'_{n+1} = f_n \cup \{(a_{n+1}, y')\}.$$

In all cases, the proof that f'_{n+1} satisfies (1) – (4) is simple.

In STEP 2 of the construction we consider the element b_{n+1} of B and we distinguish again cases: first if $b_{n+1} \in \mathrm{Image}(f'_{n+1})$ (in which case we set $f_{n+1} = f'_{n+1}$) and if not, then three cases again, in symmetry with STEP 1. We skip the details.   ⊣

The fundamental intuition about the real numbers is that on the one hand they are an ordered field, so that their arithmetic and ordering satisfy the same laws as the rationals, and on the other, they are in one-to-one correspondence with the points of the "complete" geometric line so that there are

no "gaps" between them. In formulating the property of completeness we follow Dedekind.

A.23. Definition. A linear ordering $\leq$ on a set A is **complete** if every non-empty subset of A which has an upper bound has a least upper bound. A **system of real numbers** is any **complete ordered field**, i.e., any ordered field in which the ordering is complete.

A.24. Exercise. *The ordering of the rationals is not complete, because the set*

$$X = \{r \mid r^2 < 2\}$$

is bounded from above but has no least upper bound.

A.25. Lemma. *Every complete, ordered field F has the **archimedean property***

$$(\forall x \in F)(\exists n \in \mathbb{N}_F)[x < n],$$

i.e., the set $\mathbb{N} = \mathbb{N}_F$ of its natural numbers is not bounded from above.

PROOF. Assume towards a contradiction that the set $\mathbb{N}$ has an upper bound, so that it has a least upper bound $x = \sup \mathbb{N}$ by the completeness property. The element $x - 1$ is not an upper bound of $\mathbb{N}$ because $x - 1 < x$, so there must exist some $n \in \mathbb{N}$, $x - 1 < n$: but this implies $x < n + 1$ which contradicts the assumption that x is an upper bound of $\mathbb{N}$.   ⊣

A.26. Exercise. *In every complete, ordered field F,*

$$\varepsilon > 0 \Longrightarrow (\exists n \in \mathbb{N}) \left[\frac{1}{n+1} < \varepsilon \right].$$

A.27. Exercise (Density of the rationals). *In every complete, ordered field F,*

$$x < y \Longrightarrow (\exists r \in \mathbb{Q})[x < r \,\&\, r < y],$$

where $\mathbb{Q} = \mathbb{Q}_F$ is the set of rationals in F.

We now aim to show that *there exists a complete ordered field* and that *any two complete, ordered fields are isomorphic*. Since the completeness property is geometric (or topological), these proofs of existence and uniqueness depend on geometric ideas. Specifically, we will need some basic definitions and results from the *theory of limits* which is studied in Calculus. We will review these here, briefly but without limiting ourselves to the absolute minimum list of theorems necessary to prove the existence and uniqueness of the reals: we have included several Lemmas and Exercises because they support the proposition that the notion of a complete, ordered field represents faithfully our geometric intuitions about the real numbers.

Since we will be manipulating (infinite) sequences a great deal, we will use consistently the familiar, simple notation which avoids the introduction of a distinct name for each sequence and treats the variable as an index, so that a sequence $(n \mapsto a_n)$ is denoted by $\langle a_n \rangle$ or $\langle a_0, a_1, \dots \rangle$,. For example, the sequence

$$\left\langle \frac{1}{n+1} \right\rangle = \left\langle \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \dots \right\rangle$$

is the function $f : \mathbb{N} \to \mathbb{Q}$ defined by the formula

$$f(n) = \frac{1}{n+1}.$$

The **absolute value** function is defined in every ordered field in the usual way,

$$|x| =_{\mathrm{df}} \max\{x, -x\} = \begin{cases} x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

A.28. Definition. Suppose $(F, 0, 1, +, \cdot, \leq)$ is an ordered field.
(1) A sequence $\langle x_n \rangle$ of elements of F **converges to** $x \in F$ or **has limit** x, if

$$(\forall \varepsilon \in F, \varepsilon > 0)(\exists K \in \mathbb{N})(\forall n \in \mathbb{N})[n \geq K \Longrightarrow |x - x_n| < \varepsilon].$$

We will use the notation

$$x_n \to x \iff_{\mathrm{df}} \langle x_n \rangle \text{ converges to } x.$$

(2) A sequence $\langle x_n \rangle$ has the property of Cauchy, or (simply) **is Cauchy** if

$$(\forall \varepsilon \in F, \varepsilon > 0)(\exists K \in \mathbb{N})(\forall n, m \in \mathbb{N})[n, m \geq K \Longrightarrow |x_n - x_m| < \varepsilon].$$

A.29. Definition. For all $a < b$ in an ordered field F, the set

$$(a, b) =_{\mathrm{df}} \{x \in F \mid a < x < b\} \tag{A-7}$$

is the **open interval** with endpoints a and b. A set $G \subseteq F$ is **open** if it is a (possibly empty) union of open intervals, equivalently

$$x \in G \Longrightarrow (\exists a < b)[x \in (a, b) \subseteq G].$$

We will also use the standard notations for closed and half-open intervals, e.g.,

$$(a, b] = \{x \in F \mid a < x \leq b\},$$

and also for the half-intervals with $-\infty$ or ∞ as one or both of their endpoints, e.g.,

$$(-\infty, b) = \{x \in F \mid x < b\}, \quad [a, \infty) = \{x \in F \mid a \leq x\}.$$

A.30. Exercise. *Prove that the family of open sets in an ordered field is a topology and that the definition of limits for sequences in* **A.28** *is equivalent to the topological definition of limits given in* **10.42**.

These definitions are notorious for the difficulty of understanding what they mean and learning how to use them. We emphasize that here we study them in the context of an arbitrary ordered field which need not be complete, for example, the rationals. It is useful to formulate conditions equivalent to convergence and the property of Cauchy, on the basis of the following notion.

A.31. Definition. A sequence $\langle x_n \rangle$ in an ordered field **settles** in an open interval (a, b), if after a certain stage all its terms belong to some closed subinterval $[a', b'] \subseteq (a, b)$:

$$\langle x_n \rangle \rightsquigarrow (a, b) \iff_{\mathrm{df}} (\exists K, a', b')(\forall n \geq K)[a < a' \leq x_n \leq b' < b].$$

Notice that if $\langle x_n \rangle \rightsquigarrow (a, b)$, then all its terms after a certain stage belong to (a, b),

$$\langle x_n \rangle \rightsquigarrow (a, b) \Longrightarrow (\exists K)(\forall n \geq K)[a < x_n < b];$$

this weaker property is all we need for many applications of the definition of $\langle x_n \rangle \rightsquigarrow (a, b)$.

A.32. Exercise. $\langle x_n \rangle \rightsquigarrow (a, b)$ *if and only if there exists some $\delta > 0$ such that $a + \delta < b - \delta$ and the set $\{n \in \mathbb{N} \mid x_n \notin [a + \delta, b - \delta]\}$ is finite.*

A.33. Exercise. *For all open intervals I, J, $\langle x_n \rangle \rightsquigarrow I \,\&\, I \subseteq J \Longrightarrow \langle x_n \rangle \rightsquigarrow J$.*

A.34. Exercise. *For all open intervals I, J,*

$$\langle x_n \rangle \rightsquigarrow I \,\&\, \langle x_n \rangle \rightsquigarrow J \Longrightarrow \langle x_n \rangle \rightsquigarrow I \cap J \Longrightarrow I \cap J \neq \emptyset.$$

A.35. Exercise. *Every sequence $\langle x_n \rangle$ which settles in some open interval (a, b) is **bounded**, i.e., $(\exists w)(\forall n)[x_n \leq w]$.*

The next Lemma makes it possible in many cases to avoid the so-called "method of epsilonics", which is illustrated by its proof.

A.36. Lemma. *For every sequence in any ordered field F:*

(1) *$\langle x_n \rangle$ converges to x if and only if $\langle x_n \rangle$ settles in every open interval which contains x,*

$$x_n \to x \iff (\forall a, b \in F)[a < x < b \Longrightarrow \langle x_n \rangle \rightsquigarrow (a, b)].$$

(2) *$\langle x_n \rangle$ is Cauchy if and only if for every $\varepsilon > 0$, there exists an open interval (a, b) in which $\langle x_n \rangle$ settles and such that $(b - a) \leq \varepsilon$:*

$$\langle x_n \rangle \text{ is Cauchy} \iff (\forall \varepsilon > 0)(\exists a, b)[a < b \leq a + \varepsilon \,\&\, \langle x_n \rangle \rightsquigarrow (a, b)]$$

PROOF. (1) If $x_n \to x$ and $a < x < b$, then the definition of convergence with

$$\varepsilon = \frac{\min(x - a, b - x)}{2}$$

supplies a number K such that

$$n \geq K \Longrightarrow |x - x_n| < \frac{\min(x - a, b - x)}{2},$$

which with a bit of inequality massaging implies that

$$n \geq K \Longrightarrow a < a' = \frac{a + x}{2} < x_n < \frac{x + b}{2} = b' < b,$$

so that $\langle x_n \rangle \rightsquigarrow (a, b)$. For the proof in the other direction: for every $\varepsilon > 0$, $\langle x_n \rangle \rightsquigarrow (x - \varepsilon, x + \varepsilon)$, so that for some K,

$$n \geq K \Longrightarrow |x - x_n| < \varepsilon.$$

(2) If $\langle x_n \rangle$ is Cauchy, then for every $\varepsilon > 0$ there exists some K such that $n, m \geq K \Longrightarrow |x_n - x_m| < \frac{\varepsilon}{4}$, which immediately implies that

$$\langle x_n \rangle \rightsquigarrow (x_K - \frac{\varepsilon}{2}, x_K + \frac{\varepsilon}{2}),$$

and this interval has length ε. In the other direction, for every $\varepsilon > 0$ there exists some (a, b) with $(b - a) \leq \varepsilon$, so that $\langle x_n \rangle \rightsquigarrow (a, b)$, and so for some K,

$$n, m \geq K \Longrightarrow [a < x_n < b \,\&\, a < x_m < b \Longrightarrow |x_n - x_m| < b - a \leq \varepsilon],$$

which means that $\langle x_n \rangle$ is Cauchy. $\dashv$

A.37. Corollary. *If $\langle x_n \rangle$ converges to some x, then it is Cauchy.*

PROOF. For every $\varepsilon > 0$, $\langle x_n \rangle \rightsquigarrow (x - \frac{\varepsilon}{2}, x + \frac{\varepsilon}{2})$ by (1) of **A.36**, so it is Cauchy by (2) of **A.36**. $\dashv$

A.38. Exercise. $x_n \to x \,\&\, x_n \to y \Longrightarrow x = y$. *This allows us to introduce the classical notation*

$$x = \lim_n x_n \iff_{\mathrm{df}} x_n \to x.$$

A.39. Lemma. *If $\langle x_n \rangle$ is Cauchy in a complete, ordered field F, then $\langle x_n \rangle$ has a limit, i.e., $x_n \to x$ with some x.*

PROOF. Let

$$X =_{\mathrm{df}} \{u \in F \mid (\exists v)[u < v \,\&\, \langle x_n \rangle \rightsquigarrow (u, v)]\}.$$

Since $\langle x_n \rangle$ is Cauchy, there exists some (c, d) such that $\langle x_n \rangle \rightsquigarrow (c, d)$, and hence

$$u \in X \Longrightarrow u \leq d,$$

since $d < u \Longrightarrow \langle x_n \rangle \rightsquigarrow (c, d) \cap (u, v) = \emptyset$ which is not possible; hence X is bounded from above and it must have a least upper bound

$$x = \sup X.$$

We will show that

$$a < x < b \Longrightarrow \langle x_n \rangle \rightsquigarrow (a, b)$$

which implies $x_n \to x$ by **A.36**. Using the hypothesis $a < x < b$ and the fact that $\langle x_n \rangle$ is Cauchy, we find first some (u, v) such that

$$v - u < \min(x - a, b - x), \quad \langle x_n \rangle \rightsquigarrow (u, v). \tag{A-8}$$

By definition, $u \in X$, and hence (1) $u \leq x$, because x is an upper bound of X. On the other hand, v is also an upper bound of X (because the assumption $\langle x_n \rangle \rightsquigarrow (u', v')$ with $v < u'$ implies that $\langle x_n \rangle$ settles in two disjoint intervals), and hence (2) $x \leq v$, since x is the least upper bound of X. Now (1) and (2) together yield $u \leq x \leq v$, which together with (A-8) implies $a < u \leq x \leq v < b$, so that $\langle x_n \rangle \rightsquigarrow (a, b)$. $\dashv$

The next two basic theorems relate completeness as we defined it (following Dedekind) with the notion of completeness historically associated with the name of Cantor.

A.40. Theorem (The nested interval property). *Suppose that every Cauchy sequence in an ordered field F converges, and that*

$$[x_0, y_0] \supseteq [x_1, y_1] \supseteq \cdots \tag{A-9}$$

is a nested sequence of closed intervals such that

$$\lim_n (y_n - x_n) = 0; \tag{A-10}$$

it follows that the intersection $\bigcap_n[x_n, y_n]$ is a singleton

$$\bigcap_n[x_n, y_n] = \{w\},$$

and its only member is the common limit of the sequences $\langle x_n \rangle$ and $\langle y_n \rangle$,

$$w = \lim_n x_n = \lim_n y_n.$$

Proof. The basic observation is that for every number K and every $\delta > 0$, $\langle x_n \rangle \rightsquigarrow [x_K - \delta, y_K + \delta]$ by (A-9). Now (A-10) implies that $\langle x_n \rangle$ is Cauchy and hence $x_n \to x$ for some x, using the hypothesis. Moreover, $x < x_K \Longrightarrow \langle x_n \rangle \rightsquigarrow (x - 1, x_K)$ by **A.36**, which contradicts the basic properties of the relation $\rightsquigarrow$ since $(x - 1, x_K) \cap [x_K, y_K] = \emptyset$, hence, $x_K \leq x$, for every K. By a similar argument $x \leq y_K$, for every K, so that in the end $x \in \bigcap_n[x_n, y_n]$. Symmetrically, $\langle y_n \rangle$ converges, $y = \lim_n y_n \in \bigcap_n[x_n, y_n]$, and for every n, $|x - y| \leq (y_n - x_n)$ which implies $x = y$ by (A-10).   ⊣

A.41. Theorem. *An ordered field F is complete if and only it has the archimedean property (**A.25**) and every Cauchy sequence in F has a limit.*

Proof. One direction is known from Lemmas **A.25** and **A.39**, so it is enough to show that if F has the archimedean property and every Cauchy sequence in F converges, then F is complete.

Suppose then that X is a non-empty, bounded from above set in F, so that there exists some point $x_0 \in X$ and some upper bound y_0 of X. Beginning with $[x_0, y_0]$, we define by recursion a sequence of closed intervals $[x_n, y_n]$ which satisfy the following conditions:

1. $x_n \leq x_{n+1} < y_{n+1} \leq y_n$,
2. $(y_n - x_n) = 2^{-n}(y_0 - x_0)$,
3. $[x_n, y_n] \cap X \neq \emptyset$,
4. $(\forall x \in X)[x \leq y_n]$.

In detail, to define $[x_{n+1}, y_{n+1}]$ we distinguish two cases: if $w = \frac{1}{2}(x_n + y_n)$ is an upper bound of X, we set $[x_{n+1}, y_{n+1}] = [x_n, w]$, otherwise $[x_{n+1}, y_{n+1}] = [w, y_{n+1}]$. Proof that $[x_{n+1}, y_{n+1}]$ satisfies (1) – (4) is trivial.

Lemma. $\lim_n(y_n - x_n) = 0$.

Proof. The archimedean property implies that for every $\varepsilon > 0$, there exists some natural number $K > 0$ such that

$$\frac{y_0 - x_0}{\varepsilon} < K \leq 2^K,$$

where the inequality $K \leq 2^K$ is verified easily (by induction!). It follows that for every $n \geq K$,

$$(y_n - x_n) = 2^{-n}(y_0 - x_0) \leq 2^{-K}(y_0 - x_0) < \varepsilon, \qquad \dashv \text{(Lemma)}$$

Now **A.40** implies that

$$\bigcap[x_n, y_n] = \{w\}$$

where $w = \lim_n x_n = \lim_n y_n$, and it is enough to verify that this common limit w is the least upper bound of X. We compute:

$$
\begin{aligned}
w < t &\implies \langle y_n \rangle \rightsquigarrow (w - 1, t) && \text{because } \lim_n y_n = w, \\
&\implies y_n < t && \text{for some } n, \\
&\implies t \notin X && \text{because } y_n \text{ is an upper bound of } X,
\end{aligned}
$$

so that w is an upper bound of X. Also,

$$
\begin{aligned}
t < w &\implies \langle x_n \rangle \rightsquigarrow (t, w + 1) && \text{because } \lim_n x_n = w, \\
&\implies t < x_n && \text{for some } n, \\
&\implies (\exists x \in X)[t < x] && \text{by the definition of } x_n,
\end{aligned}
$$

so that there is no upper bound of X smaller than w. $\dashv$

It is worth pointing out here that there exist *Cauchy complete* ordered fields which are not archimedean, Problem **xA.2**.

Following Cantor (up to a point), we will construct a complete, ordered field as a quotient of the Cauchy sequences on the rationals by the following, natural equivalence relation.

A.42. Definition. We call two sequences of rationals $\langle x_n \rangle$ and $\langle y_n \rangle$ **asymptotically equivalent** if their difference converges to 0, in symbols,

$$
\langle x_n \rangle \approx \langle y_n \rangle \iff_{\mathrm{df}} (x_n - y_n) \to 0.
$$

A.43. Theorem. (1) *Two Cauchy sequences $\langle x_n \rangle$ and $\langle y_n \rangle$ are asymptotically equivalent if and only if they settle in the same open intervals*:

$$
\langle x_n \rangle \approx \langle y_n \rangle \iff (\forall a < b)[\langle x_n \rangle \rightsquigarrow (a, b) \iff \langle y_n \rangle \rightsquigarrow (a, b)].
$$

(2) *If $\langle x_n \rangle$ and $\langle y_n \rangle$ are Cauchy, then*

$$
\begin{aligned}
\langle x_n \rangle \not\approx \langle y_n \rangle \implies &\ (\text{there exist open intervals } I, J) \\
&[I \cap J = \emptyset \ \& \ \langle x_n \rangle \rightsquigarrow I \ \& \ \langle y_n \rangle \rightsquigarrow J].
\end{aligned}
$$

(3) *The relation $\approx$ is an equivalence relation on the set*

$$
\mathscr{C}(F) =_{\mathrm{df}} \{ \langle x_n \rangle \mid \langle x_n \rangle \text{ is Cauchy on } F \}.
$$

PROOF. (1) Suppose first that $\langle x_n \rangle \approx \langle y_n \rangle$ and $\langle x_n \rangle \rightsquigarrow (a, b)$, so that for some K_0 and some $\delta > 0$,

$$
n \geq K_0 \implies a + \delta \leq x_n \leq b - \delta.
$$

Using $\langle x_n \rangle \approx \langle y_n \rangle$, choose K_1 such that

$$
n \geq K_1 \implies |x_n - y_n| < \frac{\delta}{2}
$$

$$
\implies x_n - \frac{\delta}{2} < y_n < x_n + \frac{\delta}{2}.
$$

From these two implications we get easily that

$$
n \geq \max(K_0, K_1) \implies a + \frac{\delta}{2} \leq y_n \leq b - \frac{\delta}{2},
$$

so that $\langle y_n \rangle \rightsquigarrow (a, b)$. In the other direction, for every $\varepsilon > 0$ there exists an open interval (a, b) with $b - a \leq \varepsilon$ such that $\langle x_n \rangle \rightsquigarrow (a, b)$ by (2) of **A.36**, so by the hypothesis, we also have

$$n \geq K \Longrightarrow [a < x_n < b \,\&\, a < y_n < b] \Longrightarrow |y_n - x_n| < \varepsilon.$$

(2) Directly from the definition of convergence,

$$\neg[(x_n - y_n) \to 0] \Longrightarrow (\exists \varepsilon > 0)(\forall K)(\exists n, m \geq K)[|x_n - y_m| \geq \varepsilon]. \quad \text{(A-11)}$$

By **A.36**, there exist open intervals I and J of length $< \frac{\varepsilon}{2}$ (with this ε) such that $\langle x_n \rangle \rightsquigarrow I$ and $\langle y_n \rangle \rightsquigarrow J$. If some $z \in I \cap J$ existed, then for K sufficiently large so that

$$n, m \geq K \Longrightarrow x_n \in I \,\&\, y_m \in J,$$

we would have

$$n, m \geq K \Longrightarrow |x_n - y_m| \leq |x_n - z| + |z - y_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

which contradicts (A-11).

(3) The reflexivity and symmetry of $\approx$ are trivial. To show its transitivity, notice that by hypothesis and (1) of Lemma **A.36**, for every (a, b),

$$\langle x_n \rangle \rightsquigarrow (a, b) \iff \langle y_n \rangle \rightsquigarrow (a, b) \iff \langle z_n \rangle \rightsquigarrow (a, b),$$

so that by (1), $\langle x_n \rangle \approx \langle z_n \rangle$. $\dashv$

A.44. Exercise. *If $\langle x_n \rangle$ and $\langle y_n \rangle$ are both Cauchy and $x_n \to x$, then*

$$\langle x_n \rangle \approx \langle y_n \rangle \iff y_n \to x.$$

At this point we could appeal to the existence of *some quotient B* of the set

$$\mathscr{C} = \mathscr{C}(\mathbb{Q}) =_{\mathrm{df}} \{\langle r_n \rangle \mid \langle r_n \rangle \text{ is Cauchy in the field of rationals}\}$$

by $\approx$ and define the necessary functions and an ordering on B so that it becomes a complete, ordered field. This is one of the classical proofs of the existence of the real numbers, connected with the name of Cantor. Instead of this, we will construct *a specific quotient* of $\mathscr{C}$ by $\approx$ which simplifies the proof a bit and (more significantly) relates this construction with the other classical proof of the existence of the reals, following Dedekind. The basic idea of Dedekind was that a real number x is completely determined by (and hence can be "identified" with) the set

$$(-\infty, x) \cap \mathbb{Q} =_{\mathrm{df}} \{r \in \mathbb{Q} \mid r < x\}$$

of all rationals preceding it, and that the sets of the form $(-\infty, x) \cap \mathbb{Q}$ can be characterized directly by three simple conditions.

A.45. Definition. A **Dedekind cut** is any set X of rational numbers which satisfies the following three conditions:

    1. $X \neq \emptyset, (\mathbb{Q} \setminus X) \neq \emptyset$.
    2. $r < q \,\&\, q \in X \Longrightarrow r \in X$.
    3. $q \in X \Longrightarrow (\exists r)[q < r \,\&\, r \in X]$.

We set

$$\mathscr{D} =_{df} \{X \subseteq \mathbb{Q} \mid X \text{ is a Dedekind cut}\}.$$

A.46. Exercise. *A set $X \subseteq \mathbb{Q}$ is a Dedekind cut if and only if it is non-empty, bounded from above, with no largest member and "downward closed", i.e., such that $r < q \,\&\, q \in X \Longrightarrow r \in X$.*

The next theorem is basic for the proof of existence of the real numbers.

