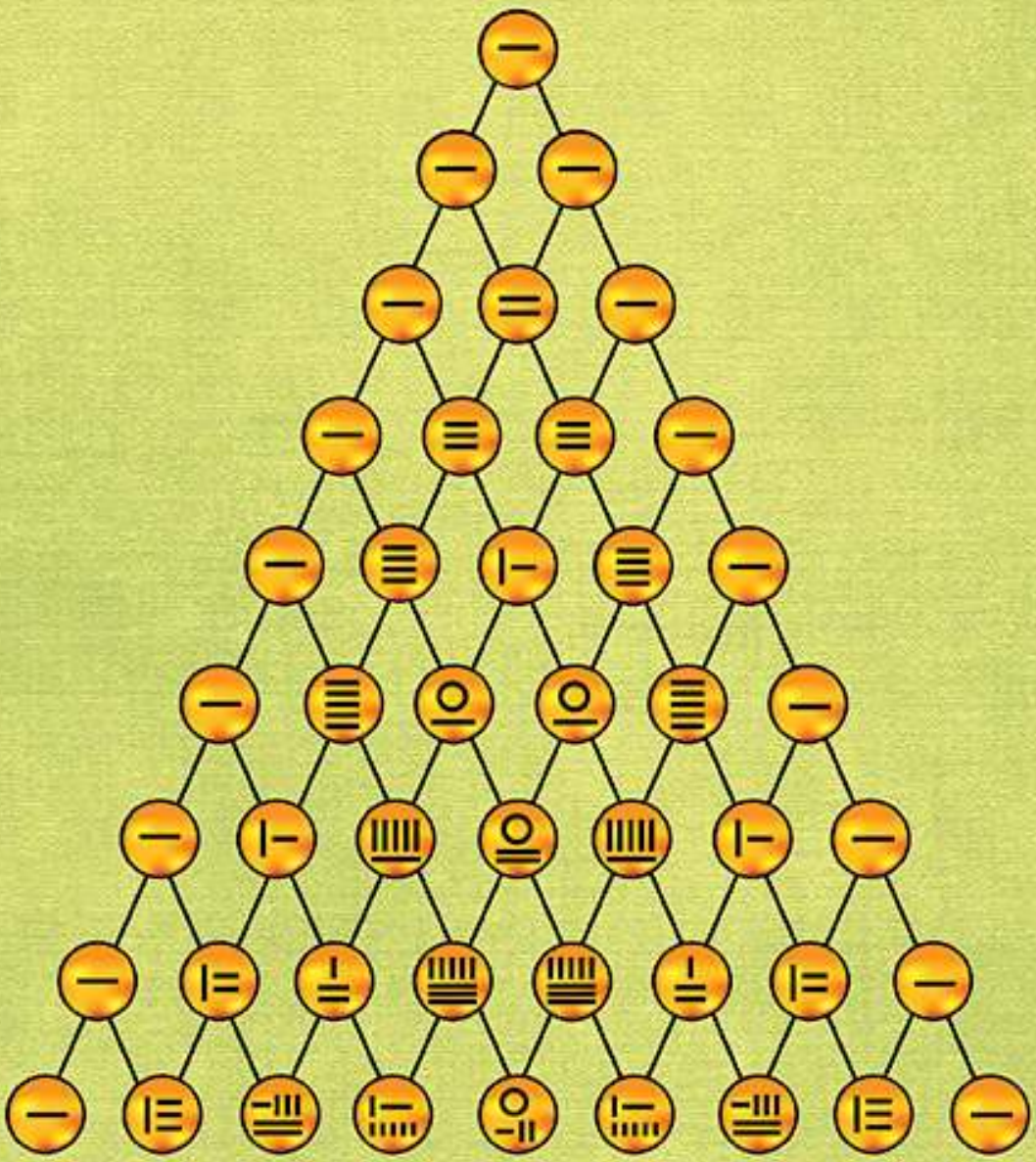
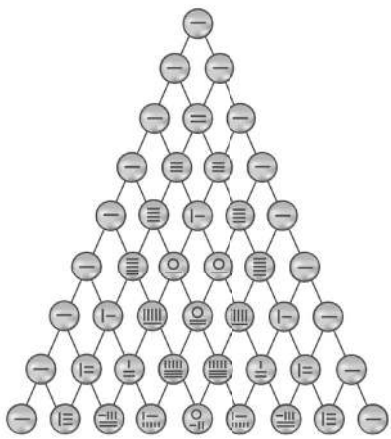
Basic Language of Mathematics