A.47. Theorem. *For each Cauchy sequence of rationals $\langle x_n \rangle$, let*

$$\pi(\langle x_n \rangle) =_{df} \{a \in \mathbb{Q} \mid (\exists b)[a < b \,\&\, \langle x_n \rangle \rightsquigarrow (a, b)]\};$$

it follows that each value $\pi(\langle x_n \rangle)$ is a Dedekind cut and that the function

$$\pi : \mathscr{C} \twoheadrightarrow \mathscr{D}$$

is a surjection which determines the equivalence relation $\approx$, so that $\mathscr{D}$ is a quotient of $\mathscr{C}$.

PROOF. That each $\pi(\langle x_n \rangle)$ is a Dedekind cut is quite easy from the definitions and the general properties of the relation $\rightsquigarrow$, and the equivalence

$$\langle x_n \rangle \approx \langle y_n \rangle \iff \pi(\langle x_n \rangle) = \pi(\langle y_n \rangle)$$

is an immediate corollary of **A.43**. The only thing which is not completely obvious is that for every cut X there exists a Cauchy sequence $\langle x_n \rangle \in \mathscr{C}$ in the rationals such that $\pi(\langle x_n \rangle) = X$. For this we construct a nested sequence of closed intervals *in the rationals*

$$[x_0, y_0] \supseteq [x_1, y_1] \supseteq \cdots$$

exactly as in the proof of **A.41**, beginning with some $x_0 \in X$ and some $y_0 \notin X$, so that, in fact, y_0 is an upper bound of X. We argue as in **A.41** that the non-decreasing sequence $\langle x_n \rangle$ is Cauchy, and that, in addition, for all n,

$$x_n \in X \,\&\, y_n \notin X,$$

because X is downward closed and has no largest member. Then we compute:

$$\begin{aligned}
a \in \pi(\langle x_n \rangle) &\implies (\exists b)[a < b \,\&\, \langle x_n \rangle \rightsquigarrow (a, b)] \\
&\implies (\exists n)[a < x_n] \\
&\implies a \in X \text{ because } x_n \in X.
\end{aligned}$$

To see that $X \subseteq \pi(\langle x_n \rangle)$, notice that if $a \in X$, then there must exist some natural number K such that $a < x_K$, because the opposite supposition $(\forall n)[x_n \leq a]$ implies (easily) that a is the largest point in X, and X does not have a largest point. Thus, for $n \geq K$, $x_K \leq x_n \leq y_K$, hence, $\langle x_n \rangle \rightsquigarrow (x_K, y_K + 1)$ and $a \in \pi(\langle x_n \rangle)$.   ⊣

A.48. Theorem (Existence of the real numbers). *There exists a complete, ordered field.*

PROOF. For the domain of the required field we take the set $\mathscr{D}$ of Dedekind cuts and for 0 and 1 we take the obvious:

$$\begin{aligned}
0 &=_{df} \{r \in \mathbb{Q} \mid r < 0\} = \pi(\langle 0 \rangle), \\
1 &=_{df} \{r \in \mathbb{Q} \mid r < 1\} = \pi(\langle 1 \rangle).
\end{aligned}$$

In order to define addition and multiplication on $\mathscr{D}$ we need the following two Lemmas, where all the sequences are Cauchy in the rationals.

$$\langle x_n \rangle \approx \langle x'_n \rangle \;\&\; \langle y_n \rangle \approx \langle y'_n \rangle \implies \langle x_n + y_n \rangle \approx \langle x'_n + y'_n \rangle \qquad \text{(A-12)}$$
$$\langle x_n \rangle \approx \langle x'_n \rangle \;\&\; \langle y_n \rangle \approx \langle y'_n \rangle \implies \langle x_n \cdot y_n \rangle \approx \langle x'_n \cdot y'_n \rangle \qquad \text{(A-13)}$$

These are the useful interpretations of the classical theorems from the theory of limits,

$$\lim_n (x_n + y_n) = \lim_n x_n + \lim_n y_n,$$
$$\lim_n (x_n \cdot y_n) = \lim_n x_n \cdot \lim_n y_n,$$

in the case at hand, when the limits need not exist since the sequences are in the incomplete ordered field of the rationals. They are not hard to verify after all the preparatory work we have done and we will skip the details. The implications (A-12) and (A-13) assert that $\approx$ is a congruence in $\mathscr{C}$ for the functions

$$(\langle x_n \rangle, \langle y_n \rangle) \mapsto \langle x_n + y_n \rangle,$$
$$(\langle x_n \rangle, \langle y_n \rangle) \mapsto \langle x_n \cdot y_n \rangle,$$

so that by **A.2** there exist functions $+$ and $\cdot$ on the quotient $\mathscr{D}$ which satisfy the identities

$$\pi(\langle x_n \rangle) + \pi(\langle y_n \rangle) = \pi(\langle x_n + y_n \rangle),$$
$$\pi(\langle x_n \rangle) \cdot \pi(\langle y_n \rangle) = \pi(\langle x_n \cdot y_n \rangle).$$

We take these $+$ and $\cdot$ for the operations of addition and multiplication in $\mathscr{D}$.

Next we must show that $(\mathscr{D}, 0, 1, +, \cdot)$ is a field, but this part of the proof is quite trivial, if a bit tiring in its details (which we will skip). The existence of additive inverses, for example, follows from the obvious

$$\pi(\langle x_n \rangle) + \pi(\langle -x_n \rangle) = \pi(\langle x_n + (-x_n) \rangle) = \pi(\langle 0 \rangle) = 0,$$

where the only "delicate" point is the observation that if $\langle x_n \rangle$ is Cauchy, then so is $\langle -x_n \rangle$. To check the corresponding property for multiplication, given a Cauchy sequence $\langle x_n \rangle \not\approx \langle 0 \rangle$, set

$$y_n =_{\mathrm{df}} \begin{cases} 1/x_n, & \text{if } x_n \neq 0, \\ 1, & \text{if } x_n = 0, \end{cases} \qquad \text{(A-14)}$$

verify that $\langle y_n \rangle$ is also Cauchy and then compute:

$$\pi(\langle x_n \rangle) \cdot \pi(\langle y_n \rangle) = \pi(\langle x_n \cdot y_n \rangle) = \pi(\langle 1 \rangle) = 1.$$

The basic observation (from (2) of **A.43**) is that

$$\langle x_n \rangle \not\approx \langle 0 \rangle \implies (\exists \delta > 0)[\langle x_n \rangle \not\rightarrow (-\delta, \delta)]$$
$$\implies (\exists \delta > 0, K)(\forall n \geq K)|x_n| \geq \delta$$

with which we begin the proof that $\langle y_n \rangle$ is Cauchy, but some epsilonics are unavoidable. The related result from the theory of limits is the assertion

$$\lim_n x_n \neq 0 \Longrightarrow \lim_n \left(\frac{1}{x_n} \right) = \frac{1}{\lim_n x_n},$$

traditionally known as the first hard theorem in Calculus, when it is taught rigorously.

Next we define on $\mathscr{D}$ the relation

$$X \leq Y \iff_{\text{df}} X \subseteq Y,$$

which is certainly a partial ordering; it is also linear, because for any two Dedekind cuts X and Y, directly from the definition,

$$r \in (Y \setminus X) \implies (\forall q \in X)[q < r] \,\&\, (\forall q < r)[q \in Y]$$
$$\implies X \subseteq Y,$$

and, of course, $X \neq Y \implies (\exists r)[[r \in (X \setminus Y)] \vee [r \in (Y \setminus X)]]$. Appealing once more to the definition of Dedekind cuts, we can also show easily that

$$I \subseteq \mathscr{D} \implies \bigcup I = \mathbb{Q} \vee \bigcup I \in \mathscr{D}. \tag{A-15}$$

(For example, if r were the largest point in the union $\bigcup I$, then $r \in X$ for some $X \in I$, and then r would also be the largest point in X, since $X \subseteq I$—but X has no largest member.) From (A-15) we infer that every set $I \subseteq \mathscr{D}$ which is bounded above has a least upper bound, because

$$(\forall X \in I)[X \leq Z] \implies \bigcup I \subseteq Z \subsetneq \mathbb{Q} \implies \bigcup I \in \mathscr{D},$$

and the union $\bigcup I$ is obviously the least upper bound of I in the relation $\leq = \subseteq$.

It remains to verify that for all $X, Y, Z \in \mathscr{D}$

$$X < Y \implies X + Z < Y + Z, \tag{A-16}$$
$$[Z > 0 \,\&\, X < Y] \implies Z \cdot X < Z \cdot Y, \tag{A-17}$$

since these implications imply then immediately their versions with $\leq$. Considering the more difficult (A-17) as an example, choose first Cauchy sequences such that

$$\pi(\langle z_n \rangle) = Z, \ \pi(\langle x_n \rangle) = X, \ \pi(\langle y_n \rangle) = Y,$$

and verify (easily) from the definitions (and **A.43**) that there exist rationals

$$x^0 < x^1 < y^0 < y^1$$

satisfying

$$\langle x_n \rangle \rightsquigarrow (x^0, x^1), \quad \langle y_n \rangle \rightsquigarrow (y^0, y^1),$$

and that for each $\varepsilon > 0$ we can find some z^0 and z^1 such that

$$0 < z^0 < z^1, \quad (z^1 - z^0) < \varepsilon, \quad \langle z_n \rangle \rightsquigarrow (z^0, z^1).$$

It follows that

$$\langle z_n x_n \rangle \rightsquigarrow (z^0 x^0, z^1 x^1), \quad \langle z_n y_n \rangle \rightsquigarrow (z^0 y^0, z^1 y^1),$$

and the desired conclusion $Z \cdot X < Z \cdot Y$ will follow quite easily, if we could choose z^0, z^1 so that

$$z^1 x^1 < z^0 y^0,$$

or equivalently,

$$x^1(z^1 - z^0) < z^0(y^0 - x^1). \tag{A-18}$$

Now (A-18) is obvious if $x^1 \leq 0$, because in that case $x^1(z^1 - z^0) \leq 0$ and $z^0(y^0 - x^1) > 0$. If $x^1 > 0$, we find first some $\delta > 0$ such that $\langle z_n \rangle \rightsquigarrow (\delta, \infty)$ and then z^0, z^1 such that

$$0 < \delta < z^0 < z^1, \quad \langle z_n \rangle \rightsquigarrow (z^0, z^1), \quad (z^1 - z^0) < \frac{\delta(y^0 - x^1)}{x^1}$$

which imply (A-18):

$$x^1(z^1 - z^0) < x^1 \frac{\delta(y^0 - x^1)}{x^1} \leq z^0(y^0 - x^1).$$

Verification of (A-16) is substantially simpler and completes the proof of the theorem. ⊣

A.49. Exercise. *Prove that for all Dedekind cuts X, Y and Z:*

$$X \cdot (Y + Z) = X \cdot Y + X \cdot Z.$$

(*Use the formal definitions of $+$ and $\cdot$ given in the proof of* **A.48**.)

A.50. Exercise. *Prove* (A-12) *and* (A-13).

A.51. Exercise. *Show that if $\langle x_n \rangle$ is Cauchy in an ordered field and $\langle x_n \rangle \not\approx \langle 0 \rangle$, then the sequence $\langle y_n \rangle$ defined by* (A-14) *is also Cauchy and $\langle x_n y_n \rangle \approx \langle 1 \rangle$.*

A.52. Theorem (Uniqueness of the real numbers). *For any two complete, ordered fields F_1 and F_2, there exists exactly one bijection*

$$\pi^* : F_1 \rightarrowtail\!\!\!\rightarrow F_2$$

which is an isomorphism, i.e.,

1. $\pi^*(0) = 0, \quad \pi^*(1) = 1,$
2. $\pi^*(x + y) = \pi^*(x) + \pi^*(y), \quad \pi^*(xy) = \pi^*(x)\pi^*(y),$
3. $x \leq y \iff \pi^*(x) \leq \pi^*(y).$

Proof. By the uniqueness of the rationals **A.14**, there exists (exactly one) isomorphism

$$\pi : \mathbb{Q}_1 \rightarrowtail\!\!\!\rightarrow \mathbb{Q}_2,$$

where $\mathbb{Q}_1, \mathbb{Q}_2$ are the sets of rationals in the two fields F_1, F_2, and the problem is to extend this π to the whole of F_1.

Lemma 1. *For each $x \in F_1$, there exists a sequence $\langle x_n \rangle$ of rationals in F_1, such that $\lim_n x_n = x$.*

Proof. Using the density of the rationals (Exercise **A.27**), we can find for each $n \in \mathbb{N}$ a rational $x_n \in \mathbb{Q}_1$ such that $|x - x_n| < 1/(n+1)$, and then (easily, using problem **A.26**) $\lim_n x_n = x$. ⊣ (Lemma 1)

Lemma 2. *For each sequence $\langle x_n \rangle$ of rationals in F_1,*

$$(\exists x \in F_1)[\lim_n = x] \Longrightarrow (\exists x^* \in F_2)[\lim_n \pi(x_n) = x^*].$$

Proof. We know that

$$u < v \iff \pi(u) < \pi(v) \quad (u, v \in \mathbb{Q}_1)$$

because π is an isomorphism, so that for all $a, b \in \mathbb{Q}_1$, $a < b$,

$$\langle x_n \rangle \rightsquigarrow (a, b) \iff \langle \pi(x_n) \rangle \rightsquigarrow (\pi(a), \pi(b)). \tag{A-19}$$

If $\langle x_n \rangle$ converges, then it is Cauchy, so that $\langle \pi(x_n) \rangle$ is also Cauchy by (A-19) and **A.36** (using **A.26** once more), and therefore $\langle \pi(x_n) \rangle$ converges because F_2 is complete. $\dashv$ (Lemma 2)

We can use the same simple idea to verify the third basic fact we need, whose proof we will omit:

Lemma 3. *For any two Cauchy sequences in the rationals of F_1,*

$$[\lim_n x_n = \lim_n y_n] \Longrightarrow \lim_n \pi(x_n) = \lim_n \pi(y_n).$$

The Lemmas guarantee that we can define unambiguously for each $x \in F_1$,

$$\pi^*(x) =_{\mathrm{df}} \lim_n \pi(x_n), \quad \text{where } \lim_n x_n = x, \text{ with } x_n \in \mathbb{Q}_1. \tag{A-20}$$

Since for each rational $x \in \mathbb{Q}_1$, $\lim_n x = x$, we have

$$\pi^*(x) = \lim_n \pi(x) = \pi(x),$$

so that π^* is an extension of π. It remains to verify that π^* is an isomorphism of F_1 with F_2.

Suppose first that

$$x = \lim_n x_n, \quad y = \lim_n y_n, \quad x_n, y_n \in \mathbb{Q}_1, \quad x < y.$$

This implies immediately (by **A.36**) that there exist rationals a, b, c and d satisfying

$$\langle x_n \rangle \rightsquigarrow (a, b), \quad \langle y_n \rangle \rightsquigarrow (c, d), \quad b < c,$$

and hence by (A-19)

$$\langle \pi(x_n) \rangle \rightsquigarrow (\pi(a), \pi(b)), \quad \langle \pi(y_n) \rangle \rightsquigarrow (\pi(c), \pi(d)).$$

It follows that

$$\pi^*(x) = \lim_n \pi(x_n) < \lim_n \pi(y_n) = \pi^*(y)$$

because $\pi(b) < \pi(c)$, and this completes the proof that π^* respects the strict relation $<$:

$$x < y \Longrightarrow \pi^*(x) < \pi^*(y).$$

Directly from this, π^* is an injection and it respects the ordering,

$$x \leq y \iff \pi^*(x) \leq \pi^*(y).$$

The rest is trivial (if tiresome) and follows mostly from the *limit theorems* of the Calculus, which hold in every complete, ordered field—and can be proved easily with the tools we have developed. As an example:

$$\pi^*(x + y) =_{\text{df}} \lim_n \pi(x_n + y_n), \quad \text{where } x_n \to x, y_n \to y, x_n, y_n \in \mathbb{Q}_1,$$
$$= \lim_n [\pi(x_n) + \pi(y_n)]$$
$$= \lim_n \pi(x_n) + \lim_n \pi(y_n)$$
$$= \pi^*(x) + \pi^*(y).$$

The crucial step in this computation is the identity

$$\lim_n [\pi(x_n) + \pi(y_n)] = \lim_n \pi(x_n) + \lim_n \pi(y_n).$$

We skip the details.   ⊣

A.53. Exercise. *Work out the details of the proof of*

$$\pi^*(x + y) = \pi^*(x) + \pi^*(y).$$

A.54. The real numbers. As we did for the natural numbers and the rationals, we now fix some complete, ordered field

$$(\mathbb{R}, 0, 1, +, \cdot, \leq), \tag{A-21}$$

whose members we will call **real numbers**. We emphasize once more that fixing some $\mathbb{R}$ is only a convenience and the specific choice of $\mathbb{R}$ is of no importance: the fundamental mathematical fact for the development of analysis is that a complete, ordered field exists and that any two complete, ordered fields are isomorphic.

A.55. Exercise. *Prove Corollary* **2.14** *in Chapter* **2** *from the axioms.*

The open sets of real numbers defined in **A.29** form a topology by the easy Exercise **A.30**, so we have notions of Borel sets of reals and Borel measurable functions from $\mathbb{R}$ to other topological spaces and vice versa, by Definitions **10.25** and **10.39**. The notion of Borel isomorphism was defined in Definition **10.40**. The next theorem makes it possible to transfer results about Baire space to the reals, and it is the main tool for analyzing the set theoretic properties of $\mathbb{R}$. We will omit its proof, which is quite simple, using Problems **x10.12** and **x10.13**.

A.56. Theorem. *As a topological space,* $\mathbb{R}$ *is Borel isomorphic with the Baire space* $\mathcal{N}$, *and, in particular,* $\mathbb{R} =_c \mathcal{N}$.

With the true (von Neumann) cardinals developed in Chapter **12** from **ZFC**, this gives the congenial equation $|\mathbb{R}| = \mathfrak{c} = 2^{\aleph_0}$.

Problems for Appendix A

***xA.1.** Let F be the set of all real, rational functions, i.e., all (partial) real functions which can be represented as quotients of polynomials with coefficients in $\mathbb{R}$, and prove that it is a field with the obvious algebraic operations.

Show that the relation

$$f \leq g \iff_{\text{df}} (\exists x)(\forall y \geq x)[f(y) \leq g(y)]$$

is a linear ordering of F, and with it F is a non-archimedean ordered field. HINT. The field $\mathbb{Q}_F$ comprises the constant functions with rational value, and the identity $(x \mapsto x)$ is above all of them.

*xA.2. Prove that there exists an ordered field which is not complete, but in which every Cauchy sequence has a limit. HINT. Show that every ordered field has a *Cauchy completion*.

xA.3. Every open set of reals is a countable union of disjoint open intervals.

xA.4. Every closed interval of real numbers $[a, b]$ is compact, in the topological sense, **10.43**.

xA.5. Every closed set of real numbers is a countable union of compact sets.

xA.6. Every closed set of real numbers F can be written uniquely as the disjoint union of a perfect and a countable set.

*xA.7. Prove Theorem **A.56**.

APPENDIX B

AXIOMS AND UNIVERSES

The serious study of *models* of axiomatic set theories depends heavily on methods from *mathematical logic* which are outside the scope of these Notes.[32] Here we will consider only SET UNIVERSES, generalizations of the Zermelo and the **ZFDC**-universes of Chapter **11**, which are very special models and can be studied by standard mathematical techniques, as we study fields or topological spaces. First we will prove that THE ZERMELO UNIVERSES of Chapter **11** ARE MODELS OF **ZDC** and THE **ZFDC**-UNIVERSES ARE MODELS OF **ZFDC**; this will give us a better understanding of these universes, and it will also yield some simple CONSISTENCY AND INDEPENDENCE RESULTS for the corresponding theories. In the main part of this chapter we will construct some new set universes with quite different properties, including the ANTIFOUNDED UNIVERSE of Aczel which contains a rich variety of ill founded sets. We will glean some consistency results from these models too, but consistency results are not our main concern: our primary interest is to explore and understand several natural, intuitive notions of SET and compare them with the standard conception of PURE, GROUNDED SET discussed in **12.34**.

We begin with a result about the least Zermelo universe $\mathcal{Z}$ which is somewhat surprising, given how much we promoted $\mathcal{Z}$ in **11.25** as a rich collection of sets which contains all objects of interest of classical mathematics.

B.1. Theorem. *The set V_ω of all pure, grounded, hereditarily finite sets is not a member of $\mathcal{Z}$; in fact, there is no set $A \in \mathcal{Z}$ such that*

$$\emptyset \in A \,\&\, (\forall X)[X \in A \Longrightarrow \mathcal{P}(X) \in A]. \tag{B-1}$$

PROOF. For each $x \in \mathcal{Z}$, we let

$$\text{level}(x) = \text{the least } n \text{ such that } x \in \mathcal{Z}_n, \tag{B-2}$$

---

[32]For those who do know logic, we remark here that the most natural way to formalize the theories we have studied is in a *many sorted predicate logic with identity*, with separate variables for objects, definite conditions and definite operations of every arity, and relation symbols Set and $\in$. Notice that we did not assume in **3.18** any extensionality principles for conditions or operations, and we have never appealed to such principles. This means that to get (a notational variant of) the classical *Gödel-Bernays Theory* from the **ZFC** of **11.31** we must add *extensionality for conditions*. On the other hand, precisely because we did not include any extensionality axioms for conditions, we can view the present **ZFC** as a notational variant of the classical *Zermelo-Fraenkel Theory*, by interpreting "conditions" to be just "formulas with set parameters".

so that members of $\mathbb{N}_0$ have level 0, but

$$\text{level}(x) > 0 \Longrightarrow \text{level}(\{x\}) = \text{level}(x) + 1;$$

because if $x \in \mathcal{Z}_n$, then $\{x\} \in \mathcal{Z}_{n+1}$, and if $\{x\} \in \mathcal{Z}_n$ with $n > 0$, then $\{x\} \subseteq \mathcal{Z}_{n-1}$ by (11-20), which implies $x \in \mathcal{Z}_{n-1}$. Define now by recursion the sets

$$A_0 = \{\emptyset, \{\emptyset\}\}, \quad A_{n+1} = \{A_n\}.$$

Clearly, each $A_n \in V_\omega$, $\text{level}(A_0) = 1$, since $A_0 \subseteq \mathbb{N}_0$ but $A_0 \notin \mathbb{N}_0$, and by induction, $\text{level}(A_n) = n + 1$.

Suppose now that there is some $A \in \mathcal{Z}$ which satisfies (B-1), and notice that $A_0 \in \mathcal{P}(\mathcal{P}(\emptyset)) \in A$, and then by induction, for each n, $A_n \in A$. If $A \in \mathcal{Z}_m$, then each $A_n \in \mathcal{Z}_m$ because $\mathcal{Z}_m$ is transitive, and hence $\text{level}(A_n) \leq m$ for each n, which violates what we just proved. $\dashv$

A similar argument shows that (with Kuratowski tupling) $\bigcup_{n=2}^{\infty} \{\mathbb{N}_0\}^n \notin \mathcal{Z}$, Problem **xB.2**, and it is quite easy to extend it to show that $\mathcal{Z}$ misses many more very simple sets, but the fact that it lacks V_ω is undoubtedly the most startling of the lot. The construction of V_ω is so direct, it seems to follow so naturally from our most basic intuitions about sets, that it is really hard to believe that we developed all this set theory in Chapters **3 – 10** and Appendix **A** from axioms which do not guarantee its existence. One may try to write off this feature of the Zermelo axioms as a small oversight of Zermelo and strengthen the axioms in some minor way to ensure the existence of V_ω, but this is the wrong way to go. On the one hand, we know the natural extension of **ZDC+AC**, it is the addition of the Replacement Axiom which can be justified by arguments only marginally different from those used to justify the Separation Axiom. And on the other hand, the importance of **ZDC+AC** lies precisely in its two, contrasting features: that it can prove so much about classical mathematics (which is its real domain), while it can be interpreted in such simple, easy to comprehend models like $\mathcal{Z}$. Whatever doubts may have lingered about the soundness of set theory from the paradoxes should be at least moderated by the strength of **ZDC+AC** and its ease of interpretation.

We still need to make precise the sense in which $\mathcal{Z}$ is a "model" of **ZDC**. Perhaps the simplest way to introduce the key, new idea we need, is to try and reinterpret **B.1** as an independence result.

B.2. Theorem? *We cannot prove in* **ZDC** *the proposition that some set A exists, which contains the empty set and is closed under the powerset operation.*

PROOF. Spelled out symbolically for precision, the proposition in question is

$$\phi \Longleftrightarrow_{\text{df}} (\exists A)[\emptyset \in A \,\&\, (\forall X)[X \in A \Longrightarrow \mathcal{P}(X) \in A]].$$

Suppose, towards a contradiction, that ϕ is a theorem of **ZDC**. Since the least Zermelo universe $\mathcal{Z}$ has all the closure properties demanded by the axioms of **ZDC**, any proof of ϕ from these axioms could be translated into a proof of the interpretation of ϕ in $\mathcal{Z}$, which is

$$\phi^{(\mathcal{Z})} \Longleftrightarrow (\exists A \in \mathcal{Z})[\emptyset \in A \,\&\, (\forall X \in \mathcal{Z})[X \in A \Longrightarrow \mathcal{P}(X) \in A]].$$

As a consequence, $\phi^{(\mathcal{Z})}$ is true, so there exists a set $A \in \mathcal{Z}$ satisfying

$$\emptyset \in A \,\&\, (\forall X \in \mathcal{Z})[X \in A \Longrightarrow \mathcal{P}(X) \in A],$$

which by the transitivity of $\mathcal{Z}$ is equivalent to

$$\emptyset \in A \,\&\, (\forall X)[X \in A \Longrightarrow \mathcal{P}(X) \in A]; \tag{B-3}$$

but by **B.1**, no $A \in \mathcal{Z}$ satisfies (B-3).   ⊣

The argument is unfamiliar, unless you know a lot of logic, and in any case it is incomplete, only a sketch. What we need to elucidate is the meaning of "interpreting a proposition in $\mathcal{Z}$", the move from ϕ to $\phi^{(\mathcal{Z})}$ above, and it may help to consider an example. Suppose we have located a traditional mathematician (perhaps an old-fashioned analyst) who disclaims any interest in general set theory beyond its applications to classical mathematics, he has studied Chapters **3 – 10** and Appendix **A** and he is convinced that all the objects he cares about live in $\mathcal{Z}$. In an effort to simplify his world, he declares that henceforth by "object" he will mean "member of $\mathcal{Z}$": that is his universe. Suppose further that this person now utters the Powerset Axiom and claims that he believes it. *What does he mean?* Spelled out in terms of the primitive notions of sethood and membership, the powerset axiom reads as follows:

(**IV**): *For each set A, there exists a set B, such that for every X,*

$$X \in B \iff \mathrm{Set}(X) \,\&\, (\forall t)[t \in X \Longrightarrow t \in A]. \tag{B-4}$$

Replacing "*set*" by "*member of $\mathcal{Z}$*" in this, we get something rather different.

(**IV**)$^{(\mathcal{Z})}$: *For each set $A \in \mathcal{Z}$, there exists some set $B \in \mathcal{Z}$, such that for every $X \in \mathcal{Z}$,*

$$X \in B \iff \mathrm{Set}(X) \,\&\, (\forall t \in \mathcal{Z})[t \in X \Longrightarrow t \in A]. \tag{B-5}$$

This is what our friend really means when he tries to tell us that every set has a powerset, and it differs enough from the Powerset Axiom that its truth is not immediately apparent. To prove it, for each $A \in \mathcal{Z}$, we naturally take $B = \mathcal{P}(A)$, which is also in $\mathcal{Z}$ and satisfies (B-4), for every X. Notice next that for each $X \in \mathcal{Z}$,

$$(\forall t \in \mathcal{Z})[t \in X \Longrightarrow t \in A] \iff (\forall t)[t \in X \Longrightarrow t \in A],$$

easily, by the transitivity of $\mathcal{Z}$. This reduces (B-5) to

$$X \in B \iff \mathrm{Set}(X) \,\&\, (\forall t)[t \in X \Longrightarrow t \in A], \tag{B-6}$$

which is true for every X and hence (in particular) for every $X \in \mathcal{Z}$.

As a matter of fact, all the axioms of **ZDC** yield true propositions when we replace in them "*object*" by "*member of $\mathcal{Z}$*". It follows that all the theorems derived from the axioms of **ZDC** *by logic alone* also yield true propositions when we understand them as assertions about $\mathcal{Z}$ in the same way, so our stipulated classical friend can safely work in **ZDC** and be assured that he is proving statements which are true of his world. This not entirely trivial proposition expresses the fact that the universe $\mathcal{Z}$ is a "model" of **ZDC**.

B.3. A **set universe** $\mathcal{M}$ is any triple M, S, E, of a class (which may be a set) M, a subclass $S \subseteq M$, and a binary definite condition E such that $E(x, y) \Longrightarrow x, y \in M$ and for each $x \in M$,

$$\boldsymbol{b}_{\mathcal{M}}(x) =_{\mathrm{df}} \{t \mid t\,E\,x\} \text{ is a set,} \tag{B-7}$$

which means that for some set X and all t,

$$t \in X \iff E(t, x).$$

We write synonymously

$$t\,E\,x \iff E(t, x)$$

for the condition E, which interprets the $\in$ condition in $\mathcal{M}$, and we call $\boldsymbol{b}_{\mathcal{M}}(x)$ the **body** of each $x \in M$. The $\mathcal{M}$-**objects** are the members of M and the $\mathcal{M}$-**sets** are the members of S. An n-ary definite condition is an $\mathcal{M}$-**condition** if it only holds of objects in M, i.e.,

$$P(x_1, \ldots, x_n) \Longrightarrow x_1, \ldots, x_n \in M;$$

and a definite n-ary operation F is an $\mathcal{M}$-**operation** if it assigns $\mathcal{M}$-objects to $\mathcal{M}$-objects, i.e.,

$$x_1, \ldots, x_n \in M \Longrightarrow F(x_1, \ldots, x_n) \in M.$$

A universe $\mathcal{M}$ is **natural** if the class M is transitive and S, E are the standard sethood and membership conditions, i.e.,

$$X \in M \implies X \subseteq M,$$
$$x \in S \iff x \in M \,\&\, \mathrm{Set}(x),$$
$$x\,E\,y \iff x, y \in M \,\&\, x \in y.$$

A natural universe $\mathcal{M}$ is completely determined by the transitive class M of its objects and we will identify it with that class, i.e., when we refer to *the universe M* for a transitive class M, we will mean the natural universe with objects the members of M. Notice that in a natural universe the body of each object is the set of its members,

$$\boldsymbol{b}_M(x) = \{t \mid t \in x\} \quad (x \in M); \tag{B-8}$$

this means that $\boldsymbol{b}_M(x) = x$, if x is a set, but $\boldsymbol{b}_M(x) = \emptyset$ if x is an atom.

The **relativization to** a universe $\mathcal{M}$ of a proposition θ is the proposition $\theta^{(\mathcal{M})}$ constructed by replacing "object" by "$\mathcal{M}$-object", "set" by "$\mathcal{M}$-set", "$x \in y$" by "$x\,E\,y$", "condition" by "$\mathcal{M}$-condition" and "operation" by "$\mathcal{M}$-operation", consistently wherever these expressions occur in θ. If $\theta^{(\mathcal{M})}$ is true, we say that θ **is true in the universe** $\mathcal{M}$ or (synonymously) that $\mathcal{M}$ **satisfies** or **models** θ.

A universe $\mathcal{M}$ is a **model** of an axiomatic system T if every axiom of T is true in $\mathcal{M}$, i.e., if the relativization $\theta^{(\mathcal{M})}$ of every axiom θ of T is a true proposition. If this holds, we call $\mathcal{M}$ a **universe of T**, or simply a T-**universe**.

B.4. Propositions and relativizations. By *proposition* we mean any ordinary, definite mathematical statement or assertion, just like the axioms, hypotheses and theorems we have considered so far. This is not completely precise, everyday mathematical English not being a perfectly specified language. The basic idea is that for all objects x and y, the expressions $x = y$, $x \in y$ and $\mathrm{Set}(x)$ are propositions; for each definite condition P and objects $x_1, \ldots, x_n$, the expression $P(x_1, \ldots, x_n)$ is a proposition; and that propositions may be combined by the basic operations of logic, $\neg$, $\vee$, $\exists$ and the like. *Relativizations*[33] to a set universe $\mathcal{M}$ are computed as one might expect, e.g.,

$$(x = y)^{(\mathcal{M})} : x = y,$$
$$\mathrm{Set}(x)^{(\mathcal{M})} : x \in S,$$
$$(x \in y)^{(\mathcal{M})} : x \, E \, y,$$
$$P(x_1, \ldots, x_n)^{(\mathcal{M})} : P(x_1, \ldots, x_n),$$
$$(\neg \theta)^{(\mathcal{M})} : \neg(\theta^{(\mathcal{M})}),$$
$$(\phi \, \& \, \psi)^{(\mathcal{M})} : \phi^{(\mathcal{M})} \, \& \, \psi^{(\mathcal{M})},$$
$$((\forall x)\phi)^{(\mathcal{M})} : (\forall x \in M)\phi^{(\mathcal{M})},$$
$$((\forall P)\phi)^{(\mathcal{M})} : (\forall P)[(\forall x, y)[P(x, y) \Longrightarrow x, y \in M] \Longrightarrow \phi^{(\mathcal{M})}],$$

where P varies over binary, definite conditions, etc. In the proofs which follow, we will take care to spell out laboriously the relativizations of all the propositions that concern us. This will produce many examples which illustrate the notion, and it will also ensure that the specific results we claim will be precise and rigorously established even if the general notion of proposition and the relativization process are not precisely delimited. On the other hand, much of the significance of the results comes from the following general principle. It says simply that logical consequences of true (in $\mathcal{M}$) propositions are true (in $\mathcal{M}$) for any universe $\mathcal{M}$, and, of course, it holds for arbitrary models of arbitrary axiomatic theories, not just set universes.

B.5. Principle of Soundness of Logical Inference. *If a proposition θ is a theorem of an axiomatic system T* (*i.e., it can be proved by logic alone from the axioms of T*), *then every universe of T satisfies θ.*

B.6. Universes vs. general models. According to the discussion in **8.22**, to define a model of an axiomatic set theory we must specify a domain of objects,

---

[33]To those who know formal logic, it might appear that the completely precise (syntactical) relativization operation on *formulas* is much easier to understand than this relativization operation, which is applied directly to propositions of "mathematical English". But to apply the formal relativization process, we must first "formalize" the ordinary language propositions in which we are ultimately interested, and a moment's thought will convince you that the formalization process is *exactly* as "vague" as the present operation of relativization. It may be argued that we know how to formalize a certain proposition precisely if we can interpret it in some arbitrary structure, i.e., precisely if we can understand its informal relativizations.

define on it the conditions of membership and sethood and also specify which conditions and operations on the domain will be considered definite. Set universes are very special models in two ways.

(1) When we view a set universe $\mathcal{M}$ as a model for an axiomatic set theory, we take its *definite conditions* to be all the definite conditions of our basic domain $\mathcal{W}$, and we take for its *definite operations* all the $\mathcal{M}$-operations, i.e., the definite operations of $\mathcal{W}$ which take $\mathcal{M}$-objects to $\mathcal{M}$-objects.[34] It is routine to verify that all the axioms for definite conditions and operations listed in **3.18** hold with this interpretation, Problem **xB.1**: thus, to prove that a set universe $\mathcal{M}$ is a model of (say) **ZFDC**, it is enough to prove the relativizations to $\mathcal{M}$ of axioms (**I**) – (**VIII**).

(2) Because we assume that the body $\boldsymbol{b}_{\mathcal{M}}(x)$ of each $x \in M$ is a set, $\mathcal{M}$-sets cannot be "larger" than the sets of the intended universe $\mathcal{W}$: for example, there cannot be an $\mathcal{M}$-set x such that for all t, $t\,E\,x$.

Natural universes are even more special, of course, as they only restrict the domain of objects to some transitive class—and this makes it especially simple to understand the meaning of the relativization operation for them. From the mathematical point of view, natural universes are *subuniverses* of $\mathcal{W}$ and they inherit their structure from $\mathcal{W}$, much like subgroups, subposets, topological subspaces and the like are specified by a subset of some given space and inherit the relevant structure from it. The additional subtlety here is that we need to interpret (relativize) in these subuniverses propositions which are logically quite complex, more complex than the typical identities or inclusions which come up in Algebra or Topology.

We begin with the verification of the axioms we have been studying in natural universes which we have already introduced, where the notions are most familiar.

B.7. Lemma. *Every transitive class M satisfies the Axiom of Extensionality.*

Proof. The relativization to M of the Extensionality Axiom reads as follows:

$(\mathbf{I})^{(M)}$ Extensionality: *For all sets $A, B \in M$,*

$$A = B \iff (\forall x \in M)[x \in A \iff x \in B].$$

If $A = B$, then A and B have the same members, so they certainly have the same members in M. To prove the converse, we must show that if $A \neq B$, then there must exist some $x \in M$, such that either $x \in A \setminus B$ or $x \in B \setminus A$; but if $A \neq B$, then there certainly exists some

$$x \in (A \setminus B) \cup (B \setminus A) \subseteq A \cup B,$$

$A, B \subseteq M$ because M is transitive, and hence this x is also in M.   ⊣

---

[34]More pedantically, the definite conditions of $\mathcal{M}$ are the restrictions to M of the definite conditions in the intended universe $\mathcal{W}$, and similarly for the operations.

B.8. Lemma. *Every Zermelo universe M is a natural universe of axioms* (**I**) – (**VI**).

PROOF. We consider in turn each of the axioms (**I**) – (**VI**) other than (**I**) just shown and (**IV**) for which we gave the argument in the example above. The reader is advised to compare the relativization of each axiom to M which we must prove with the original statement of the axiom in Chapter **3**.

(**II**)$^{(M)}$ Emptyset and Pairset: (a) *There is a special object $y \in M$ which is a set but has no members in M.* (b) *For all $x, y \in M$, there is a set $z \in M$ such that*

$$(\forall t \in M)[t \in z \iff [t = x \vee t = y]]. \tag{B-9}$$

(a) $\emptyset \in M$ by hypothesis, it has no members whatsoever, so it certainly has no members in M. (b) If $x, y \in M$, then $z = \{x, y\} \in M$ by hypothesis, and it obviously satisfies (B-9), since it satisfies the stronger

$$(\forall t)[t \in z \iff [t = x \vee t = y]].$$

(**III**)$^{(M)}$ Separation Axiom: *For each set $A \in M$ and each unary, definite condition P, there exists a set $B \in M$ which satisfies the equivalence*

$$(\forall x \in M)[x \in B \iff x \in A \,\&\, P(x)]. \tag{B-10}$$

Suppose $A \in M$. By the Axiom of Separation, there exists some B which satisfies

$$(\forall x)[x \in B \iff x \in A \,\&\, P(x)]. \tag{B-11}$$

Now $B \subseteq A$, so $B \in \mathcal{P}(A) \in M$, so $B \in M$ because M is transitive, and (B-11) implies the weaker (B-10).

(**V**)$^{(M)}$ Unionset Axiom: *For each set $\mathscr{E} \in M$, there exists a set $B \in M$ which satisfies the equivalence*

$$(\forall t \in M)[t \in B \iff (\exists X \in M)[X \in \mathscr{E} \,\&\, t \in X]]. \tag{B-12}$$

Again, we naturally take $B = \bigcup \mathscr{E}$, which is in M by hypothesis and satisfies the equivalence

$$(\forall t)[t \in B \iff (\exists X)[X \in \mathscr{E} \,\&\, t \in X]]. \tag{B-13}$$

Using once more the transitivity of M, immediately, for every t,

$$(\exists X)[X \in \mathscr{E} \,\&\, t \in X] \iff (\exists X \in M)[X \in \mathscr{E} \,\&\, t \in X],$$

so that (B-12) reduces to

$$(\forall t \in M)[t \in B \iff (\exists X)[X \in \mathscr{E} \,\&\, t \in X]],$$

which is implied by the stronger (B-13).

(**VI**)$^{(M)}$ Axiom of Infinity: *There exists a set $I \in M$ such that*

$$\emptyset \in I \,\&\, (\forall x \in M)[x \in I \Longrightarrow \{x\} \in I].$$

This is quite simple, taking $I = \mathbb{N}_0 \in M$. $\qquad \dashv$

B.9. Theorem. (1) *Every Zermelo universe is a natural universe of* **ZDC**, *and every* **ZFDC***-universe is a natural universe of* **ZFDC**.

(2) (**AC**) *Every Zermelo universe is a natural universe of* **ZDC+AC**, *and every* **ZFDC***-universe is a natural universe of* **ZFDC+AC**.

In particular, $\mathcal{Z}$ *and every* $M(I)$ *such that* $\mathbb{N}_0 \subseteq I$ *are natural universes of* **ZDC**, *or universes of* **ZAC**, *granting* **AC**.

(3) (von Neumann) (**AC**) *The von Neumann universe* $\mathcal{V}$ *is a natural universe of* **ZFC**.

PROOF. The relativizations of **DC** and **AC** were proved in **11.23** and the relativization of the Axiom of Replacement is exactly the defining condition of a **ZFDC**-universe in **11.33**. For Part (3), we have already shown that $\mathcal{V}$ is a **ZFDC**-universe in **11.34**, so it only remains to prove the relativizations to $\mathcal{V}$ of the Principles of Purity **3.25** and Foundation **11.29**. The first of these is

Purity$^{(\mathcal{V})}$: *For every* $x \in \mathcal{V}$, $\mathrm{Set}(x)$,

and it is true simply because every object in $\mathcal{V}$ is a set, and the interpretation of "sethood" in $\mathcal{V}$ is the standard one. For the second, it is easiest to relativize the elementary version of the Foundation Principle in **11.30**.

Foundation$^{(\mathcal{V})}$: *For every set* $X \in \mathcal{V}$, *there exists some* $m \in X$ *such that* $m \cap X$ *is empty in* $\mathcal{V}$, *i.e.,*

$$(\forall t \in \mathcal{V})[t \notin m \vee t \notin X].$$

The negation of this would give us an $X \in \mathcal{V}$ and some $a \in X$ such that

$$(\forall m \in X)(\exists t \in X)[t \in m],$$

which by **DC** (as in the proof of **11.30**) implies that X is ill founded, contradicting the assumption $X \in \mathcal{V}$. ⊣

The Soundness of Logical Inference **B.5** combines well with **B.9** to yield many simple but interesting independence results about **ZDC** and **ZDC+AC**: to prove that a proposition θ cannot be a theorem of **ZDC**, it is enough to find some $I \supseteq \mathbb{N}_0$ such that $\theta^{(M(I))}$ is false. This is exactly the way we proved **B.2**, a bit clumsily without the precise notions. We have included in the problems several examples of this kind.

By the same reasoning, Part (3) of Theorem **B.9** implies that *we cannot refute in* **ZFAC** *the Principles of Purity or Foundation*, because $\mathcal{V}$ is a model of **ZFAC** which satisfies these principles. It should also be obvious that we cannot prove these principles in **ZFAC**, but to establish this rigorously we need to construct universes of **ZFAC** which are not natural. The basic tool for such constructions is the next simple notion.

B.10. Definition. A **Rieger universe** is any set universe $\mathcal{M} = M, S, E$ such that for every set $Y \subseteq M$, there exists exactly one $\mathcal{M}$-set X satisfying $Y = \boldsymbol{b}_{\mathcal{M}}(X)$. For each $Y \subseteq M$ we set

$$\rho_{\mathcal{M}}(Y) =_{\mathrm{df}} \text{the unique } X \in S \text{ such that } \boldsymbol{b}_{\mathcal{M}}(X) = Y, \tag{B-14}$$

so that $\rho_{\mathcal{M}}$ is a definite operation and immediately from its definition, for every $Y \subseteq M$,

$$X = \rho_{\mathcal{M}}(Y) \iff X \in S \,\&\, \boldsymbol{b}_{\mathcal{M}}(X) = Y, \qquad (\text{B-15})$$

$$t \, E \, \rho_{\mathcal{M}}(Y) \iff t \in Y. \qquad (\text{B-16})$$

B.11. Rieger's Theorem. *Every Rieger universe is a universe of* **ZFDC**.

PROOF. Fix $\mathcal{M} = M, S, E$ with the Rieger property and let $\boldsymbol{b}(x) = \boldsymbol{b}_{\mathcal{M}}(x)$, skipping the subscript since $\mathcal{M}$ is the only universe around. We verify in turn the relativizations of all the axioms of **ZFDC**.

$(\mathbf{I})^{(\mathcal{M})}$ Extensionality Axiom: *For all $A, B \in M$, if $S(A)$ and $S(B)$, then*

$$A = B \iff (\forall x \in M)[x \, E \, A \iff x \, E \, B].$$

If $A = B$, then surely, for all x, $x \, E \, A \iff x \, E \, B$. Conversely, if for all $x \in M$, $x \, E \, A \iff x \, E \, B$, then $\boldsymbol{b}(A) = \boldsymbol{b}(B)$; by the Rieger property, there is exactly one $C \in S$ such that $\boldsymbol{b}(A) = \boldsymbol{b}(C)$; so we must have $C = A$, and similarly $C = B$, hence $A = B$.

$(\mathbf{II})^{(\mathcal{M})}$ Emptyset and Pairset: (a) *There is a special object $y \in S$ such that* $(\forall t \in M)\neg t \, E \, y$. (b) *For all $x, y \in M$, there is some $z \in S$ such that*

$$(\forall t \in M)[t \, E \, z \iff t = x \vee t = y].$$

For (a), choose y so that $y \in S$ and $\boldsymbol{b}(y) = \emptyset$, and for (b) choose $z \in S$ so that $\boldsymbol{b}(z) = \{x, y\}$, both times by applying directly the Rieger property. In the case of (b), for example, we compute:

$$t \, E \, z \iff t \in \boldsymbol{b}(z) \iff [t = x \vee t = y],$$

which is the required conclusion.