Juan Jorge Schäffer

World Scientific

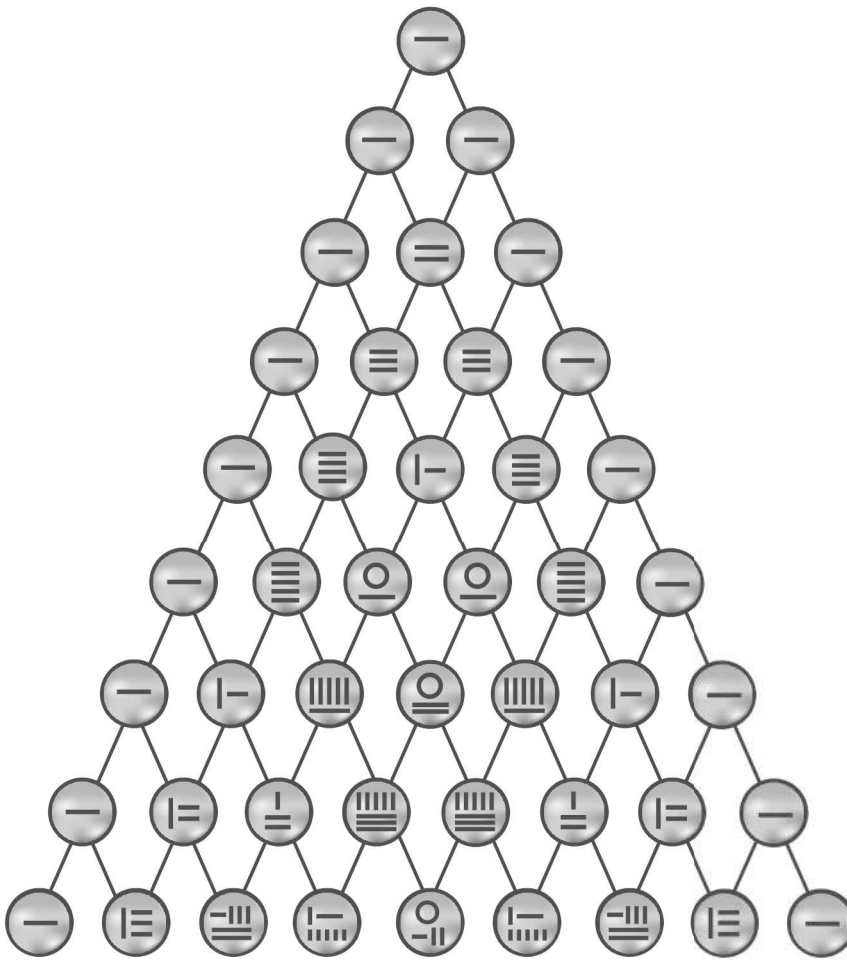
Basic Language of Mathematics



This page intentionally left blank

Basic Language of Mathematics

Juan Jorge Schäffer
Carnegie Mellon University, USA



World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Schäffer, Juan Jorge, author.

Basic language of mathematics / by Juan Jorge Schäffer (Carnegie Mellon University, USA).

pages cm

Includes indexes.

ISBN 978-9814596091 (hardcover : alk. paper)

1. Mathematics. 2. Mathematical analysis. I. Title.

QA37.3.S33 2014

510--dc23

2014009124

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

The image on the book cover was adapted from the triangle of binomial coefficients depicted in *Si Yuan Yu Jian*, that was published in 1303 by Zhu Shijie.

Printed in Singapore

PREFACE

This work is intended to provide a presentation (with some carefully chosen alternatives) of essential ingredients of mathematical discourse. It deals with concepts: *set mapping, family, order, natural number, real number, finite and infinite sets, countability; with procedures: proof by induction, recursive definition, fixed-point theorem*, and with some immediate applications. It concludes with a sketch of maximality axioms for infinite sets, with the Axiom of Choice and other useful equivalents.

The resulting account was designed to support a component of a comprehensive three-semester honors program, "Mathematical Studies", conducted for years at Carnegie Mellon University, by the author with the inspiring partnership of Walter Noll.

The author owes special thanks to Ms Nancy J. Watson for her superb and tireless work on the preparation of the manuscript.

The author also records his appreciation of Ms Kwong Lai Fun of World Scientific for her unerring eye and tireless and insightful attention to every detail that have given this work its final polish.

Some Symbols

The symbols for logical connectives and quantifiers used in this work are:

$\Rightarrow$ read "only if" or "implies"

$\Leftrightarrow$ read "if and only if" or "is equivalent to"

$\forall$ read "for all" or "for every"

$\exists$ read "for some" or "there exists such that".

The following symbols denote certain sets of numbers, regarded as endowed with the usual arithmetical operations and order relations:

$\mathbb{N}$ the set of all natural numbers (including zero)

$\mathbb{Z}$ the set of all (positive and negative) integers

$\mathbb{Q}$ the set of all rational numbers

$\mathbb{R}$ the set of all real numbers

$\mathbb{P}$ the set of all positive real numbers (including zero).

These sets, with their structures, will be formally introduced in due course, but they are familiar enough to be used, before that introduction, in examples and remarks.

Finally, some symbols are used to facilitate the reading:

■ indicates the end of a proof or of some formally labelled example, remark, etc. It is omitted after the statement of a proposition for which no proof is provided.

* placed after the heading of an example or remark signals the use of material not yet formally introduced or even lying outside the purview of this work. (For another use of the asterisk see p. 6.)

▼ placed in the margin encompass a paragraph, a passage, or even almost a whole
▲ chapter that the reader is encouraged to skip, at least initially, unless particularly motivated.

• records that the proposition or passage or phrase to which it is prefixed uses the Axiom of Choice in one of its variant forms. Little is lost, at least initially, if the presence of this symbol is ignored. (For the Axiom of Choice, see p. 59 and Chapter 17.)

CONTENTS

Chapter 1

SETS

11. Introduction

In this chapter we present the essential terminology, notation, and facts pertaining to *sets* and *members* of sets, the most elementary ingredients of current mathematical discourse. Successive chapters will introduce other fundamental ingredients, such as mappings, relations, numbers ...

We do not aim either at a philosophical elucidation of these concepts, or at a rigorous account of the foundations of mathematics as they are currently understood. These are specialized subjects, attractive in their own right, but of little immediate concern to most practicing mathematicians or users of mathematics; to engage seriously in their study requires considerable mathematical experience and maturity. We merely intend to clarify the usage of the fundamental concepts, derive their simplest properties and relationships, and make the language they constitute available for use.

Underlying all mathematical discourse are concepts and rules of logic. For an exposition of the relevant logical tools that is particularly well suited to our approach, we refer to Chapter 2 of A. M. Gleason, *Fundamentals of Abstract Analysis*. The book as a whole is recommended for its choice of contents and its professional style. Although the usage adopted in it differs in many particulars from ours, the book is an excellent aid to understanding. Chapter 1 and part of Chapter 3 of that book should also be studied in conjunction with the present chapter.

Our use of *equality* is exclusively as follows: the assertion $a = b$ means that the object (designated by the symbol) a and the object (designated by the symbol) b are one and the same. In practice, what stands on either side of $=$ may be a complicated array of typographical symbols. The negation of the assertion $a = b$ is denoted by $a \neq b$, and if it holds we say that a and b are **distinct** objects, or that a is **distinct from** b.