$(\mathbf{IV})^{(\mathcal{M})}$ Powerset Axiom: *For each set $A \in M$, there exists some $B \in S$, such that for every $X \in M$,*

$$X \, E \, B \iff X \in S \,\&\, (\forall t \in M)[t \, E \, X \Longrightarrow t \, E \, A].$$

Given $A \in M$, choose $B \in S$ by the Rieger property so that

$$\boldsymbol{b}(B) = \{\rho(Y) \mid Y \subseteq \boldsymbol{b}(A)\}, \qquad (\text{B-17})$$

where $\rho(Y) = \rho_{\mathcal{M}}(Y)$ is the Rieger operation associated with $\mathcal{M}$ by (B-14), and compute:

$$\begin{aligned}
X \, E \, B &\iff X \in \boldsymbol{b}(B) \\
&\iff (\exists Y)[Y \subseteq \boldsymbol{b}(A) \,\&\, X = \rho(Y)] \\
&\iff (\exists Y)[Y \subseteq \boldsymbol{b}(A) \,\&\, [X \in S \,\&\, \boldsymbol{b}(X) = Y]] \quad \text{by (B-15)} \\
&\iff X \in S \,\&\, \boldsymbol{b}(X) \subseteq \boldsymbol{b}(A) \\
&\iff X \in S \,\&\, (\forall t \in M)[t \, E \, X \Longrightarrow t \, E \, A].
\end{aligned}$$

Verifications of the remaining axioms in $\mathcal{M}$ are similar, following the same ideas as in the proof of Theorem **B.9**, and we leave them for the exercises. $\dashv$

B.12. Exercise. *Prove that a set universe $\mathcal{M} = M, S, E$ is a Rieger universe if and only if* (1) *it satisfies the Axiom of Extensionality, and* (2) *for every $Y \subseteq M$, there exists some $X \in S$ such that $\boldsymbol{b}_{\mathcal{M}}(X) = Y$.*

B.13. Exercise. *Every Rieger universe is a model of the Axioms of Separation* (**III**) *and Replacement* (**VIII**).

B.14. Exercise. (**AC**) *Every Rieger universe is a model of* **ZFAC**.

B.15. Relativization of "faithfully modeled" notions. The Choice Principles **DC** and **AC** are formulated in terms of the notions of "function" which was defined in Chapter **4** using an "arbitrary but fixed" ordered pair operation **4.1**, and "system of natural numbers", which was also "arbitrary but fixed" in **5.9**. It is not completely obvious how to relativize propositions involving such "faithfully modeled" notions, since (for example) a given universe $\mathcal{M}$ may not be closed under the chosen ordered pair operation. We avoided the problem in **11.23** and **B.9** by assuming that the fixed ordered pair is the Kuratowski pair under which every universe satisfying (**I**) – (**VI**) is closed, but there may be some lingering vagueness on how to deal with this problem in general. There is an easy solution for Rieger universes, which we outline in Problems **xB.13** and **xB.14**. In discussing Rieger universes from now on, we will assume tacitly that we have fixed an ordered pair operation, a system of natural numbers, etc. for each of them, and we will relativize propositions which involve these faithfully modeled notions in terms of these fixed operations. Problems **xB.13** and **xB.14** make it clear that which particular definitions are chosen is irrelevant for the results.

B.16. Proposition. *There exists a Rieger universe $\mathcal{M}_{at}$ in which every set is equinumerous with a set of atoms; in particular, we cannot prove in* **ZFDC** *that every set is pure.*

Proof. The idea is to code every set A by the pair $(0, A)$ in $\mathcal{M}_{at}$, and to declare that every object which does not code a set in this way is an atom. We set

$$x \in M_{at} \iff_{df} x = x, \text{ so } M_{at} = \mathcal{W},$$
$$x \in S_{at} \iff_{df} (\exists A)[\text{Set}(A) \,\&\, x = (0, A)],$$
$$x \, E_{at}\,(0, A) \iff_{df} x \in A,$$

and proceed to verify that $\mathcal{M}_{at} = M_{at}, S_{at}, E_{at}$ is a set universe, that it has the Rieger property and that it satisfies the proposition *"every set is equinumerous with a set of atoms"*.

Skipping the subscript $\mathcal{M}_{at}$, clearly

$$\boldsymbol{b}(x) = \begin{cases} \emptyset, & \text{if for all sets } A, \ x \neq (0, A), \\ A, & \text{if for some (necessarily unique) set } A, \ x = (0, A), \end{cases}$$

so each $\boldsymbol{b}(x)$ is a set. For the Rieger property, we notice first that for each set $A \subseteq M_{at}$, $(0, A) \in S_{at}$ and $\boldsymbol{b}((0, A)) = A$; if $x \in S$ and $\boldsymbol{b}(x) = A$, then

$x = (0, B)$ and $\boldsymbol{b}(x) = B$ for some set B, by the definition, so we have $A = B$ and $x = (0, A)$, as required. The Rieger operation is very simple in this case,

$$\rho(Y) = (0, Y).$$

The relativization to $\mathcal{M}_{at}$ of the proposition "every set is equinumerous with a set of atoms" is the assertion that *for every $\mathcal{M}_{at}$-set $(0, A)$, there is a bijection in $\mathcal{M}_{at}$ between $(0, A)$ and some $\mathcal{M}_{at}$-set of atoms $(0, B)$.* We take

$$B =_{df} \{(1, t) \mid t \in A\};$$

now $(0, B) \in S_{at}$ and, by the definition,

$$x\, E_{at}\, (0, B) \Longrightarrow x \in B \Longrightarrow (\forall y)[x \neq (0, y)] \Longrightarrow x \notin S_{at},$$

i.e., each $\mathcal{M}_{at}$-member of $(0, B)$ is an atom in $\mathcal{M}_{at}$. If $(x, y)_{at}$ is the ordered pair operation of $\mathcal{M}_{at}$ and

$$f =_{df} \rho\{(t, (1, t))_{at} \mid t \in A\},$$

then $\mathcal{M}_{at}$ recognizes f as a bijection between the $\mathcal{M}_{at}$-sets $(0, A)$ and $(0, B)$.⊣

By the anthropomorphic "$\mathcal{M}_{at}$ *recognizes f ...* " we simply mean that if $A' = (0, A)$ and $B' = (0, B)$ are the objects in M_{at} which code the sets A and B, respectively, and if we set

$$\theta \Longleftrightarrow_{df}\ f \subseteq A \times B \,\&\, (\forall x \in A')(\exists y \in B')(x, y) \in f$$
$$\&\, (\forall y \in B')(\exists! x \in A')[(x, y) \in f],$$

then the relativization $\theta^{(\mathcal{M}_{at})}$ of the proposition θ to $\mathcal{M}_{at}$ is true. This relativization can be computed in principle, but it is quite messy. It is best to develop a machinery for arguing about relativizations without actually writing them out, and for this the following, traditional "model theoretic" notation is very useful. For each set universe $\mathcal{M}$ and each proposition θ,

$$\mathcal{M} \models \theta \Longleftrightarrow_{df} \theta \text{ is true in } \mathcal{M} \Longleftrightarrow \theta^{(\mathcal{M})}. \qquad \text{(B-18)}$$

We read $\mathcal{M} \models \theta$ "*$\mathcal{M}$ models θ*", but also "*$\mathcal{M}$ thinks that θ*", "*$\mathcal{M}$ believes that θ*", etc. as befits the occasion. For example, for each pair of classes M, S and each binary condition E, let

$$\text{Setuniv}(M, S, E) \Longleftrightarrow_{df}\ (\forall u, v)[u\, E\, v \Longrightarrow u, v \in M\,]$$
$$\&\, (\forall t)[t \in S \Longrightarrow t \in M\,]$$
$$\&\, (\forall x)[x \in M \Longrightarrow$$
$$(\exists X)[\text{Set}(X)\,\&\,(\forall t)[t \in X \Longleftrightarrow t\, E\, x]]$$

be the fairly complex proposition which asserts that M, S, E comprise a set universe. Consider also

$$\text{Rieger}(M, S, E) \Longleftrightarrow_{df}\ (\forall Y)[(\text{Set}(Y)\,\&\,Y \subseteq M)$$
$$\Longrightarrow (\exists! X \in S)[Y = \boldsymbol{b}_{\mathcal{M}}(X)]]$$
$$\Longleftrightarrow\ (\forall Y)[(\text{Set}(Y)\,\&\,Y \subseteq M)$$
$$\Longrightarrow (\exists! X \in S)(\forall t \in M)[t \in Y \Longleftrightarrow t\, E\, X]],$$

which asserts that $\mathcal{M}$ is a Rieger universe. These are propositions about M, S and E, which may be true or false; whether they are true or not, it makes sense to interpret them (relativize them) in some universe $\mathcal{M}'$ and ask if $\mathcal{M}'$ *thinks that* M, S, E comprise a Rieger universe! There is an obvious question here, which has a simple and useful answer.

B.17. Theorem. *Suppose* $\mathcal{M}_1 = M_1, S_1, E_1$ *is a Rieger universe,* $S_2 \subseteq M_2 \subseteq M_1$ *are classes and* E_2 *is a binary condition such that* $x \, E_2 \, y \Longrightarrow x, y \in M_2$; *if*

$$\mathcal{M}_1 \models \Big[\mathrm{Setuniv}(M_2, S_2, E_2) \, \& \, \mathrm{Rieger}(M_2, S_2, E_2) \Big],$$

then $\mathcal{M}_2 = M_2, S_2, E_2$ *is also a Rieger universe.*

Proof. To show first that $\mathcal{M}_2$ is a set universe, we must verify that for every $x \in M_2$,

$$(\exists Y \in \mathrm{Set})(\forall t)[t \in Y \iff t \in M_2 \, \& \, t \, E_2 \, x]. \tag{B-19}$$

Fix some $x \in M_2$. The proposition (B-19) is true in $\mathcal{M}_1$ since

$$\mathcal{M}_1 \models \mathrm{Setuniv}(M_2, S_2, E_2),$$

which means that some $Y_1 \in S_1$ exists such that for all $t \in M_1$,

$$t \, E_1 \, Y_1 \iff t \in M_2 \, \& \, t \, E_2 \, x; \tag{B-20}$$

and since $\mathcal{M}_1$ is a set universe, there exists some set Y such that for all t,

$$t \in Y \iff t \in M_1 \, \& \, t \, E_1 \, Y_1; \tag{B-21}$$

now (B-20) and (B-21) together imply what (B-19) demands for the given $x \in M_2$ and some Y.

To show that $\mathcal{M}_2$ is a Rieger universe, we must show first that if $Y \subseteq M_2$, then there exists some $X_2 \in S_2$ such that for every $t \in M_2$,

$$t \in Y \iff t \, E_2 \, X_2, \tag{B-22}$$

and then verify that this X_2 is unique. Since $Y \subseteq M_2 \subseteq M_1$ and $\mathcal{M}_1$ is a Rieger universe, there exists some $X_1 \in S_1$ such that for all $t \in M_1$,

$$t \in Y \iff t \, E_1 \, X_1; \tag{B-23}$$

working in $\mathcal{M}_1$, we apply the Rieger property for $\mathcal{M}_2$ to the $\mathcal{M}_1$-set X_1, to get some $X_2 \in S_2$ such that

$$\mathcal{M}_1 \models (\forall t)[t \in M_2 \Longrightarrow [t \in X_1 \iff t \, E_2 \, X_2]]; \tag{B-24}$$

and computing the relativization, this means that

$$(\forall t \in M_1)[t \in M_2 \Longrightarrow [t \, E_1 \, X_1 \iff t \, E_2 \, X_2]]. \tag{B-25}$$

Compute now, for any $t \in M_2 \subseteq M_1$:

$$\begin{aligned} t \in Y &\iff t \, E_1 \, X_1 \quad \text{by (B-23)}, \\ &\iff t \, E_2 \, X_2 \quad \text{by (B-25)}, \end{aligned}$$

which proves (B-22). The fact that at most one $X_2 \in S_2$ can satisfy (B-22) for every $t \in M_2$ is proved similarly.   ⊣

REMARK: Notice that in relativizing (B-19) in this proof, we left the clause $t \in M_2$ alone. In computing relativizations, "primitive" propositions of the form $P(x_1, \ldots, x_n)$ (which express that a definite condition P holds of the objects $x_1, \ldots, x_n$) are their own relativizations.

This simple theorem makes it possible to construct universes within universes within universes, each time using the properties of the model just constructed. Consider, for example, the next Corollary of **B.16**.

B.18. Proposition. *There exists a* **ZFDC**-*universe* $\mathcal{M}$ *which has exactly one atom.*

PROOF. Suppose c is an atom and let

$$M_c =_{df} \{x \mid (\forall t \in \mathrm{TC}(x))[\neg\mathrm{Set}(t) \Longrightarrow t = c]\} \qquad \text{(B-26)}$$

be the class of objects *supported by* $\{c\}$ in the sense of Problem **11.38**. It is quite easy to verify that M_c is transitive and for each set X,

$$X \subseteq M_c \Longrightarrow X \in M_c,$$

so the natural universe M_c has the Rieger property and by **B.11** it is a model of **ZFDC**; and it is quite clear that it has exactly one atom, c.

So far so good, as long as there exists at least one atom, which may or may not be true in our intended domain $\mathcal{W}$. But there are lots of atoms in the Rieger universe $\mathcal{M}_{at}$ of **B.16**, so what we need to do is *to interpret the proof of the preceding paragraph in the universe* $\mathcal{M}_{at}$. This argument runs as follows.

Let

$$\phi(M, S, E, c) \Longleftrightarrow_{df} \mathrm{Setuniv}(M, S, E) \,\&\, \mathrm{Rieger}(M, S, E) \qquad \text{(B-27)}$$
$$\&\, \mathrm{Atom}(c) \,\&\, (\forall t \in M)[t \notin S \Longrightarrow t = c]$$

be the proposition which asserts of M, S, E, c that they have the properties we are interested in, and let

$$\theta \Longleftrightarrow_{df} (\exists M)(\exists S)(\exists E)(\exists c)\phi(M, S, E, c) \qquad \text{(B-28)}$$

be the proposition which asserts that some M, S, E, c with these properties exist. We have proved θ from the hypothesis that some atom exists, and other than that we have only used the axioms of **ZFDC**—what else is there! Hence this θ is true in every universe of **ZFDC** which has an atom, in particular $\mathcal{M}_{at}$, i.e.,

$$\mathcal{M}_{at} \models \theta.$$

This means that for some $\mathcal{M}_{at}$-classes M, S, some binary $\mathcal{M}_{at}$-condition E and some $\mathcal{M}_{at}$-object c,

$$\mathcal{M}_{at} \models \phi(M, S, E, c),$$

and in particular

$$\mathcal{M}_{at} \models \Big[\mathrm{Setuniv}(M, S, E) \,\&\, \mathrm{Rieger}(M, S, E)\Big],$$
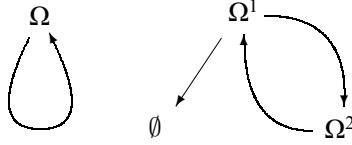
FIGURE **B.1**. Two decorated, ill-founded graphs.

so by **B.17**, $\mathcal{M} = M, S, E$ is a Rieger universe. In addition

$$\mathcal{M}_{\mathrm{at}} \models \Big[\mathrm{Atom}(c) \,\&\, (\forall t \in M)[t \notin S \Longrightarrow t = c] \Big],$$

which means precisely that

$$\mathcal{M} \models \mathrm{Atom}(c) \,\&\, (\forall t)[\neg \mathrm{Set}(t) \Longrightarrow t = c],$$

the required conclusion that $\mathcal{M}$ believes that exactly one atom exists. ⊣

It is not hard to manufacture Rieger universes with various types of ill founded sets, by a combination of the techniques in **B.16** and **B.17**. Some of the problems are about such results. Here we will concentrate on the construction of Aczel's *Antifounded Universe* $\mathcal{A}$, which has a rich variety of ill founded sets with well understood structure.

The idea for $\mathcal{A}$ comes from the Mostowski Collapsing Lemma **11.36**, which gives a "structural" characterization of pure, grounded sets. Recall that by **11.35**, a decoration of a graph G is any surjection $d : G \twoheadrightarrow d[G]$ such that

$$d(x) = \{d(y) \mid y \leftarrow x\} \quad (x \in G), \tag{B-29}$$

where $\rightarrow$ is the edge relation on G and $\leftarrow$ is its inverse,

$$y \leftarrow x \iff y \text{ is a child of } x \iff x \rightarrow y.$$

Each grounded graph G admits a unique decoration d_G, and the pure, grounded sets are all the values $d_G(x)$ of these decorations. Can we also "decorate" the nodes of ill founded graphs to get pure, ill founded sets which are related to ill founded graphs in the same way that pure, grounded sets are related to grounded graphs?

B.19. Antifoundation Principle, AFA. *Every graph admits a unique decoration*.

In Figure **B.1** we have labeled the nodes of two ill founded graphs by the values of their unique decorations, assuming that such exist. By the definition of decoration,

$$\Omega = \{\Omega\}, \quad \Omega^1 = \{\emptyset, \Omega^2\}, \quad \Omega^2 = \{\Omega^1\}, \tag{B-30}$$

i.e., Ω, Ω^1 and Ω^2 are the "ultimately frustrating gifts" we discussed in **11.32**. We can refer to "the" frustrating gifts, because, in fact, the equations in (B-30)—or the graphs in Figure **B.1**—characterize these sets under **AFA**, as follows.

B.20. Proposition. (**AFA**) (1) *There is exactly one set* Ω *which is its own singleton.* (2) *There is exactly one pair of sets* Ω^1, Ω^2 *such that* $\Omega^1 = \{\emptyset, \Omega^2\}$ *and* $\Omega^2 = \{\Omega^1\}$.

PROOF. (1) If $X = \{X\}$ and $Y = \{Y\}$, then we can use either X or Y to decorate the single node graph in Figure **B.1**; but this graph has only one decoration by **AFA**, so $X = Y$. The proof of (2) is similar. ⊣

This "uniqueness" part of the Antifoundation Principle we just applied makes it possible to specify and analyze the structure of ill founded sets with diverse properties, and is the main advantage of the antifounded universe $\mathcal{A}$ over other models which contain ill founded sets. We now proceed to its construction.

B.21. Definition. A **pointed graph** is a pair (G, p_G) of a graph and a node in it, in full detail, a structured set $(G, \rightarrow_G, p_G)$ where $p_G \in G$ and $\rightarrow_G$ is a binary relation on the field G. The designated node p_G is the **point** of the pointed graph.

B.22. Pictures. A pointed graph (G, p) is a **picture** of a set A, if there exists a decoration $d : G \rightarrow d[G]$ of G such that $d_G(p) = A$. The **canonical picture** of a pure set A is the pointed graph $(\mathrm{TC}(A), \ni, A)$, where $\mathrm{TC}(A)$ is the transitive closure of A and $\ni$ is the restriction of the inverse membership condition to $\mathrm{TC}(A)$. This is a picture of A, because the identity function $d(x) = x$ is obviously a decoration of it, $A \in \mathrm{TC}(A)$ and $d(A) = A$.

B.23. A bijection $\pi : G \rightarrowtail\!\!\!\rightarrow H$ between two graphs is an **isomorphism** if it respects the edge relations,

$$x \rightarrow_G y \iff \pi(x) \rightarrow_H \pi(y) \quad (x, y \in G);$$

and an **isomorphism between two pointed graphs** (G, p) and (H, q) is a graph isomorphism $\pi : G \rightarrow H$ such that $\pi(p) = q$. We call G **isomorphic** with H if there exists an isomorphism $\pi : G \rightarrowtail\!\!\!\rightarrow H$ of the appropriate kind.

It is easy to construct non-isomorphic pointed graphs which picture the same set, even grounded ones, e.g., see Figure **B.2** where we have labelled the nodes with the values of the unique decorations. On the other hand, we would expect that if a pointed graph (G, p) admits a unique decoration d_G, then the set $A = d_G[G]$ captures some important invariant of (G, p). The next fundamental definition identifies that invariant.

B.24. Definition. A relation $R \subseteq G \times H$ is a **bisimulation** between two pointed graphs $(G, \rightarrow_G, p_G)$ and $(H, \rightarrow_H, p_H)$, if it relates the points, i.e., $p_G \, R \, p_H$ and also satisfies the implication

$$x \, R \, y \implies (\forall u \,_G\!\leftarrow x)(\exists v \,_H\!\leftarrow y)u \, R \, v \qquad \text{(B-31)}$$
$$\& \, (\forall v \,_H\!\leftarrow y)(\exists u \,_G\!\leftarrow x)u \, R \, v.$$
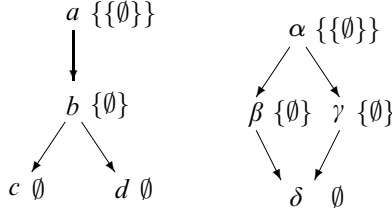
FIGURE **B.2**. Non-isomorphic, bisimilar, grounded graphs.

Two pointed graphs G, H are **bisimilar** if some bisimulation between them exists,

$$G =_{\text{bs}} H \iff_{\text{df}} (\exists R \subseteq G \times H)[R \text{ is a bisimulation}]. \qquad (\text{B-32})$$

As usual with structured sets, we will often refer to "a pointed graph G", skipping the explicit reference to the edge relation $\rightarrow_G$ or the point when it is obvious or irrelevant—we already did this in (B-32).

B.25. Exercise. *Isomorphic pointed graphs are bisimilar, and so are the non-isomorphic, grounded, pointed graphs G and H in Figure **B.2**.*

B.26. Exercise. *If a is a minimal node in a graph G and b is a minimal node in H, then $\{(a,b)\}$ is a bisimulation of the pointed graphs (G,a) and (H,b).*

B.27. Exercise. *Let L be the "single loop" graph on a singleton $\{a\}$, with the one edge pair (a,a), and on the set of integers $\mathbb{N}$ define the successor edge relation*

$$n \rightarrow_s m \iff_{\text{df}} n + 1 = m.$$

Show that the relation $\{(a,i) \mid i \in \mathbb{N}\}$ is a bisimulation of (L,a) with $(\mathbb{N},n)$, for every n.

B.28. Lemma. *The condition $=_{\text{bs}}$ is an equivalence condition on the class of all pointed graphs.*

PROOF. Of the three properties of an equivalence condition (defined in **12.42**), only the transitivity of $=_{\text{bs}}$ is not immediate. To prove that, suppose G_1, G_2 and G_3 are pointed graphs, R_1 is a bisimulation of G_1 with G_2 and R_2 is a bisimulation of G_2 with G_3, and let

$$x \, R \, z \iff_{\text{df}} (\exists y)[x \, R_1 \, y \ \& \ y \, R_2 \, z] \qquad (\text{B-33})$$

be the "product relation" of R_1 and R_2. It is clear that R relates the points p_1 and p_3 of G_1 and G_3, because $p_1 \, R_1 \, p_2$ and $p_2 \, R_2 \, p_3$ hold. Suppose that $x \, R \, z$, so there exists some $y \in G_2$ such that $x \, R_1 \, y$ and $y \, R_2 \, z$. If $u \, _1\!\leftarrow x$, then by (B-31) for R_1, there exists some $v \, _2\!\leftarrow y$ such that $u \, R_1 \, v$; and then by (B-31) for R_2, there exists some $w \, _3\!\leftarrow z$ such that $v \, R_2 \, w$, which together with $u \, R_1 \, v$ establish $u \, R \, w$. This is half of (B-31), and the other half is equally easy.   ⊣

With these definitions, we can now prove that for a grounded graph G and any $p \in G$, the properties of (G,p) coded by the value $d_G(p)$ of its unique decoration are exactly those preserved under bisimulation of pointed graphs.
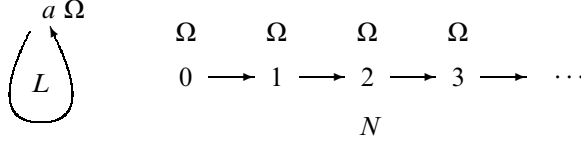
FIGURE **B.3**. Many pictures of $\Omega : (L, a)$ and each $(\mathbb{N}, n)$.

B.29. Theorem. *For all grounded graphs G and H with associated decorations d_G and d_H, and for all $p \in G$ and $q \in H$,*

$$d_G(p) = d_H(q) \iff (G, p) =_{\mathrm{bs}} (H, q). \qquad (\text{B-34})$$

PROOF. We verify first that the relation

$$R = \{(x, y) \in G \times H \mid d_G(x) = d_H(y)\} \qquad (\text{B-35})$$

satisfies (B-31), as follows:

$$x \, R \, y \implies \{d_G(u) \mid u \,_{G}{\leftarrow}\, x\} = \{d_H(v) \mid v \,_{H}{\leftarrow}\, y\} \text{ by the def. of decoration}$$
$$\implies (\forall u \,_{G}{\leftarrow}\, x)(\exists v \,_{H}{\leftarrow}\, y)[d_G(u) = d_H(v)]$$
$$\& \; (\forall v \,_{H}{\leftarrow}\, y)(\exists u \,_{G}{\leftarrow}\, x)[d_G(u) = d_H(v)]$$
$$\implies (\forall u \,_{G}{\leftarrow}\, x)(\exists v \,_{H}{\leftarrow}\, y)u \, R \, v \; \& \; (\forall v \,_{H}{\leftarrow}\, y)(\exists u \,_{G}{\leftarrow}\, x)u \, R \, v.$$

Hence, if $p \in G$, $q \in H$, and $d_G(p) = d_H(q)$, then the relation R of (B-35) establishes that $(G, p) =_{\mathrm{bs}} (H, q)$.

For the converse, suppose towards a contradiction that p is minimal in G such that there exists some $q \in H$ and a bisimulation R of the pointed graphs (G, p) and (H, q), but $d_G(p) \neq d_H(q)$. Now

$$(\forall u \,_{G}{\leftarrow}\, p)(\exists v \,_{H}{\leftarrow}\, q)u \, R \, v,$$

hence,

$$(\forall u \,_{G}{\leftarrow}\, p)(\exists v \,_{H}{\leftarrow}\, q)[d_G(u) = d_H(v)]$$

by the choice of p, and, similarly, $(\forall v \,_{H}{\leftarrow}\, q)(\exists u \,_{G}{\leftarrow}\, p)[d_G(u) = d_H(v)]$, which proves $d_G(p) = d_H(q)$, contradicting the choice of p. $\dashv$

It is a crucial property of **AFA** that it yields the same characterization of bisimulation for all graphs, by quite a different argument.

B.30. Theorem (Aczel). (**AFA**) *For all graphs G and H with associated decorations d_G and d_H, and for all $p \in G$ and $q \in H$,*

$$d_G(p) = d_H(q) \iff (G, p) =_{\mathrm{bs}} (H, q). \qquad (\text{B-36})$$

PROOF. The left-to-right implication in (B-36) is proved exactly as in **B.29**, as that part of the argument did nor depend on the given graphs being grounded.

Suppose now that the edge relations of G and H are $\to_G$ and $\to_H$ and $R \subseteq G \times H$ is a bisimulation of G with H. We can turn R into a pointed graph, with point the pair (p_G, p_H) and edge relation the product of $\to_G$ and $\to_H$:

$$(p, q) \to (u, v) \iff_{\mathrm{df}} p \to_G u \; \& \; q \to_H v.$$

If d_G is the unique decoration of the graph G (forgetting the point) given by **AFA**, define on R the function

$$d_G^R(p, q) =_{df} d_G(p),$$

and compute:

$$\begin{aligned}
x \in d_G^R(p, q) &\iff x \in d_G(p) \\
&\iff (\exists u \;{}_G{\leftarrow}\; p)[x = d_G(u)] \\
&\iff (\exists u \;{}_G{\leftarrow}\; p)(\exists v \;{}_H{\leftarrow}\; q)[u \, R \, v \, \& \, x = d_G(u)] \quad \text{(B-37)} \\
&\iff (\exists (u, v) \leftarrow (p, q))[x = d_G^R(u, v)],
\end{aligned}$$

where the key equivalence (B-37) holds because R is a bisimulation and $p \, R \, q$, and hence for each $u \;{}_G{\leftarrow}\; p$, there exists some $v \;{}_H{\leftarrow}\; q$ satisfying $u \, R \, v$. Thus, the function d_G^R is a decoration of R, and the corresponding extension

$$d_H^R(p, q) =_{df} d_H(q)$$

of the decoration d_H of H is also a decoration of R, by the same argument. By **AFA** then, for all $(p, q) \in R$,

$$d_G(p) = d_G^R(p, q) = d_H^R(p, q) = d_H(q),$$

which completes the proof.   ⊣

This characterization under **AFA** of the properties of (G, p) which are coded into the value $d_G(p)$ of its unique decoration, suggests a method for the construction of $\mathcal{A}$.

B.31. The Antifounded Universe. Let

$$\mathcal{A}_0 =_{df} \{(G, \to_G, p_G) \in \mathcal{V} \mid \to_G \subseteq G \times G \, \& \, p_G \in G\} \quad \text{(B-38)}$$

be the class of all pointed graphs on pure, grounded sets, and on $\mathcal{A}_0$ define the binary definite condition

$$\begin{aligned}
(G, p_G) \, &\varepsilon_0 \, (H, p_H) \\
&\iff_{df} (\exists q \in H)[q \;{}_H{\leftarrow}\; p_H \, \& \, (G, p_G) =_{bs} (H, q)], \quad \text{(B-39)}
\end{aligned}$$

skipping the edge relations in the notation.

First we note that ε_0 respects bisimulation:

$$G_1 \, \varepsilon_0 \, H_1 \, \& \, G_1 =_{bs} G_2 \, \& \, H_1 =_{bs} H_2 \implies G_2 \, \varepsilon_0 \, H_2. \quad \text{(B-40)}$$

To prove this, suppose $\to_1$, p_1 are the edge relation and the designated node of H_1, and similarly with $\to_2$, p_2 for H_2. The hypothesis of (B-40) gives us some $q_1 \;{}_1{\leftarrow}\; p_1$ such that

$$G_2 =_{bs} G_1 =_{bs} (H_1, q_1), \quad \text{(B-41)}$$

and a bisimulation R of (H_1, p_1) with (H_2, p_2). By the basic property of bisimulations, there must exist some $q_2 \;{}_2{\leftarrow}\; H_2$ such that $q_1 \, R \, q_2$; this means that R is a bisimulation of (H_1, q_1) with (H_2, q_2), and then (B-41) with the transitivity of $=_{bs}$ gives $G_2 =_{bs} (H_2, q_2)$, hence $G_2 \, \varepsilon_0 \, H_2$.

Now each $G \in \mathcal{A}_0$ is a pure, grounded *set* (a triple in $\mathcal{V}$), even if it is ill founded *as a graph*, and the bisimulation condition $=_{\mathrm{bs}}$ is an equivalence condition on $\mathcal{A}_0$ by **B.28**. By Problem **x12.45**, there exists a definite operation α which is determining for $=_{\mathrm{bs}}$, i.e., for $G, H \in \mathcal{A}_0$

$$G =_{\mathrm{bs}} H \iff \alpha(G) = \alpha(H).$$

The domain of the antifounded universe is the quotient class of $\mathcal{A}_0$ by $=_{\mathrm{bs}}$,

$$\mathcal{A} =_{\mathrm{df}} \{\alpha(G) \mid G \in \mathcal{A}_0\}. \tag{B-42}$$

We define on $\mathcal{A}$ the membership condition

$$x \,\varepsilon\, y \iff_{\mathrm{df}} (\exists G, H)[x = \alpha(G) \,\&\, y = \alpha(H) \,\&\, G \,\varepsilon_0\, H], \tag{B-43}$$

unambiguously by (B-40), and finally we take the *Pure Antifounded Universe* to be the triple $\mathcal{A}, \mathcal{A}, \varepsilon$. We will refer to it by the name of its domain $\mathcal{A}$, which is also the collection of its sets—there are no atoms in $\mathcal{A}$.

B.32. Theorem (Aczel). (**AC**) $\mathcal{A}$ *is a Rieger universe, which further satisfies the Antifoundation Principle* **AFA**, *the Axiom of Choice* **AC** *and the Principle of Purity.*

Proof. The key property of $\mathcal{A}$ is that for each graph $H \in \mathcal{V}$ with edge relation $\rightarrow_H$ and each node $p \in H$,

$$\boldsymbol{b}_{\mathcal{A}}(\alpha(H, p)) = \{\alpha(H, q) \mid q \,_H{\leftarrow}\, p\}, \tag{B-44}$$

which follows from the following trivial computation:

$$
\begin{aligned}
x \in \boldsymbol{b}_{\mathcal{A}}(\alpha(H, p)) &\iff (\exists G \in \mathcal{A}_0)(\exists q \,_H{\leftarrow}\, p)[x = \alpha(G) \,\&\, G =_{\mathrm{bs}} (H, q)] \\
&\iff (\exists q \,_H{\leftarrow}\, p)[x = \alpha(H, q)].
\end{aligned}
$$

This implies, in particular, that each $\boldsymbol{b}_{\mathcal{A}}(x)$ is a set, so $\mathcal{A}$ is a set universe.

For the Rieger property, suppose $Y \subseteq \mathcal{A}$ and (using **AC**) choose for each $y \in Y$ a pointed graph $G_y \in \mathcal{A}_0$, such that (1) $\alpha(G_y) = y$. By replacing each G_y by an isomorphic copy if necessary, we can also ensure that (2) $y \neq z \Longrightarrow G_y \cap G_z = \emptyset$, and (3) for all $y \in Y, \emptyset \notin G_y$. Let

$$
\begin{aligned}
H &=_{\mathrm{df}} \bigcup\{G_y \mid y \in Y\} \cup \{\emptyset\}, \\
u \rightarrow_H v &\iff_{\mathrm{df}} (\exists y \in Y)[u \rightarrow_y v \vee (u = \emptyset \,\&\, v = p_y)],
\end{aligned}
$$

where $\rightarrow_y$ and p_y are the edge relation and the point of G_y. The pointed graph H with edge relation $\rightarrow_H$ and point $\emptyset$ is obviously in $\mathcal{A}_0$, and for each $y \in Y$,

$$(H, p_y) =_{\mathrm{bs}} G_y \tag{B-45}$$

by the trivial (identity) bisimulation

$$\{(u, v) \in H \times G_y \mid u = v\};$$

thus, by (B-44), and the definition,

$$\boldsymbol{b}_A(\alpha(H, \emptyset)) = \{\alpha(H, q) \mid q \mathrel{{}_H\!\leftarrow} \emptyset\}$$
$$= \{\alpha(H, p_y) \mid y \in Y\}$$
$$= \{\alpha(G_y) \mid y \in Y\} \quad \text{by (B-45)}$$
$$= Y.$$

To prove the uniqueness of $\alpha(H)$, suppose H' is any pointed graph in $\mathcal{A}_0$ with edge relation $\rightarrow'$ and point q' such that

$$G \,\varepsilon_0\, H' \iff \alpha(G) \in Y.$$

By (B-44) again,

$$\boldsymbol{b}_A(\alpha(H', q')) = \{\alpha(H', q) \mid q \mathrel{'\!\leftarrow} q'\}$$
$$= \{\alpha(H, p_y) \mid y \in Y\} \quad \text{by hyp.}$$

Thus, for each $y \in Y$,

$$y = \alpha(G_y) = \alpha(H, p_y)\,\varepsilon\,\alpha(H'),$$

and we can choose (by **AC**) some $q_y \mathrel{'\!\leftarrow} q'$ and a bisimulation S_y of (H, p_y) with (H', q_y); and conversely, by the same argument, for each $q \mathrel{'\!\leftarrow} q'$ we can choose some $y_q \in Y$ and some bisimulation T_q of (H, q_y) with (H', q). It is now easy to verify that the union

$$R = \bigcup\{S_y \mid p_y \mathrel{{}_H\!\leftarrow} \emptyset\} \cup \{T_q \mid q \mathrel{'\!\leftarrow} q'\} \cup \{(\emptyset, q')\}$$

is a bisimulation which establishes that $H =_{\mathrm{bs}} H'$, i.e., $\alpha(H) = \alpha(H')$.

Finally, to verify **AFA** for $\mathcal{A}$, suppose G is a graph in $\mathcal{A}$ with edge relation $\rightarrow_G \in \mathcal{A}$. To prove that G admits a decoration in $\mathcal{A}$, it is enough to define an $\mathcal{A}$-operation δ such that

$$\mathcal{A} \models (\forall p \in G)[\delta(p) = \{\delta(q) \mid q \mathrel{{}_G\!\leftarrow} p\}], \tag{B-46}$$

since $\mathcal{A}$ is a **ZFDC**-universe, so it "knows" from (B-46) that the restriction of δ to G is a function, which is then a decoration of G. Let

$$H =_{\mathrm{df}} \boldsymbol{b}_A(G) \in \mathcal{V},$$

and make H into a graph in $\mathcal{V}$ with the edge relation

$$x \rightarrow_H y \iff_{\mathrm{df}} \mathcal{A} \models x \rightarrow_G y \quad (x, y \in H). \tag{B-47}$$

For each $p \in H$, set

$$\delta(p) =_{\mathrm{df}} \alpha(H, p) \in \mathcal{A} \quad (p \in H) \tag{B-48}$$

and compute:

$$\boldsymbol{b}_A(\delta(p)) = \{\alpha(H, q) \mid q \mathrel{{}_H\!\leftarrow} p\} \qquad \text{by (B-44)}$$
$$= \{\alpha(H, q) \mid \mathcal{A} \models q \mathrel{{}_G\!\leftarrow} p\} \quad \text{by (B-47)}$$
$$= \{\delta(q) \mid \mathcal{A} \models q \mathrel{{}_G\!\leftarrow} p\} \qquad \text{by (B-48).}$$

Put another way, for each $p \, \varepsilon \, G$,

$$x \varepsilon \delta(p) \iff \mathcal{A} \models (\exists q \; {}_G{\leftarrow} \, p)[x = \delta(q)],$$

which is equivalent to (B-46).

It remains to show that G admits at most one decoration in $\mathcal{A}$, and for this it suffices (as above) to show that if δ' is any $\mathcal{A}$-operation such that

$$\mathcal{A} \models (\forall p \in G)[\delta'(p) = \{\delta'(q) \mid q \; {}_G{\leftarrow} \, p\}], \qquad \text{(B-49)}$$

then $\delta'(p) = \delta(p)$, for every $p \in H$. Given such a δ', choose (by **AC**) a pointed graph H_p' with point r_p for each $p \in H$, such that

$$\delta'(p) = \alpha(H_p') \quad (p \in H),$$

and make sure as in the proof of the Rieger property that these graphs are all pairwise disjoint. If $H' = \bigcup \{H_p' \mid p \in H\}$ is the union of all the graphs, then

$$\delta'(p) = \alpha(H', r_p) \quad (p \in H),$$

since (trivially) the identity relation $\{(q, q) \mid q \in H\}$ is a bisimulation of (H_p', r_p) with (H, r_p). We now claim that *the relation*

$$R =_{\mathrm{df}} \{(p, s) \in H \times H' \mid \alpha(H', s) = \delta'(p)\}$$

is a bisimulation of (H, p) *with* (H', r_p) *for each* $p \in H$. This will complete the proof, because for $p \in H$, $\alpha(H', r_p) = \delta'(p)$, hence $p \, R \, r_p$, and hence

$$\delta(p) = \alpha(H, p) = \alpha(H', r_p) = \delta'(p).$$

To show the somewhat less trivial half of the italicized statement, let $\to'$ be the edge relation of H', assume $p \, R \, s$ and compute:

$$
\begin{aligned}
t \; '{\leftarrow} \, s &\implies \alpha(H', t) \varepsilon \, \alpha(H', s) \\
&\implies \alpha(H', t) \varepsilon \, \delta'(p) && \text{because } p \, R \, s \\
&\implies \text{for some } q \; {}_H{\leftarrow} \, p, \alpha(H', t) = \delta'(q) && \text{by (B-49)} \\
&\implies \text{for some } q \; {}_H{\leftarrow} \, p, q \, R \, t.
\end{aligned}
$$

The Axiom of Choice for $\mathcal{A}$ follows from **B.14**, and the Principle of Purity is trivial.   ⊣

Problems for Appendix B

xB.1. Prove that for each set universe $\mathcal{M} = M, S, E$, the axioms for definite conditions and operations listed in **3.18** become true, if we replace in them "condition" by "$\mathcal{M}$-condition", "operation" by "$\mathcal{M}$-operation", $\in$ by E, Set by S and $(\forall y)$ by $(\forall y \in M)$.

****xB.2.** Suppose pairs and Cartesian products are defined by the Kuratowski operation of **4.2**. Show that $\bigcup_{n=2}^{\infty} \{\mathbb{N}_0\}^n \notin \mathcal{Z}$ and infer that the following

proposition is not a theorem of **ZDC**: *for each set A, there exists a function* $f : \mathbb{N} \to f[A]$ *such that*

$$f(0) = A \times A, \quad f(n+1) = f(n) \times A.$$

***xB.3.** Construct a definite operation $(x, y)'$ with the following properties. (1) $(x, y)'$ is an ordered pair operation, i.e., it satisfies (OP1) and (OP2) of **4.1**. (2) If $X, Y \in \mathcal{Z}$, then their Cartesian product $X \times Y$ is also in $\mathcal{Z}$. (3) If $\bigcup_{n=2}^{\infty} A^n$ is defined using this pair, then for each $A \in \mathcal{Z}$, $\bigcup_{n=2}^{\infty} A^n \in \mathcal{Z}$.

***xB.4.** Show that the implication (OP1) $\Longrightarrow$ (OP2) in **4.1** is not a theorem of **ZDC** for any definite operation (x, y).

***xB.5.** Verify that if I is a transitive set, then

$$A \in M(I) \Longrightarrow \mathrm{TC}(A) \in M(I).$$

xB.6. For each I, define $K_n(I)$ by the recursion

$$K_0(I) = I, \quad K_{n+1}(I) = K_n(I) \cup \mathcal{P}(K_n(I)). \tag{B-50}$$

Show that

$$M_n(I) \subseteq K_n(\mathrm{TC}_{n+1}(I)) \subseteq M(I),$$

where $\mathrm{TC}_n(I)$ and $M_n(I)$ are defined by (11-16) and (11-18), respectively.

***xB.7.** Find some $I \supset \mathbb{N}_0$ such that $\mathrm{TC}(I) \notin M(I)$. Infer that **ZDC** cannot prove that "every set has a transitive closure".

***xB.8.** The implication (C1) $\Longrightarrow$ (C3) in **4.20** cannot be proved for an arbitrary, definite operation $|A|$ in **ZDC**. (Cf. Problem **x11.6**.)

***xB.9.** The equivalence in Problem **x11.7** is not a theorem of **ZDC**.

***xB.10.** Assume that the full Axiom of Choice and the Generalized Continuum Hypothesis is true, so for all cardinals κ, $2^{\kappa} =_c \kappa^+$. Prove that

$$(n \mapsto \aleph_n) \subseteq \mathcal{Z}, \quad (n \mapsto \aleph_n) \notin \mathcal{Z}.$$

Infer that **ZDC**+**AC** cannot prove the existence of an infinite, increasing sequence of infinite cardinals, i.e., the proposition

$$\theta : (\exists f : \mathbb{N} \to f[\mathbb{N}])(\forall n \in \mathbb{N})[\mathbb{N} \leq_c f(n) <_c f(n+1)].$$

***xB.11.** Show that **ZDC** cannot prove the proposition "the well ordered set $\mathbb{N}$ of integers is similar with an ordinal". Hint. Use Problem **x12.27**. The less trivial part of the problem is how to compute (or avoid computing) the relativization of this fairly complex proposition.

***xB.12.** (**AC**) Show that **ZFC** cannot prove that strongly inaccessible cardinals exist. Hint. Go by contradiction and interpret the meaning of the alleged theorem in $\mathcal{V}_\kappa$, where κ is the least strongly inaccessible cardinal.

xB.13. An **ordered pair operation** in a Rieger universe $\mathcal{M}$ is any binary $\mathcal{M}$-operation C such that for all $x, y, x', y' \in M$,

$$C(x, y) = C(x', y') \iff x = x' \,\&\, y = y'. \tag{B-51}$$

Cartesian products and function spaces relative to C are defined by

$$A \times_C B =_{\text{df}} \rho\{C(x, y) \mid x \,E\, A, y \,E\, B\},$$
$$(A \to_C B) =_{\text{df}} \rho\{f \in M \mid (\forall t \,E\, f)[t \,E\, A \times_C B]$$
$$\&\, (\forall x \,E\, A)(\exists! y \,E\, B)[C(x, y) \,E\, f]\},$$

where $\rho(Y)$ is the Rieger operation of $\mathcal{M}$ defined in (B-14). Verify that these definitions make sense (i.e., ρ is applied to appropriate arguments) and hence $A \times_C B$ and $A \to_C B$ are $\mathcal{M}$-operations.

****xB.14.** Define triples, structured sets and systems of natural numbers in an arbitrary Rieger universe $\mathcal{M}$, relative to an arbitrary ordered pair operation $C(x, y)$ in $\mathcal{M}$. Formulate the Choice Principles **DC**, **AC**$_\mathbb{N}$ and **AC** using these notions and prove that every Rieger universe $\mathcal{M}$ satisfies **DC**, and if **AC** is also true, then $\mathcal{M}$ also satisfies **AC**.

xB.15. Show that the Rieger universe $\mathcal{M}_{\text{at}}$ of **B.16** has an ordered pair operation C such that for all x and y, the "pair" $C(x, y)$ is an atom.

****xB.16.** (**AC**) Define a Rieger universe $\mathcal{M}$ which satisfies the following two propositions.
  (a) There exists a binary, definite condition $\leq_{\text{at}}$ which well orders the class of atoms, in the sense that (1) for all atoms a, b, c,

$$a \leq a, \quad [a \leq b \,\&\, b \leq c] \Longrightarrow a \leq c, \quad [a \leq b \,\&\, b \leq a] \Longrightarrow a = c,$$

(2) for every two atoms a, b, either $a \leq b$ or $b \leq a$, and (3) every non-empty set of atoms has a $\leq$-least member.
  (b) Every set X is equinumerous with an $\leq$-*initial segment* of atoms, i.e., for some atom b, $X =_c \{a \mid \text{Atom}(a) \,\&\, a < b\}$. Hint. Make the ordinals atoms in some Rieger universe.