The symbols $:=$ and $=:$ are used in definitions: $a := b$ or, equivalently, $b =: a$ means that a is *defined* to be (equal to) b. The colon stands on the side of the *definiendum* a, the term to be defined, in order to contrast it with the *definiens* b, the defining term. When the definition of an object is given in words, we distinguish

the words constituting the definiendum by means of boldface type. For example, "A **square** is defined to be a rectangle with equal sides," or "A rectangle with equal sides is called a **square**."

Similar usages occur in the definition of a property of an object by means of an appropriate predicate. Thus, "A number n is said to be **even** if 2 divides n" (in this style of definition it would be redundant to add "and only if"). In slightly more symbolic form we might write

$$\text{For every number } n, (n \text{ is } \mathbf{even}) \ :\Leftrightarrow \ (2 \text{ divides } n);$$

the symbol $:\Leftrightarrow$ may be read "means by definition that" or "is equivalent by definition to".

12. Sets and their members

In our presentation we give no formal definition of the concept of *set*. Speaking vaguely, whenever objects are thought of as collected into a definite whole, a set describes this state of affairs. A **set** is determined whenever there is an unambiguous answer to the question whether any given object belongs to it, even if this answer may be very difficult to ascertain in practice, or is currently unknown.

Each of the objects constituting a set S is called a **member of** S; sometimes the term *element of* S is encountered. The assertion that an object s is a member of the set S is denoted by $s \in S$, and this assertion may be equivalently expressed by "s **is contained in** S", or "S **contains** s", or, more informally, by "s is in S", or "s belongs to S". Objects are often introduced as members of given sets: phrases such as "let $s \in S$ be given", "choose $s \in S$ such that", "for all $s \in S$" are common; in them, the symbol $\in$ has to be construed to give a grammatically correct reading, as, e.g., "choose a member s of S such that". The negation of the assertion $s \in S$ is denoted by $s \notin S$.

We regard it as an essential feature of the notion of *set* that the identity of a set is determined by its membership: sets that have precisely the same members are one and the same. More formally, for given sets S and T we have

$$S = T \quad \Leftrightarrow \quad (\forall x, \quad x \in S \quad \Leftrightarrow \quad x \in T)\,. \tag{12.1}$$

It is useful to have a synonym for the term *set*. Thus a **collection** is a set, and this term is used, in particular, when the members of the set in question are themselves sets (or at least when it is material to state this fact): "collection of sets" is more usual than "set of sets". On occasion, the term *class* is also used as a synonym of the term *set*. The term *family*, however, has acquired a quite different meaning (see Section 41), and should never be used as a synonym for *set*.

13. Inclusion

Let the sets A and B be given. If every member of A is also a member of B, we say that A **is included in** B, or that B **includes** A, or that A is a **subset of** B, and we write $A \subset B$ or $B \supset A$. In symbols,

$$A \subset B \quad :\Leftrightarrow \quad B \supset A \quad :\Leftrightarrow \quad (\forall x \in A, \quad x \in B) \quad \Leftrightarrow \quad (\forall x, \quad x \in A \quad \Rightarrow \quad x \in B)\,.$$

We note that $A \subset A$ for every set A. If A is included in B, but is not equal to B, we say that A **is properly included in** B or that B **properly includes** A, or that A is a **proper subset** of B, and we write $A \subsetneqq B$ or $B \supsetneqq A$.

Many mathematicians use a different notation: they write $\subseteq$, $\supseteq$, $\subset$, $\supset$ where we, with many other mathematicians, write $\subset$, $\supset$, $\subsetneqq, \supsetneqq$, respectively. Unless the context is quite unambiguous, it is still advisable to declare one's choice between these incompatible conventions in any piece of mathematical writing.

We note the careful distinction between the terms *contain* and *include*: "S contains s" means $s \in S$, i.e., "s is a member of S"; "S includes T" means $T \subset S$, i.e., "every member of T is a member of S". (In very unusual circumstances, a set S may both contain another set T — as a member — and include T — as a subset — simultaneously!) This useful distinction was recommended by Paul Richard Halmos. We shall adhere to it strictly; most mathematical writing does not.