****xB.17.** Define a Rieger universe which has at least two, distinct self-singletons, i.e., sets a and b such that $a \neq b$, $a = \{a\}$ and $b = \{b\}$. Hint. Start with a universe which has two atoms and imitate the coding construction in **B.16**.

****xB.18.** Define a Rieger universe which contains an infinite sequence of distinct sets $x_0, x_1, \ldots$, such that for each i, $x_i = \{x_{i+1}\}$.

xB.19. Given a graph G and a node $p \in G$, let

$$G \upharpoonright p = \{x \in G \mid x = p \vee p \Rightarrow x\}$$

consist of p and all the nodes on a path below it. Consider $G \upharpoonright p$ as a subgraph of G, with the restriction of the edge relation $\to_G$ to it, and prove that

$$(G, p) =_{\text{bs}} (G \upharpoonright p, p).$$

B.33. Definition. A **partial bisimulation** between two graphs G and H is any relation $R \subseteq G \times H$ which is a bisimulation of the pointed graphs (G, p) and (H, q) for every $(p, q) \in R$; it is a **total bisimulation** if in addition

$$(\forall p \in G)(\exists q \in H)\, p\, R\, q \,\&\, (\forall q \in H)(\exists p \in G)\, p\, R\, q.$$

Two graphs are **bisimilar** if there exists a total bisimulation between them.

xB.20. For all pairs of graphs G, H, there exists a largest (under $\subseteq$) partial bisimulation R between G and H, and $G =_{\text{bs}} H$ if and only if this largest bisimulation is total.

xB.21. Two graphs G, H are bisimilar if and only if every pointed graph (G, p) with $p \in G$ is bisimilar with some (H, q), $q \in H$, and conversely, every (H, q) is bisimilar with some (G, p).

xB.22. (**AFA**) Prove that there exist distinct, pure sets x, y and z such that

$$x \ni y \ni z \ni x,$$

and draw a picture of them.

****xB.23.** (**AFA**) Prove that there are only two, transitive, pure singletons. How many transitive, pure doubletons are there? Draw pictures of them.

xB.24. (**AFA**) With the Kuratowski pair, prove that there exists a pure set x such that

$$x = (\emptyset, x),$$

and draw a picture of it.

xB.25. (**AFA**) With the Kuratowski pair, prove that there exists a pure set x such that

$$x = \{(n, x) \mid n \in \mathbb{N}\},$$

and draw a picture of it.

SOLUTIONS TO THE EXERCISES IN CHAPTERS 1 – 12

2.5. The identity function $(x \mapsto x)$ shows that $A \leq_c A$. And if $f : A \rightarrowtail B$ and $g : B \rightarrowtail C$, then the composition $h(x) = g(f(x))$ is an injection of A into C, so that $A \leq_c C$.

2.8. The assumptions give $A \leq_c \mathbb{N}$ and $B \leq_c A$, so $B \leq_c \mathbb{N}$.

2.9. If A is empty, then $B = f[\emptyset] = \emptyset$. If A is not empty, then there exists a surjection $\pi : \mathbb{N} \twoheadrightarrow A$, and then the composition $h(i) = f(\pi(i))$ is a surjection of $\mathbb{N}$ onto B, so B is countable.

2.20. Suppose $f : A \rightarrowtail\!\!\!\twoheadrightarrow B$ is a bijection, and let $\pi : \mathcal{P}(A) \to \mathcal{P}(B)$ be the image map,

$$\pi(X) = f[X] \quad (X \subseteq A).$$

To prove that π is an injection, note that if $x \in X \setminus Y$, then $f(x) \in f[X]$ but it is not possible that $f(x) \in f[Y]$; because if $f(x) = f(y)$ for some $y \in Y$, then $y = x$ since f is an injection, and so $x = y \in Y$, which is not true. Thus

$$x \in X \setminus Y \Longrightarrow f(x) \in f[X] \setminus f[Y],$$

and by the symmetric argument,

$$y \in Y \setminus X \Longrightarrow f(y) \in f[Y] \setminus f[X].$$

The two implications together show that

$$X \neq Y \Longrightarrow f[X] \neq f[Y],$$

so that the image map is an injection.

To see that the image map is also a surjection, given $Y \subseteq B$, let $X = f^{-1}[Y]$, so that, immediately $f[X] \subseteq Y$; but each $y \in Y$ is the value $f(x)$ of some $x \in A$ since f is a surjection, and then $x = f^{-1}(y) \in X$, so that $f[X] = Y$.

2.23. Given bijections $f : A_1 \rightarrowtail\!\!\!\twoheadrightarrow A_2$ and $g : B_1 \rightarrowtail\!\!\!\twoheadrightarrow B_2$, we let for each $p : A_1 \to B_1$,

$$\pi(p) = q \text{ where, for each } x \in A_2, \ q(x) = g(p(f^{-1}(x))).$$
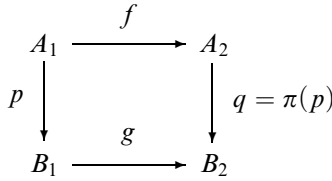
FIGURE 1. Diagram for Exercise 2.23.

In short: q is the unique function which makes the diagram in Figure 1 *commutative*, i.e., such that

$$g(p(x)) = q(f(x)) \qquad (x \in A_1).$$

We include the details of the argument for completeness, but they can all be read from the diagram.

To check that if $p : A_1 \to B_1$, then $\pi(p) = q : A_2 \to B_2$, compute:

$$x \in A_2 \Longrightarrow f^{-1}(x) \in A_1$$
$$\Longrightarrow p(f^{-1}(x)) \in B_1 \Longrightarrow q(x) = g(p(f^{-1}(x))) \in B_2.$$

The proof that π is a bijection is very similar to that of the corresponding Exercise **2.20** for powersets, only a bit simpler.

First, if $p_1, p_2 : A_1 \to B_1$ and $p_1 \neq p_2$, then there is some $y \in A_1$ such that $p_1(y) \neq p_2(y)$. If $x = f(y)$, then

$$p_1(f^{-1}(x)) = p_1(y) \neq p_2(y) = p_2(f^{-1}(x))$$

using the fact that f^{-1} is an injection; and so applying g to both sides and using the fact that g is an injection, we get that

$$\pi(p_1)(x) \neq \pi(p_2)(x),$$

so that $\pi(p_1) \neq \pi(p_2)$. Thus π is an injection. To check that π is also a surjection, define for each $q : A_2 \to B_2$ the function

$$p(y) = g^{-1}(q(f(y))) \quad (y \in A_1),$$

so that $p : A_1 \to B_1$, and for each $x \in A_2$,

$$\pi(p)(x) = g(p(f^{-1}(x))) = g(g^{-1}(q(f(f^{-1}(x))))) = q(x),$$

so that $\pi(p) = q$ and π is a surjection.

3.9. We must show that these two sets do not have the same members, and this is true because $\emptyset \in \{\emptyset\}$ while $\emptyset \notin \emptyset$.

3.13. The empty set is a subset of every set, so $\emptyset \in \mathcal{P}(\emptyset)$; and if $X \subseteq \emptyset$, then X must be empty, since any member of X would have to be a member of $\emptyset$—and there are none such. For the second claim, it is again immediate for every set A that $\emptyset, A \subseteq A$, and so $\{\emptyset, \{\emptyset\}\} \subseteq \mathcal{P}(\{\emptyset\})$; and for the other direction, if $X \subseteq \{\emptyset\}$, then X can have only one member, $\emptyset$, and so $X = \emptyset$ or $X = \{\emptyset\}$.

3.14. Using the powerset and subset axioms, we set

$$B = \{X \in \mathcal{P}(A) \mid (\exists t \in A)[X = \{t\}]\}.$$

3.16. The emptyset has no members, and so $\bigcup \emptyset = \emptyset$; and the only member of the singleton $\{\emptyset\}$ is $\emptyset$, which has no members, so again, $\bigcup \{\emptyset\} = \emptyset$.

3.20. The unary condition P defined by

$$P(x) \iff x \in A$$

is coextensive with A, and so by (3-10), $\{x \mid x \in A\} = A$. Similarly, if

$$Q(X) \iff \text{Set}(X) \,\&\, X \subseteq A,$$

then $Q =_e \mathcal{P}(A)$, and so by (3-10), $\{X \mid \text{Set}(X) \,\&\, X \subseteq A\} = \mathcal{P}(A)$.

3.21. That Set is not a set is part of Theorem **3.11**. For $\mathcal{W}$ we argue by contradiction: if it were a set, so would Set be a set by the Axiom of Subsets, since Set $= \{x \in \mathcal{W} \mid \text{Set}(x)\}$.

3.22. If $A = \{w \mid w \text{ is a singleton}\}$ were a set, then its unionset $\bigcup A$ would also be a set; but every object x is a member of its singleton $\{x\}$, and so $\bigcup A = \mathcal{W}$, which is not a set by the preceding Exercise.

3.23. This is primarily an exercise in terminology, and we can prove it by a simple round-robin argument. We assume that A is a class.

If A is a set, then $A \in \{A\}$ and the singleton $\{A\}$ is a set, and hence a class, so A belongs to a class.

If A belongs to some class B, then it must be an object (atom or set), since only objects are put into classes by (3-8); and since A is a class by the hypothesis, it is not an atom, and so it must be a set. It follows that $A \subseteq A$, and so A is a subset of a set.

Finally, suppose that A is a class and $A \subseteq X$, for some set X. If A is a set, we are done; otherwise, by (3-10), $A = P$ for some definite condition P, but $A \subseteq X$, so that only members of X can satisfy P and $A = \{x \in X \mid P(x)\}$; and this makes A a set by the Subset Axiom, contradicting the case hypothesis.

4.3. If Pair(z), then $z = (x, y)$ for some x, y, and by the definitions,

$$x = \text{First}(z), \, y = \text{Second}(z);$$

and if $z = (\text{First}(z), \text{Second}(z))$, then $z = (x, y)$ with these x, y, and so Pair(x, y) holds.

4.4. If $(x, y, z) = (x, (y, z)) = (x', (y', z')) = (x', y', z')$, then by the basic property of pairs we have $x = x'$ and $(y, z) = (y', z')$; and so applying the same property once more, we have $y = y'$ and $z = z'$. The other direction is trivial.

4.6. $A \uplus \emptyset = \{(\underline{\text{blue}}, x) \mid x \in A\}$, and $(\underline{\text{blue}}, x) \in A \uplus B$ for every $x \in A$ and every B. For the second claim, notice that for any A, B, the members of $A \uplus B$

are ordered pairs; but if we take $A = B = \{\emptyset\}$, then (with the Kuratowski pair), $\emptyset$ is not an ordered pair, and so we do not have $\{\emptyset\} \subseteq \{\emptyset\} \uplus \{\emptyset\}$.

4.9. There is nothing much to prove for (2) either, since $P_{A,B}$ is a set by the Subset Axiom (**III**) and it is a subset of $A \times B$ by its definition—so it is a relation on A, B.

4.11. The claim is trivial for the first two examples. For the third one, the reflexiveness $x \sim_{A/B} x$ and symmetry $x \sim_{A/B} y \Longrightarrow y \sim_{A/B} x$ are again trivial. For the transitivity, observe first that

$$[x \sim_{A/B} y \,\&\, y = z] \Longrightarrow x \sim_{A/B} z;$$

this is because the hypothesis tells us that either $x = y$, and then with $y = z$ we have $x = z$, so that $x \sim_{A/B} z$; or $x, y \in B$, and so with $y = z$ we have $x, z \in B$, i.e., $x \sim_{A/B} z$ again. Suppose now that $x \sim_{A/B} y$ and $y \sim_{A/B} z$, and assume that all three x, y, z are distinct, since each of the assumptions $x = y$, $y = z$ or $y = z$ implies immediately that $x \sim_{A/B} z$ by the observation just made; but if $x \neq y$ and $y \neq z$, the first hypothesis $x \sim_{A/B} y$ gives $x \in B \,\&\, y \in B$, and the second hypothesis $y \sim_{A/B} z$ gives $y \in B \,\&\, z \in B$, so that $x \in B \,\&\, z \in B$, which implies $x \sim_{A/B} z$.

4.13. If $=_A$ is the restriction of the identity relation on A, then each $x \in A$ is equivalent only to itself and so $[x/=_A] = \{x\}$. For the universal relation $\sim_A$ in which $x \sim_A y$ for all $x, y \in A$, clearly $[x/\sim_A] = A$, for any $x \in A$. Finally, for the more complex $\sim_{A/B}$, we have $[x/\sim_{A/B}] = \{x\}$ if $x \in A \setminus B$, and $[x/\sim_{A/B}] = B$ if $x \in B$.

4.15. If f is a function, then there exist sets A, B such that $f \subseteq A \times B$, and so

$$\text{Domain}(f) = \{x \in A \mid (\exists y)[(x, y) \in f]\}$$

and $\text{Domain}(f)$ is a set by the Subset Axiom (**III**). The same argument shows that $\text{Image}(f)$ is a set (a subset of B), and the last implication is trivial.

4.17. The proof is exactly that in Exercise **2.20**, so the point is to read that argument carefully and see exactly what axioms are needed to justify it. Recall that the required equinumerosity between $\mathcal{P}(A)$ and $\mathcal{P}(B)$ was the image function

$$\pi(X) = f[X] \quad (X \subseteq A),$$

so we must check what axioms are needed to construct this function as a set of ordered pairs. First, the values are sets by the Subset Axiom (**III**), since

$$f[X] = \{y \in B \mid (\exists x \in X)[(x, y) \in f]\};$$

but this part of the argument uses an assumed ordered pair operation, and so we are implicitly appealing to all the axioms required to construct such an operation, i.e., those used in the proof of Lemma **4.2**. Looking carefully at that argument, it uses the Pair Axiom (**II**) to define the Kuratowski ordered pair; the Extensionality Axiom (**I**) to prove that this operation satisfies (OP1); and

to prove (OP2), first the Pair and Unionset axioms (**II**) and (**V**) to construct $A \cup B$ and then the Powerset axiom (**IV**). In short, to make precise and to show that if $f : A \to B$, then the image $f[X]$ of any $X \subseteq A$ is a set, we need all the axioms except for the Axiom of Infinity (**VI**).

To finish the proof we define π as a set of ordered pairs:

$$\pi = \{(X, Y) \in \mathcal{P}(A) \times \mathcal{P}(B) \mid (\forall y)[y \in Y \iff (\exists x \in X)[(x, y) \in f]]\}.$$

This is a set by the Subset Axiom (**III**) again, once we have the basic properties of the ordered pair operation as above.

4.18. The detailed proofs for this are almost identical to the argument in Exercise **4.17** and it is not worth repeating them. We confine ourselves to defining the functions required to show these equinumerosities, assuming that $f : A \twoheadrightarrow A'$ and $g : B \twoheadrightarrow B'$.

To show that $A \uplus B =_c A' \uplus B'$, we set

$$\pi((i, x)) = \begin{cases} (\underline{\text{blue}}, f(x)), & \text{if } i = \underline{\text{blue}}, \\ (\underline{\text{white}}, g(x)), & \text{otherwise, i.e., if } i = \underline{\text{white}}. \end{cases}$$

To show that $A \times B =_c A' \times B'$ we set

$$\pi((x, y)) = (f(x), g(y)).$$

Finally, to show that $(A \to B) =_c (A' \to B')$, we set for $p : A \to B$,

$$\pi(p) = \{(x, y) \in A' \times B' \mid y = g(p(f^{-1}(x)))\},$$

following the proof in Exercise **2.23**.

4.22. Using successively (C1), the hypothesis and (C1) again,

$$A =_c |A| = |B| =_c B.$$

4.23. Given the two bijections $f : A \twoheadrightarrow |A|$ and $g : B \twoheadrightarrow |B|$, we define $h : A \cup B \rightarrowtail A \uplus B$ by

$$h(x) = \begin{cases} (\underline{\text{blue}}, f(x)), & \text{if } x \in A, \\ (\underline{\text{white}}, g(x)), & \text{otherwise (i.e., if } x \in B \setminus A). \end{cases}$$

If $A \cap B = \emptyset$, then the two cases in the definition of h are mutually exclusive, and so h is a bijection which witnesses

$$A \cup B =_c A \uplus B =_c |A| + |B|.$$

4.24. These identities follow immediately from the definitions, the basic (C1) and Exercise **4.18**: e.g.,

$$\kappa_1 + \kappa_2 =_c \kappa_1 \uplus \kappa_2 =_c \lambda_1 \uplus \lambda_2 (\text{by Exercise } \mathbf{4.18}) =_c \lambda_1 + \lambda_2.$$

4.26. If κ is a cardinal number, then $\kappa = |A|$ for some set A, and by (C1), $\kappa =_c A$; but then (C2) (which holds of a strong cardinal assignment) implies that $|\kappa| = |A|$, i.e., $|\kappa| = \kappa$. For the non-trivial direction of the second claim: if $\kappa =_c \lambda$, then $|\kappa| = |\lambda|$ by (C2), and so $\kappa = \lambda$ by the first claim.
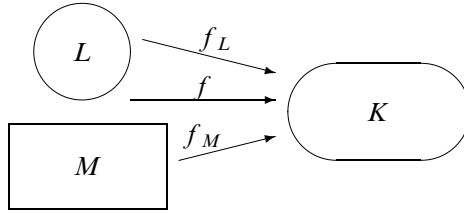
FIGURE 2. Diagram for Exercise **4.28**.

4.27. By the definition, each member of $K \uplus (L \uplus M)$ is in exactly one of these forms, where k, l, m are uniquely determined members of K, L, M respectively:

$$x = (\underline{\text{blue}}, k), \text{ or } x = (\underline{\text{white}}, (\underline{\text{blue}}, l)), \text{ or } x = (\underline{\text{white}}, (\underline{\text{white}}, m)).$$

In the same way, each member of $(K \uplus L) \uplus M$ is in exactly one of the following forms, where k, l, m are uniquely determined members of K, L, M respectively:

$$y = (\underline{\text{blue}}, (\underline{\text{blue}}, k)), \text{ or } y = (\underline{\text{blue}}, (\underline{\text{white}}, l)), \text{ or } y = (\underline{\text{white}}, m).$$

The required bijection $f : K \uplus (L \uplus M) \rightarrowtail\!\!\!\rightarrow (K \uplus L) \uplus M$ is defined by matching the corresponding elements in the obvious way:

$$f(\underline{\text{blue}}, k) = (\underline{\text{blue}}, (\underline{\text{blue}}, k)),$$
$$f(\underline{\text{white}}, (\underline{\text{blue}}, l)) = (\underline{\text{blue}}, (\underline{\text{white}}, l)),$$
$$f(\underline{\text{white}}, (\underline{\text{white}}, m)) = (\underline{\text{white}}, m).$$

4.28. If $L \cap M = \emptyset$, then $(L \cup M) \times K$ splits into the two disjoint sets $L \times K$ and $M \times K$, and so each function $f \subseteq (L \cup M) \times K$ splits into two functions

$$f_L = f \cap (L \times K) : L \to K, \quad f_M = f \cap (M \times K) : M \to K.$$

This defines a mapping $\pi : ((L \cup M) \to K) \to (L \to K) \times (M \to K)$ given by

$$\pi(f) = (f_L, f_M),$$

and it is very easy to check that π is, in fact, a bijection.

To get the cardinal identity (4-17) from this, use $\lambda \uplus \mu = L \cup M$ where $L = \{\underline{\text{blue}}\} \times \lambda$ and $M = \{\underline{\text{white}}\} \times \mu$, so that $L \cap M = \emptyset$ and $\lambda =_c L$, $\mu =_c M$, and compute:

$$
\begin{aligned}
\kappa^{(\lambda+\mu)} &=_c (L \cup M) \to K &&\text{(def., (C1) and \textbf{4.18})} \\
&=_c (L \to K) \times (M \to K) \\
&=_c \kappa^\lambda \cdot \kappa^\mu &&\text{(def., (C1) and \textbf{4.18}).}
\end{aligned}
$$

5.18. We prove by induction on n, that each function

$$f_n(s) = n + s$$

is an injection. This is trivial at the basis, since $f_0(s) = s$, and for the induction step,

$$f_{n+1}(s) = (n+1) + s = (n+s) + 1 = f_n(s) + 1,$$

and so f_{n+1} is the composition of f_n and the successor function; both of these are injections, and so f_{n+1} is an injection.

5.25. For the first claim, suppose that there is a $k \in [n, n)$, so that there exist t and s satisfying

$$n + t = k, \ k + s = n, \ k \neq n;$$

it follows that $n+t+s = n$, which by three applications of Exercise **5.18** implies that $t = 0$ and $s = 0$; and that gives $k = n$ contradicting the hypothesis.

The second claim is essentially a restatement of Lemma **5.22**. First we compute:

$$
\begin{aligned}
k < Sm &\iff k \leq Sm \ \& \ k \neq Sm &&\text{(def.)}, \\
&\iff (k \leq m \vee k = Sm) \ \& \ k \neq Sm &&\text{(by Lemma **5.22**)}, \\
&\iff k \leq m \\
&\iff (k \leq m \ \& \ k \neq m) \vee k = m \\
&\iff k < m \vee k = m,
\end{aligned}
$$

where for the third equivalence we have used the fact that

$$k \leq m \Longrightarrow k \neq Sm;$$

this holds because its negation implies that $Sm \leq m$ which is absurd (if $Sm + t = m$, then $m + St = m$, so $St = 0$ by Exercise **5.18**, which is false). It follows that for all k and $n \leq m$,

$$
k \in [n, Sm) \iff n \leq k \ \& \ k < Sm \iff n \leq k \ \& \ (k < m \vee k = m)
$$

$$
\iff (n \leq k \ \& \ k < m) \vee (n \leq k \ \& \ k = m) \iff k \in \Big([n, m) \cup \{m\} \Big),
$$

the last inference using the hypothesis $n \leq m$.

6.3. For each $x \in P$, we have $x \leq M$ since M is the maximum; and if M' is any upper bound of P, then $M \leq M'$ because $M \in P$.

6.4. (1) $\{e, f\}$ has no upper bound. (2) $\{a, c\}$ has upper bounds (b, d, e and f) but no least upper bound. (3) $\{b, d\}$ has a least upper bound (e) but no maximum.

6.5. Every member of P is an upper bound of $\emptyset$, because it satisfies (formally) the condition for being an upper bound,

$$(\forall x)[x \in \emptyset \Longrightarrow x \leq M];$$

it follows that M is *the least upper bound* of $\emptyset$ if and only if it is the least element of P.

6.6. The conditions

$$X \subseteq X, \; [X \subseteq Y \,\&\, Y \subseteq Z] \Longrightarrow X \subseteq Z, \; [X \subseteq Y \,\&\, Y \subseteq X] \Longrightarrow X = Y$$

are all trivial. If $\mathscr{E} \subseteq \mathcal{P}(A)$ is any set of subsets of A, then the unionset $\bigcup \mathscr{E}$ is an upper bound of $\mathscr{E}$, because $X \in \mathscr{E} \Longrightarrow X \subseteq \bigcup \mathscr{E}$; and it is the least upper bound, because for any M', if $X \subseteq M'$ for every $X \in \mathscr{E}$, then $\bigcup \mathscr{E} \subseteq M'$.

6.9. If E is empty, then $(A \rightharpoonup E) = (A \to E) = \{\emptyset\}$ and there is nothing to prove, so assume that $e_0 \in E$. If $f : A \rightharpoonup E$ and $X = \mathrm{Domain}(f)$, we set

$$f'(x) = \begin{cases} f(x), & \text{if } f(x)\downarrow, \\ e_0, & \text{otherwise,} \end{cases}$$

so that $f' : A \to E$ and $f' \restriction X = f$; and if $f : A \to E$ and $X \subseteq A$, then $f \restriction X : A \rightharpoonup E$ with $\mathrm{Domain}(f \restriction X) = X$, directly from the definitions.

6.11. The empty set satisfies (formally) the condition for being a chain,

$$(\forall x, y)[(x \in \emptyset \,\&\, y \in \emptyset) \Longrightarrow x \le y \lor y \le x].$$

6.12. The only chains in a flat poset P are singletons (including $\{\bot\}$) and doubletons of the form $\{\bot, a\}$, and every one of these sets has a maximum, which is also its supremum.

A discrete poset with at least two elements a, b, cannot be inductive, because it has no least element; and so the only inductive, discrete posets are the singletons.

6.13. If $u, v \in \{x_n \mid n \in \mathbb{N}\}$, then there exist n, m such that $u = x_n$ and $v = x_m$; if $n \le m$, then $u = x_n \le x_m = v$, and, similarly, if $m \le n$, then $v \le u$.

6.15. With the usual ordering, $\mathbb{N}$ is a chain and it has no upper bound; and the poset of Figure **6.2** does not have a least element.

6.16. We claim that if $X \subseteq P = E^* \cup (\mathbb{N} \to E)$ and X is a chain, then the union set $\bigcup X$ is a function with domain some subset of $\mathbb{N}$: because if $(x, y), (x, y') \in \bigcup X$, then there exist $p, q \in X$ such that $(x, y) \in p$ and $(x, y') \in q$; and if $p \subseteq q$, we have $(x, y), (x, y') \in q$ so that $y = y'$, while if $q \subseteq p$ we have $(x, y), (x, y') \in p$, so that, again, $y = y'$. Moreover, the domain of $\bigcup X$ is either all of $\mathbb{N}$ or a finite, initial segment of $\mathbb{N}$, because it is closed under $<$:

$$\begin{aligned} x < x' \in \mathrm{Domain}(\textstyle\bigcup X) &\Longrightarrow \text{for some } p \in X, x' \in \mathrm{Domain}(p) \\ &\Longrightarrow x \in \mathrm{Domain}(p) \\ &\Longrightarrow x \in \mathrm{Domain}(\textstyle\bigcup X). \end{aligned}$$

Thus $\bigcup X \in P$, and it is clear that it is the least upper bound of X.

6.17. Suppose $S \subseteq (A \rightarrowtail B)$ is a chain of *partial injections* and let $p^* = \bigcup S$ be the unionset of S. Since p^* is obviously the least upper bound of S under $\subseteq$, it is enough to show that p is a partial injection; i.e., we must show that

$$(x, y), (x, y') \in p^* \Longrightarrow y = y, \tag{a}$$

$$(x, y), (x', y) \in p^* \Longrightarrow x = x'. \tag{b}$$

For (a): from the hypotheses, there are $p, p' \in S$ such that $(x, y) \in p$ and $(x, y') \in p'$; if $p \leq p'$, then $(x, y) \in p'$, and since p' is a partial function and it also contains (x, y'), we must have $y = y'$; and if $p' \leq p$, then the analogous symmetric reasoning applies and leads to the same conclusion.

The argument for (b) is similar.

6.20. If $x_0 \leq_P x_1 \leq_P \cdots$, then the set $X = \{x_n \mid n \in \mathbb{N}\}$ of values of the sequence $\{x_n\}_{n \in \mathbb{N}}$ and its image $\pi[X] = \{\pi(x_n) \mid n \in \mathbb{N}\}$ under a monotone π are both non-empty, countable chains, an observation that was needed in order to define limits of monotone sequences in the first place,

$$\lim_n x_n = \sup X, \quad \lim_n \pi(x_n) = \sup \pi[X];$$

it follows that if π is countably continuous, then

$$\pi(\lim_n x_n) = \pi(\sup(X)) = \sup \pi[X] = \lim_n \pi(x_n),$$

as required for one direction of the Exercise. For the other direction, suppose that $S = \{s_0, s_1, \dots\}$ is a non-empty, countable chain in P, and define the function (sequence) $x : \mathbb{N} \to S$ by the recursion

$$x_0 = s_0, \quad x_{n+1} = \max_P(x_n, s_{n+1}).$$

The definition makes sense, since each x_n is a member of the chain S and hence comparable with s_{n+1}. By a trivial induction on n,

$$x_n \in S, \; s_n \leq_P x_n, \; x_n \leq_P x_{n+1},$$

from which we get immediately that

$$\sup S = \sup\{x_n \mid n \in \mathbb{N}\} = \lim_n x_n;$$

and the same holds for the images by any monotone π,

$$\sup \pi[S] = \sup\{\pi(x_n) \mid n \in \mathbb{N}\} = \lim_n \pi(x_n).$$

Thus, if π respects limits of monotone sequences, we have the required

$$\sup \pi[S] = \lim_n \pi(x_n) = \pi(\lim_n x_n) = \pi[\sup S].$$

6.24. Directly from the definition,

$$\pi(f)(n) = \pi(f_0)(n),$$

where f_0 is the restriction of f to the finite set $\{0, \dots, n\}$. For the second part, if $f(n) = 2n$, then

$$\pi(f)(2) = f(0) + f(1) + f(2) = 0 + 2 + 4 = 6.$$

6.26. If $f : X \to Y$ is continuous and $F \subseteq Y$ is closed, then $G = Y \setminus F$ is open, and $F = Y \setminus G$, so

$$f^{-1}[F] = f^{-1}[Y \setminus G] = f^{-1}[Y] \setminus f^{-1}[G] = X \setminus f^{-1}[G],$$

so $f^{-1}[F]$ is closed. The symmetric argument shows that if $f^{-1}[F]$ is closed for every closed $F \subseteq Y$, then $f^{-1}[G]$ is open for every open $G \subseteq Y$, and so f is continuous.

6.32. Directly from the definition,

$$\pi(f)(n) = \pi(f_0)(n),$$

where $f_0 = \{(n, 0)\}$ if $n \in A$, and if $n \notin A$, then

$$f_0 = \{(n, h(w)) \mid (n + 1, w) \in f\}.$$

This f_0 is finite, in fact it is defined on exactly one number if $n \in A$ or $n \notin A \,\&\, f(n+1) \downarrow$, and it is the empty partial function otherwise; and so by **6.23**, π is a continuous map.

7.3. Suppose $\leq_C$ is a wellordering of C and $f : A \rightarrowtail C$ is an injection, and let

$$x \leq_A y \iff f(x) \leq_C f(y) \qquad (x, y \in A).$$

We prove that the relation $\leq_A$ is a wellordering of A.

Proof that $\leq_A$ is a linear ordering is trivial, basically by inspection. For example,

$$
\begin{aligned}
x \leq_A y \,\&\, y \leq_A z &\implies f(x) \leq_C f(y) \,\&\, f(y) \leq_C f(z) \\
&\implies f(x) \leq_C f(z) \quad \text{(because } \leq_C \text{ is transitive)} \\
&\implies x \leq_A z \quad \text{(by def.)},
\end{aligned}
$$

so that $\leq_A$ is transitive.

To check that $\leq_A$ wellorders A, suppose $X \subseteq A$ and $X \neq \emptyset$, and let $f[X]$ be the image of X; now $f[X]$ is a non-empty subset of C, and so it has a least element $f(m)$; and then m is the $\leq_A$-least element of X, because for every $x \in X$, $f(m) \leq_C f(x)$, and so $m \leq_A x$, by the definition.

7.4. If $(C, \leq_C)$ is a well ordered set and $f : C \twoheadrightarrow A$, define "an inverse" function $g : A \to C$ by

$$g(x) = \min_C \{y \mid f(y) = x\}.$$

This is an injection; because if $x \neq x'$, then

$$\{y \mid f(y) = x\} \cap \{y \mid f(y) = x'\} = \emptyset,$$

and $g(x)$ is in the first of these two sets while $g(x')$ is in the second.

We now appeal to Exercise **7.3** using this injection.

7.6. This is trivial: the restriction $\leq_I$ of $\leq_U$ to I inherits the properties of reflexiveness, transitivity and antisymmetry, and every non-empty $X \subseteq I$ has a least member in $I \cap U$, which is its least member in I.

7.8. By the definition, $\mathbf{seg}(0) = \{x \in U \mid x <_U 0\}$, and so $\mathbf{seg}(0) = \emptyset$ since 0 is the least element.

Again by the definition,

$$S(x) = \min\{y \in U \mid x <_U y\},$$

so that there is no $z \in U$ such that $x <_U z <_U S(x)$; this means that

$$y <_U S(x) \iff y <_U x \vee y = x,$$

i.e., $\mathbf{seg}(S(x)) = \mathbf{seg}(x) \cup \{x\}$.

7.10. If V is the set of all initial segments of U, then

$$V = V_p \cup \{U\},$$

where the "proper" initial segments in V_p are those of the form $\mathbf{seg}_U(x)$ with $x \in U$. Moreover, the mapping

$$f(I) = \text{the unique } x \in U \text{ such that } I = \mathbf{seg}_U(x)$$

is obviously order preserving from V_p into U (in fact it is a similarity) and so V_p is well ordered. Now V is obtained by adding a point (U) at the top of V_p, and so it remains well ordered—for a detailed proof of this, see Exercise **7.18** below.

7.12. Recall from Definition **6.18** that a mapping $\pi : P \to Q$ is monotone if

$$x \le y \Longrightarrow \pi(x) \le \pi(y),$$

and so, it is order-preserving if, *in addition*, the converse holds,

$$\pi(x) \le \pi(y) \Longrightarrow x \le y;$$

hence, order-preserving mappings are monotone. For a counterexample to the converse, take any constant mapping $\pi(t) = c$, with P chosen so that it has (at least) two members $x \ne y$; constant mappings are trivially monotone, but this one cannot be order preserving, since $\pi(x) = \pi(y) = c$, which would imply (if π were order preserving) that $x \le y$ and $y \le x$, i.e., $x = y$.

7.13. Suppose (towards a contradiction) that P and Q are linear, $f : P \to Q$ is order-preserving, $x <_P y$ but $f(x) \not<_Q f(y)$; it follows by the linearity of Q that $f(y) \le_Q f(x)$; but this implies $y \le_P x$, since f is order-preserving; and this contradicts the hypothesis $x <_P y$.

For the converse, suppose (with P and Q linear again) that f is strictly monotone but not order-preserving. Easily, $x \le_P y \Longrightarrow f(x) \le_Q f(y)$, taking cases on whether $x = y$ or $x <_P y$, and so, since f is not order-preserving, there must exist $x, y \in P$ such that $f(x) \le_Q f(y)$ but $x \not\le_P y$; but then $y <_P x$ by the linearity of P; and so $f(y) <_Q f(x)$, since f is strictly monotone, contradicting the other assumption, $f(x) \le_Q f(y)$.

7.14. The identity mapping $\pi(x) = x$ on P is a similarity, so $P =_o P$.

To verify the symmetry property, suppose $\pi : P \rightarrowtail\!\!\!\rightarrow Q$ is a similarity, and let $\pi^{-1} : Q \rightarrowtail\!\!\!\rightarrow P$ be the inverse bijection. Now, if $x, y \in Q$, then

$$\pi^{-1}(x) \leq_P \pi^{-1}(y) \iff \pi(\pi^{-1}(x)) \leq_Q \pi(\pi^{-1}(y))$$

since π is order-preserving, and the order-preserving property for π^{-1} follows because $\pi(\pi^{-1}(z)) = z$ for every $z \in Q$.

For the transitivity property, we first observe that *the composition of order-preserving mappings is order-preserving*:

$$x \leq_P y \iff \pi(x) \leq_Q \pi(y) \iff \rho(\pi(x)) \leq_R \rho(\pi(y)).$$

Since the composition of bijections is a bijection, it follows that *the composition of similarities is a similarity*, and so $=_o$ is transitive.

7.17. If $\pi : P \rightarrowtail\!\!\!\rightarrow Q$ is a similarity of P with Q, then its extension by one value $p \cup \{(t_P, t_Q)\}$ is easily a similarity of $\mathrm{Succ}(P)$ with $\mathrm{Succ}(Q)$.

7.18. It is easy to show that $\leq_{\mathrm{Succ}(U)}$ is a linear order of $\mathrm{Field}(U) \cup \{t_U\}$. To check the wellordering property, suppose $\emptyset \neq X \subseteq \mathrm{Field}(U) \cup \{t_U\}$; if $X = \{t_U\}$, then t_U is the least member of X in $\mathrm{Succ}(U)$, and if $X \cap \mathrm{Field}(U) \neq \emptyset$, then the $\leq_U$-least member of this set is the $\leq_{\mathrm{Succ}(U)}$-least member of X.

7.28. Let $\sigma = \rho\pi$ be the composition of π and ρ, so that

$$\sigma(x) = \rho(\pi(x)) \quad (x \in U).$$

To prove that $\sigma : U \to W$ is order preserving, we compute, using the fact that π and ρ are order preserving:

$$\begin{aligned}
x \leq_U y &\iff \pi(x) \leq_V \pi(y) \\
&\iff \rho(\pi(x)) \leq_W \rho(\pi(x)) \\
&\iff \sigma(x) \leq_W \sigma(y).
\end{aligned}$$

To show that $\sigma[U]$ is an initial segment of W, suppose $w \in \sigma[U] = \rho[\pi[U]]$, so that $w = \rho(v)$ for some $v \in \pi[U]$, and $w' <_W w$. Since ρ is an initial similarity, there must exist some $v' \in V$ such that $w' = \rho(v')$; and since $\rho(v') <_W \rho(v)$ and ρ is order preserving, we must have $v' <_V v$. Since $v \in \pi[U]$, we have at this point some $u \in U$ such that

$$v = \pi(u) \text{ and } v' <_V v.$$

But π is also an initial similarity, so that (as above) there must exist some $u' <_U u$ such that $v' = \pi(u')$, from which we get

$$\sigma(u') = \rho(\pi(u')) = \rho(v') = w',$$

which completes the proof.

8.3. There is no choice set S for a family $\mathscr{E}$ which contains $\emptyset$, since $S \cap \emptyset$ is not singleton. For the second example, if S is a choice set for $\mathscr{E}$, then $S \cap \{a\} = \{a\}$ and also $S \cap \{b\} = \{b\}$, so that $\{a, b\} \subseteq S$ and $S \cap \{a, b\}$ is not a singleton.

8.7. Suppose $(\forall x)(\exists y)P(x, y)$, and let ε be a choice function for B, so that

$$[X \subseteq B \& X \neq \emptyset] \Longrightarrow \varepsilon(X) \in X.$$

Define $f : A \to B$ by

$$f(x) = \varepsilon(\{y \mid P(x, y)\});$$

this makes sense since for each $x \in A$, $\{y \mid P(x, y)\} \neq \emptyset$ by the hypothesis. Moreover, $f(x) \in \{y \mid P(x, y)\}$ for any $x \in A$, since ε is a choice function for B, and so $P(x, f(x))$, as required.

8.15. Suppose $a \in A$, $P \subseteq A \times A$, and $(\forall x \in A)(\exists y \in A)P(x, y)$. Let

$$P^*(u, x) \iff u \in A^* \& x \in A$$

$$\& \left[[u = \emptyset \& x = a] \vee [\mathrm{lh}(u) > 0 \& P(u(\mathrm{lh}(u) - 1), x)]\right].$$

If $u = \emptyset$, then $P^*(u, a)$, and if $u = \langle u_0, \ldots, u_n \rangle$ with $n \geq 0$ and $P(u_n, x)$, then $P^*(u, x)$; thus the hypothesis on P guarantees that $(\forall u)(\exists x)P^*(u, x)$, and the assumed version of **DC** supplies a function $f : \mathbb{N} \to A$ such that $(\forall n)P^*(\overline{f}(n), f(n))$; which in particular implies that $f(0) = a$, when we apply it to $u = \emptyset$. Moreover, if $n > 0$,

$$P^*(\overline{f}(n), f(n)) \iff P(f(n - 1), f(n)),$$

so that $(\forall n > 0)P(f(n-1), f(n))$, which is the same as $(\forall n)P(f(n), f(n+1))$.

8.18. Suppose first that $(P, \leq)$ is linear and grounded, and let $X \subseteq P$ be a non-empty set of points; it follows that there is some $m \in X$ such that for all $x \in X$, $x \not< m$, which by the linearity of $\leq$ means that $m \leq x$, and so m is the least member of X. Conversely, if $(P, \leq)$ is a wellordering and X is any non-empty subset of P, then X has a least element m, which is certainly minimal in X.

9.4. If u is a leaf, then $u \sqsubseteq w \Longrightarrow w = u$, so that $T_u = \{w \in T \mid w \sqsubseteq u\}$; on the other hand, the second summand in (9-2) is empty (because u has no children), and so (9-2) holds. If u is not a leaf, then clearly, for every w,

$$u \sqsubseteq w \iff (\exists v)[v \text{ is a child of } u \& v \sqsubseteq w],$$

which again implies (9-2).

9.5. An infinite branch $f : \mathbb{N} \to \mathbb{N}$ would need to satisfy

$$f(0) > f(1) > \cdots,$$

and there are no infinite, descending sequences of natural numbers.

9.12. In full detail, we know that $\kappa <_c 2^\kappa$, and so $\kappa^+ \leq_c 2^\kappa$ by (9-5). Now the **GCH** asserts that no cardinal is properly between an infinite κ and 2^κ, so that $\kappa^+ =_c 2^\kappa$; and, conversely, if $\kappa^+ =_c 2^\kappa$, then there is no cardinal properly between κ and 2^κ, again by (9-5).

9.19. If κ_1 and κ_2 satisfy $(1) - (3)$, then $\kappa_1 =_c A$ for some $A \in \mathscr{E}$ by (2) for κ_1; and so $\kappa_2 \leq_c A \leq_c \kappa_1$ by (3) for κ_2. The symmetric argument gives $\kappa_1 \leq_c \kappa_2$.

9.22. This is immediate, from the theorem and the fact that for every i, $\{i\} <_c 2$, since 2 is a set with two members.

9.24. This is immediate.

10.2. Suppose $u \sqsubseteq v$ and compute:

$$x \in \mathcal{N}_v \Longrightarrow v \sqsubseteq x \Longrightarrow u \sqsubseteq x \Longrightarrow x \in \mathcal{N}_u.$$

The converse is equally obvious.

10.3. The family of neighborhoods is equinumerous with $\mathbb{N}^*$, and $\mathbb{N}^* =_c \mathbb{N}$.

10.5. If G is open and $x \in G$, then by $(10\text{-}6)$, there is a u such that $x \in \mathcal{N}_u \subseteq G$; and if each $x \in G$ belongs to some $\mathcal{N}_u \subseteq G$, then $(10\text{-}6)$ holds trivially.

10.9. By definition, $\mathcal{N}_u = [T_u]$, where the subtree T_u is defined in $(9\text{-}1)$ and it is splitting, since for each $v \in T_u$ and each n, $v \star \langle n \rangle \in T_u$.

10.13. If A is uncountable and $A = \bigcup_n A_{n \in \mathbb{N}}$, then at least one A_n must be uncountable; and if $A_n \in \Gamma$, then A_n has a non-empty, perfect subset P by the hypothesis, and so also $P \subseteq A$.

10.17. If $\pi : (\mathbb{N} \rightarrow \mathbb{N}) \rightarrow (\mathbb{N} \rightarrow \mathbb{N})$ is continuous and $x \in \mathcal{N}$, then

$$\pi(x) = \sup\{\pi(v) \mid v \in \mathbb{N}^*, v \sqsubseteq x\},$$

and so if $\pi(x) \in \mathcal{N}_u$ (which means $u \sqsubseteq \pi(x)$), there must exist some $v \sqsubseteq x$ such that $u \sqsubseteq \pi(v)$; but this implies that $u \sqsubseteq \pi(y)$ for any $y \in \mathcal{N}$ such that $v \sqsubseteq y$, so that finally,

$$y \in \mathcal{N}_v \Longrightarrow \pi(y) \in \mathcal{N}_u$$

for this v, which establishes the continuity of $f = \pi {\restriction} \mathcal{N}$ at x.

10.26. Call (temporarily) a collection of pointsets *good* if it is one of the collections whose intersection defines $\mathcal{B}(X)$, i.e.,

$$\mathscr{E} \text{ is good} \iff \mathcal{G} \subseteq \mathscr{E}$$
$$\& \, (\forall \{A_n\}_n)[(\forall n)A_n \in \mathscr{E} \Longrightarrow \bigcup_n A_n \in \mathscr{E}]$$
$$\& \, (\forall A \in \mathscr{E})[cA \in \mathscr{E}]\}.$$

To verify that $\mathcal{B}(X) = \bigcap\{\mathscr{E} \mid \mathscr{E} \text{ is good}\}$, is a σ-field which contains all the open sets (i.e., good), we must check the following three claims:

(1) *Every open set $G \subseteq X$ is in $\mathcal{B}(X)$*; this is because $G \in \mathscr{E}$ for every good $\mathscr{E}$, and so $G \in \mathcal{B}(X)$.

(2) *If $A \in \mathcal{B}(X)$, then $cA \in \mathcal{B}(X)$*: if $A \in \mathscr{E}$ for every good $\mathscr{E}$, then $cA \in \mathscr{E}$ for every good $\mathscr{E}$, by the definition of "goodness", and so $cA \in \mathcal{B}(X)$.

(3) *If $A_n \in \mathcal{B}(X)$ for every n, then $\bigcup_n A_n \in \mathcal{B}(X)$.* Fix some good collection $\mathcal{E}$. The hypothesis tells us that each $A_n \in \mathcal{E}$, and so by the definition of "goodness", $\bigcup_n A_n \in \mathcal{E}$; but $\mathcal{E}$ was arbitrary, and so $\bigcup_n A_n$ belongs to every good collection, which means that $\bigcup_n A_n \in \mathcal{B}(X)$.

10.27. By Problem **x1.3**, $\bigcap_n A_n = c(\bigcup_n cA_n)$.

11.4. By the Replacement Axiom, the image $F[X]$ of any set X is a set, so there exists a w satisfying the condition

$$(\forall y)[y \in w \iff (\exists x \in X)[y = F(x)]];$$

and by the Extensionality Axiom, only one such w can exist, so we can set

$$G(X) = \text{the unique } w \text{ such that } (\forall y)[y \in w \iff (\exists x \in X)[y = F(x)]].$$

11.7.

$$H(w,n) = \begin{cases} a, & \text{if } n = 0, \\ y, & \text{if } n > 0 \,\&\, (n-1, y) \in w \,\&\, (\forall t \neq y)[(n-1,t) \notin w], \\ \emptyset, & \text{otherwise.} \end{cases}$$

11.9. In these cases, it is easiest to verify $\bigcup M \subseteq M$ by inspection, after computing the relevant unionsets:

$$\bigcup \emptyset = \emptyset;$$
$$\bigcup \{\emptyset, \{\emptyset\}\} = \{\emptyset\};$$
$$\bigcup \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{\emptyset, \{\emptyset\}\};$$
$$\bigcup \mathbb{N}_0 = \mathbb{N}_0.$$

Finally, if M is a class of atoms, then $\bigcup M = \emptyset$, so $\bigcup M \subseteq M$.

11.11. By the definition, $A \in \mathrm{TC}(A)$ and $\mathrm{TC}(A)$ is transitive, so that $A \subseteq \mathrm{TC}(A)$ and hence $A \cup \{A\} \subseteq \mathrm{TC}(A)$, for any A. For the other direction, suppose A is transitive and $x \in A \cup \{A\}$. If $x \in A$, then $x \subseteq A \subseteq A \cup \{A\}$, and if $x = A$, then again $x \subseteq A \subseteq A \cup \{A\}$; which shows that $A \cup \{A\}$ is transitive, and hence $A \cup \{A\} \subseteq \mathrm{TC}(A)$.

11.13. If there are no atoms, then the transitive closure $\mathrm{TC}(A)$ of every set has no atoms, and so every set is pure. Conversely, if there exists some atom a, then $\mathrm{TC}(\{a\}) = \{\{a\}, a\}$, and so $\{a\}$ is not pure.

11.14. If A is transitive, then $\mathrm{TC}(A) = A \cup \{A\}$ by Exercise **11.11**, and $A \cup \{A\}$ is finite or countable exactly when A is finite or countable.

11.16. False: the singleton $\{a\}$ of an atom is transitive with $\mathrm{TC}(\{a\}) = \{\{a\}, a\}$, but it is not a subset of its powerset $\mathcal{P}(\{a\}) = \{\emptyset, \{a\}\}$, all of whose members are (by definition) sets.

11.17. The inclusion $M_n(I) \subseteq M_n(J)$ is proved by a simple induction on n, with the basis $M_0(I) = I \subseteq J = M_0(J)$ supplied by the hypothesis.

11.20. The class $\mathcal{W}$ of all objects is transitive, because if y is an object and $x \in y$, then x is also an object—simply because we have assumed that membership is a condition on pairs of objects; and it is a Zermelo universe, because we have assumed of it all the conditions we demand of a Zermelo universe—that it contains the unordered pair $\{x, y\}$ of any two objects (in it), and the unionset $\bigcup X$ and powerset $\mathcal{P}(X)$ of any set (in it), and that it also contains a set I which satisfies the Axiom of Infinity, from which $\mathbb{N}_0$ can be constructed using the closure properties of $\mathcal{W}$, as in the proof of Theorem **5.4**.

For the second claim, notice first that $\mathbb{N}_0 \subseteq M$, since M is transitive and contains $\mathbb{N}_0$, and so $\emptyset \in M$, since $\emptyset \in \mathbb{N}_0$. Moreover, if $A \in M$, then $\mathcal{P}(A) \in M$ and so $\mathcal{P}(A) \subseteq M$, which means that every subset of A is in M.

11.22. The Peano system constructed in the proof of Theorem **5.4** is a member of every Zermelo universe M, because by Proposition **11.21**, M is closed under all the operations we used in that proof to construct it.

11.24. By the Axiom of Choice, the hypothesis of (11-24) implies that there exists some $f : A \to B$ such that for all $x \in A$, $P(x, f(x))$. But the function space $(A \to B) \in M$, and so $f \in M$ since M is transitive.

11.27. No descending $\in$-chain can start with x if x has no members, which is why atoms and $\emptyset$ are grounded. For $\mathbb{N}_0$, we use the fact that it is a Peano system with $0 = \emptyset$ and $Sm = \{m\}$, and we prove by induction that for all $m \in \mathbb{N}_0$, there is no infinite, descending $\in$-chain which starts with m: this is clear if $m = 0$, since $0 = \emptyset$, and if it is true of m, then it is also true of $Sm = \{m\}$, for which the alleged chain would have to start with

$$\{m\} \ni m \ni \cdots$$

immediately yielding a chain which starts with m.

If A is grounded, then so is each $x \in A$: because if $x \ni x_1 \ni \cdots$ were an infinite, descending $\in$-chain starting with some $x \in A$, then $A \ni x \ni x_1 \ni \cdots$ would be an infinite, descending $\in$-chain starting with A; and conversely, if every member of A is grounded, then we cannot have a chain $A \ni x_1 \ni \cdots$, because the tail $x_1 \ni \cdots$ would witness that x_1 is not grounded.

Similarly, if A is grounded, then so is $\mathcal{P}(A)$: because any chain $\mathcal{P}(A) \ni X \ni x_1 \ni \cdots$ would yield a chain $x_1 \ni \cdots$ starting with $x_1 \in A$. And conversely, if $\mathcal{P}(A)$ is grounded, then so is A, because any chain $\mathcal{P}(A) \ni X \ni x \ni x_1 \ni \cdots$ starting with $\mathcal{P}(A)$ would yield a chain $x \ni x_1 \ni \cdots$ starting with some $x \in A$.

Finally, the class of all grounded sets is transitive, because, again, all elements of a grounded set grounded.

12.2. For any function $f : A \to B$, the image $f[\emptyset]$ of the empty set is empty; and so with $\boldsymbol{v}_U : U \twoheadrightarrow \mathrm{ord}(U)$, since 0_U has no predecessors,

$$\boldsymbol{v}_U(0_U) = \boldsymbol{v}_U[\{y \in U \mid y < 0_U\}] = \boldsymbol{v}_U[\emptyset] = \emptyset.$$

By the definition of the successor operation on U,

$$y <_U S(x) \iff y <_U x \vee y = x,$$

and using this, we can compute:

$$\begin{aligned}
\boldsymbol{\nu}_U(S(x)) &= \{\boldsymbol{\nu}_U(y) \mid y <_U S(x)\} \\
&= \{\boldsymbol{\nu}_U(y) \mid y <_U x\} \cup \{\boldsymbol{\nu}_U(x)\} \\
&= \boldsymbol{\nu}_U(x) \cup \{\boldsymbol{\nu}_U(x)\}.
\end{aligned}$$

12.3. If 0_U is least in U, then $0_U <_U x$, and so $\emptyset = \boldsymbol{\nu}_U(0_U) \in \boldsymbol{\nu}_U(x)$. If $\alpha \in \boldsymbol{\nu}_U(x)$, then $\alpha = \boldsymbol{\nu}_U(y)$, for some $y <_U x$; but then y has a successor $S(y) <_U x$, since x is a limit point, $\boldsymbol{\nu}_U(S(y)) \in \boldsymbol{\nu}_U(x)$, by the definition, and using the preceding Exercise **12.2**,

$$\boldsymbol{\nu}_U(S(y)) = \boldsymbol{\nu}_U(y) \cup \{\boldsymbol{\nu}_U(y)\} = \alpha \cup \{\alpha\} \in \boldsymbol{\nu}_U(x).$$

12.4. Let

$$G = \{X \mid \emptyset \in X \,\&\, (\forall \alpha \in X)[\alpha \cup \{\alpha\} \in X]\}.$$

Since ω_U is a limit point, $\boldsymbol{\nu}_U(\omega_U) \in G$ by Exercise **12.3**—and, in particular, G is a non-empty class. For the converse, suppose towards a contradiction that there is some $X \in G$ such that $\boldsymbol{\nu}_U(\omega_U) \not\subseteq X$, and let y be least in U such that $\boldsymbol{\nu}_U(y) \notin X$. Now y is not the least element of U, since $\boldsymbol{\nu}_U(0_U) = \emptyset \in X$, by the hypothesis on X. And y is not a limit point of U, since $y < \omega_U$; hence $y = S(x)$ for some $x \in U$, $\boldsymbol{\nu}_U(x) \in X$, by the choice of y, and $\boldsymbol{\nu}_U(y) = \boldsymbol{\nu}_U(x) \cup \{\boldsymbol{\nu}_U(x)\} \in X$ by the hypothesis on X, which contradicts the assumption on y.

12.6. If $\pi : U \rightarrowtail V$ is an initial similarity, then, directly from Lemma **12.5**,

$$\mathrm{ord}(U) = \{\boldsymbol{\nu}_U(x) \mid x \in U\} = \{\boldsymbol{\nu}_V(\pi(x)) \mid x \in U\} \subseteq \mathrm{ord}(V).$$

12.8. If $U = \mathrm{Succ}((\mathbb{N}, \leq_{\mathbb{N}}))$ is the next well ordered set to $(\mathbb{N}, \leq_{\mathbb{N}})$ with t added on top, then $t = \omega_U$, $\mathbb{N} = \mathrm{seg}_U(t)$ and $\mathrm{ord}(\mathbb{N}, \leq_{\mathbb{N}}) = \boldsymbol{\nu}_U(t) = \omega$. The rest follows immediately from Exercise **12.2**.

12.10. More precisely, the claim is that $\mathrm{ord}(\alpha, \leq_\alpha) = \alpha$, and the proof is as follows: if $\alpha = \mathrm{ord}(U)$, then $(\alpha, \leq_\alpha) =_o U$ by Lemma **12.9**, and so by Exercise **12.6**

$$\mathrm{ord}(\alpha) = \mathrm{ord}(U) = \alpha.$$

12.13. If $U =_o V$, then

$$\mathrm{ord}(U) =_o U =_o V =_o \mathrm{ord}(V)$$

by (12-12), and so $\mathrm{ord}(\mathrm{ord}(U)) =_o \mathrm{ord}(\mathrm{ord}(V))$ by (12-12) again, from which $\mathrm{ord}(U) = \mathrm{ord}(V)$ follows by Exercise **12.10**.

For the second claim: if $U =_o \alpha$ and $U =_o \beta$, then $\alpha =_o \beta$, and so, in the same way,

$$\alpha = \mathrm{ord}(\alpha) = \mathrm{ord}(\beta) = \beta.$$

12.16. If β is the least member of $\mathscr{E}$, then $\beta \leq \alpha$, for every $\alpha \in \mathscr{E}$, so that by Lemma **12.9**, $\beta \subseteq \alpha$; thus $\beta \subseteq \bigcap \mathscr{E}$. On the other hand, if $\gamma \in \beta$, then $\gamma < \beta$, so $\gamma < \alpha$ for every $\alpha \in \mathscr{E}$, i.e., $\gamma \in \alpha$ for every $\alpha \in \mathscr{E}$, hence $\gamma \in \bigcap \mathscr{E}$; thus $\gamma \in \beta \Longrightarrow \gamma \in \mathscr{E}$, i.e., $\beta \subseteq \bigcap \mathscr{E}$.

12.20. The successor poset $\mathrm{Succ}(\alpha)$ is defined by adding a new element $r \notin \alpha$ to the field α and placing it above all the members of α. Define $\pi : \alpha + 1 = \alpha \cup \{\alpha\} \to \mathrm{Succ}(\alpha)$ by

$$\pi(x) = \begin{cases} x, & \text{if } x \in \alpha, \\ r, & \text{otherwise (i.e., if } x = \alpha). \end{cases}$$

It follows immediately that π is a similarity, so that $\alpha + 1 =_0 \mathrm{Succ}(\alpha)$, and hence $\alpha + 1 = \mathrm{ord}(\alpha + 1) = \mathrm{ord}(\mathrm{Succ}(\alpha))$.

12.21. This is an elaboration of the preceding Exercise **12.20**. By the definition of the sum of two posets in **7.37**, we have

$$\alpha +_o \beta = (\{0\} \times \alpha \cup \{1\} \times \beta, \leq),$$

where the ordering $\leq$ is defined by

$$(i, x) \leq (j, y) \iff i < j \vee [i = j \,\&\, [x = y \vee x \in y]],$$

using the properties of the ordering on the ordinals. Fix α, and for each β, define the function

$$\pi_\beta : \alpha +_o \beta \to \mathrm{ON}$$

by

$$\pi(0, x) = x, \quad \pi(1, y) = \alpha + y;$$

we show by ordinal induction on β that π_β *is a similarity of* $\alpha +_o \beta$ *with* $\alpha + \beta$, so that

$$\mathrm{ord}(\alpha +_o \beta) = \mathrm{ord}(\alpha + \beta) = \alpha + \beta.$$

At the basis $\beta = 0$, $\alpha +_o 0 = (\{0\} \times \alpha, \leq)$, so that $\pi_0(0, x) = x$ and this is obviously a similarity of $\alpha +_o 0$ with $\alpha + 0 = \alpha$.

At the induction step, if $\beta = \gamma + 1$, then the required result is exactly the preceding Exercise **12.20**.

Finally, at the induction step when β is a limit ordinal, the induction hypothesis gives us that for each $\gamma < \beta$, the function π_γ is a similarity of $\alpha +_o \gamma$ with $\alpha + \gamma$. It is obvious from the definition of these functions that

$$\gamma < \delta < \beta \Longrightarrow \pi_\gamma \subseteq \pi_\delta;$$

and from this we get easily that the union

$$\pi_\beta = \bigcup\nolimits_{\gamma < \beta} \pi_\gamma$$

is a similarity of $\alpha +_o \beta$ with $\alpha + \beta$, as required.

The results about associativity and (not)-commutativity follow now from the problems in Chapter **7**, especially Problem **x7.4**.

Finally, (12-26) also follows from the main part of the exercise, because if $\pi : \beta \rightarrowtail \gamma$ is an initial similarity of β onto a proper initial segment of γ, then the function

$$\rho((i, x)) = \begin{cases} (i, x), & \text{if } i = 0, \\ (i, \pi(x)), & \text{if } i = 1 \end{cases}$$

is easily an initial similarity of $\alpha +_o \beta$ onto a proper, initial segment of $\alpha +_o \gamma$.

12.22. By the definition,

$$\alpha \cdot_o \beta = (\alpha \times \beta, \leq),$$

where $\leq$ is the "inverse lexicographic" ordering on the pairs,

$$(x_1, y_1) \leq (x_2, y_2) \iff y_1 < y_2 \vee [y_1 = y_2 \,\&\, x_1 \leq x_2].$$

To illustrate a different method than the one we used in the preceding exercise, we prove for each fixed α by ordinal induction on β, that

$$\alpha \cdot_0 \beta =_0 \alpha \cdot \beta,$$

which gives the required result by taking the ordinals of both sides.

At the basis $\beta = 0$, we have that $\alpha \cdot_o 0$ is the empty poset, $\alpha \cdot 0 = 0 = \emptyset$, so we have literal equality.

At the successor induction step, $\beta = \gamma + 1 = \gamma \cup \{\gamma\}$, so

$$\alpha \times \beta = (\alpha \times \gamma) \cup (\alpha \times \{\gamma\}).$$

Notice that

$$(\alpha \times \gamma) \cap (\alpha \times \{\gamma\}) = \emptyset,$$

because the pairs (x, γ) in the second part have second member γ, while every pair (x, y) in the first part has second member some $y < \gamma$. Moreover,

$$(\{(x, \gamma) \mid x \in \alpha\}, \leq) =_o \alpha$$

by the trivial similarity $\rho(x, \gamma) = x$, and so, by the definition of poset addition,

$$\alpha \cdot_o \beta =_o \alpha \cdot_o \gamma +_o \alpha;$$

but then the induction hypothesis, the preceding Exercise **12.21** and the definition of ordinal multiplication imply together that

$$\alpha \cdot_o \beta =_o \alpha \cdot \gamma + \alpha = \alpha \cdot \beta$$

which is what we needed to show.

For the limit case in the induction step, the induction hypothesis gives us for each $\gamma < \beta$ a similarity

$$\pi_\gamma : \alpha \cdot_o \gamma \rightarrowtail \alpha \cdot \gamma,$$

which is an initial similarity of $\alpha \cdot_o \gamma$ into $\alpha \cdot \beta$ because $\alpha \cdot \gamma \leq \alpha \cdot \beta$, so that $\alpha \cdot \gamma$ is an initial segment of $\alpha \cdot \beta$. We can now appeal to that fact that if U, V

are well ordered sets and $U \leq_o V$, then there is exactly one initial similarity $\pi : U \rightarrowtail V$ by Problem **x7.13**. In the present case, this means that

$$\delta < \gamma < \beta \Longrightarrow \pi_\beta \restriction \gamma = \pi_\gamma,$$

i.e., these initial similarities "cohere", so that their union is an initial similarity of $\alpha \cdot_o \beta$ into $\alpha \cdot \beta$,

$$\cup\{\pi_\gamma \mid \gamma < \beta\} = \pi : \alpha \cdot_o \beta \rightarrowtail \alpha \cdot \beta.$$

But this π is onto $\alpha \cdot \beta$ since

$$\alpha \cdot \beta = \bigcup_{\gamma<\beta} \alpha \cdot \gamma$$

by the definition of ordinal multiplication, and so π is a similarity and we have the desired $\alpha \cdot_o \beta =_o \alpha \cdot \beta$.

The results about associativity and (not)-commutativity follow now from Problems **x7.9** and **x7.8**.

Finally, if $\pi : \beta \rightarrowtail \gamma$ is a proper initial similarity, then so is (easily) the function $\rho : \alpha \cdot_o \beta \rightarrowtail \alpha \cdot_o \gamma$ defined by

$$\rho(x, y) = (x, \pi(y)),$$

and it establishes the required $\alpha \cdot_o \beta <_o \alpha \cdot_o \gamma$.

Note. Another way to do the main part of this exercise is to show directly that for all α, β, the map

$$\pi(x, y) = x \cdot y \quad ((x, y) \in \alpha \times \beta)$$

is a similarity of $\alpha \cdot_o \beta$ with $\alpha \cdot \beta$. The proof (by ordinal recursion on β) is very similar to the argument we just gave, but it involves a somewhat more detailed "chasing" of similarities.

12.23. We prove the claimed identity by ordinal induction on γ, simultaneously for all α and β: i.e., we prove by induction on γ that

$$(\forall \alpha, \beta)[\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma].$$

The basis $\gamma = 0$ is trivial.
For $\gamma = \delta + 1$, compute:

$$
\begin{aligned}
\alpha \cdot (\beta + (\delta + 1)) &= \alpha \cdot ((\beta + \delta) + 1) \\
&= \alpha \cdot (\beta + \delta) + \alpha \quad \text{(def. of $\cdot$)} \\
&= \alpha \cdot \beta + \alpha \cdot \delta + \alpha \quad \text{(ind. hyp.)} \\
&= \alpha \cdot \beta + \alpha \cdot (\delta + 1).
\end{aligned}
$$

Finally, for limit γ, we observe first that

$$\bigcup_{\eta<\gamma}\{\beta + \eta\} \text{ is a limit ordinal,}$$

because for each $\eta < \gamma$ there is a ζ such that $\eta < \zeta < \gamma$, and then $\beta + \eta < \beta + \zeta$ by (12-26). Using this to distinguish cases in the definition of multiplication,

we compute:

$$\alpha \cdot (\beta + \gamma) = \alpha \cdot (\textstyle\bigcup_{\eta<\gamma}\{\beta + \eta\})$$
$$= \textstyle\bigcup_{\eta<\gamma}\{\alpha \cdot (\beta + \eta)\}$$
$$= \textstyle\bigcup_{\eta<\gamma}\{\alpha \cdot \beta + \alpha \cdot \eta\} \quad \text{(ind. hyp.)}$$
$$= \alpha \cdot \beta + \alpha \cdot \gamma \qquad \text{(def. of +)},$$

where for the last inference we have used (12-27) to infer that if $\eta < \zeta < \gamma$, then $\alpha \cdot \eta < \alpha \cdot \zeta < \alpha \cdot \gamma$, so that $\alpha \cdot \gamma$ is a limit ordinal and $\alpha \cdot \gamma = \bigcup_{\eta<\gamma} \alpha \cdot \eta$.

12.26. Suppose first $\kappa = (\mu\xi \in \text{ON})[\xi =_c A]$, and assume towards a contradiction that $\kappa =_c \alpha$ for some $\alpha < \kappa$; but then $A =_c \alpha < \kappa$, contradicting the definition of κ. For the converse (and the second claim), if κ is not equinumerous with any $\alpha < \kappa$, then, $\kappa = |\kappa|$, since $\kappa =_c \kappa$, and so $\kappa \in \text{Card}_v$.

12.27. If $A =_c B$, then, for all $\xi \in \text{ON}$,

$$A =_c \xi \iff B =_c \xi,$$

and hence $(\mu\xi \in \text{ON})[A =_c \xi] = (\mu\xi \in \text{ON})[B =_c \xi]$. The third property of strong cardinal assignments is trivial when we assume the Axiom of Replacement: because $\{|X| \mid X \in \mathscr{E}\}$ is the image of $\mathscr{E}$ under the definite operation $X \mapsto |X|$, and so it is a set.

12.30. If **AC** holds, then every set is equinumerous with an ordinal λ, and then $|A|$ is a von Neumann cardinal, by definition. Conversely, if $|A| = \aleph_\alpha$, then $A =_c \aleph_\alpha$, and so A is well orderable.

12.31. Granting **AC**, the Generalized Continuum Hypothesis is the claim that for every infinite cardinal number κ,

$$|\mathcal{P}(\kappa)| = 2^\kappa = \kappa^+;$$

which becomes exactly the claimed identity since the cardinals are exactly the alephs and $\aleph_\alpha^+ = \aleph_{\alpha+1}$.

INDEX

Undergraduate Texts in Mathematics

Undergraduate Texts in Mathematics

Prenowitz/Jantosciak: Join Geometries.

Priestley: Calculus: A Liberal Art. Second edition.

Protter/Morrey: A First Course in Real Analysis. Second edition.

Protter/Morrey: Intermediate Calculus. Second edition.

Pugh: Real Mathematical Analysis.

Roman: An Introduction to Coding and Information Theory.

Roman: Introduction to the Mathematics of Finance: From Risk management to options Pricing.

Ross: Differential Equations: An Introduction with Mathematica®. Second Edition.

Ross: Elementary Analysis: The Theory of Calculus.

Samuel: Projective Geometry. *Readings in Mathematics*.

Saxe: Beginning Functional Analysis

Scharlau/Opolka: From Fermat to Minkowski.

Schiff: The Laplace Transform: Theory and Applications.

Sethuraman: Rings, Fields, and Vector Spaces: An Approach to Geometric Constructability.

Sigler: Algebra.

Silverman/Tate: Rational Points on Elliptic Curves.

Simmonds: A Brief on Tensor Analysis. Second edition.

Singer: Geometry: Plane and Fancy.

Singer: Linearity, Symmetry, and Prediction in the Hydrogen Atom.

Singer/Thorpe: Lecture Notes on Elementary Topology and Geometry.

Smith: Linear Algebra. Third edition.

Smith: Primer of Modern Analysis. Second edition.

Stanton/White: Constructive Combinatorics.

Stillwell: Elements of Algebra: Geometry, Numbers, Equations.

Stillwell: Elements of Number Theory.

Stillwell: The Four Pillars of Geometry.

Stillwell: Mathematics and Its History. Second edition.

Stillwell: Numbers and Geometry. *Readings in Mathematics*.

Strayer: Linear Programming and Its Applications.

Toth: Glimpses of Algebra and Geometry. Second Edition. *Readings in Mathematics*.

Troutman: Variational Calculus and Optimal Control. Second edition.

Valenza: Linear Algebra: An Introduction to Abstract Mathematics.

Whyburn/Duda: Dynamic Topology.

Wilson: Much Ado About Calculus.