13A. **Proposition.** *Let the sets A, B, C be given. Then*

$$(A \subset B \text{ and } B \subset A) \;\Leftrightarrow\; A = B \tag{13.1}$$

$$(A \subset B \text{ and } B \subset C) \;\Rightarrow\; A \subset C \tag{13.2}$$

$$(A \subsetneqq B \text{ and } B \subset C) \;\Rightarrow\; A \subsetneqq C \tag{13.3}$$

$$(A \subset B \text{ and } B \subsetneqq C) \;\Rightarrow\; A \subsetneqq C. \tag{13.4}$$

Proof. (13.1) is a restatement of (12.1) with $S := A$ and $T := B$. (13.2) follows trivially from the definition of inclusion. If $A \subset B$ and $B \subset C$ and $A = C$, then $A \subset B$ and $B \subset A$; by (13.1) it follows that $A = B = C$; this observation, together with (13.2), establishes (13.3) and (13.4). ∎

We shall abbreviate "$A \subset B$ and $B \subset C$" to "$A \subset B \subset C$", and similarly for more sets, as well as for formulas such as "$A \subsetneqq B \subset C$" and "$A \supset B \supsetneqq C$."

We remark that (13.1) describes the most frequently used strategic scheme for proving equality of given sets.

14. Set formation

Most sets are defined by specifying the properties that their members must have. Such properties are embodied in *predicates*, or sentence fragments: in "n is even", "2 divides n", the sentence fragments "is even", "2 divides " are predicates, and they become complete assertions when "the blanks are filled in". It is a matter for logical analysis to determine whether a string of words, symbols, and blanks corresponds to a well-formed predicate.

If $P(\ )$ is a predicate, the set consisting precisely of all objects x that *satisfy* the assertion $P(x)$ (i.e., all x for which $P(x)$ holds, or is true) is denoted by

$$\{x \mid P(x)\}\,. \tag{14.1}$$

This is read "the set of all x such that $P(x)$ (holds)." This set is defined more precisely by requiring that

$$\forall y, \quad y \in \{x \mid P(x)\} \quad :\Leftrightarrow \quad P(y). \tag{14.2}$$

Most frequently, the members of the set to be defined are assumed *a priori* to be members of a given set. If the set S and the predicate $P(\ )$ are given, the set consisting of all members x of S that satisfy the assertion $P(x)$ is denoted by

$$\{x \in S \mid P(x)\}\,. \tag{14.3}$$

This is read "the set of all x (contained) in S such that $P(x)$ (holds)." This set is obviously included in S. The following assertion holds:

$$\forall y \in S, \quad y \in \{x \in S \mid P(x)\} \quad \Leftrightarrow \quad P(y). \tag{14.4}$$

For instance, the set consisting of all even natural numbers is $\{n \in \mathbb{N} \mid 2 \text{ divides } n\}$, and a number $n \in \mathbb{N}$ is a member of this set (i.e., is even) if and only if 2 divides n.

In the notations (14.1) and (14.3), the symbol x is a "dummy", and may therefore be replaced, without changing the set, by any single symbol that does not appear in the explicit formula for $P(\ )$ (or in that for S).

Let the set S and the predicates $P(\ )$ and $Q(\ )$ be given. Then the definitions and (14.4) yield

$$(\{x \in S \mid P(x)\} \subset \{x \in S \mid Q(x)\}) \;\Leftrightarrow\; (\forall x \in S,\ P(x) \;\Rightarrow\; Q(x))\,. \tag{14.5}$$

We observe that, for every set S and predicate $P(\ )$,

$$\{x \in S \mid P(x)\} \stackrel{*}{=} \{x \mid x \in S \text{ and } P(x)\}\,, \tag{14.6}$$

thus potentially reducing the notation (14.3) to an instance of (14.1). The notation (14.3) is not, however, merely a convenient abbreviation. The set-forming notation

(14.1) relies on the premise that, for the given predicate $P(\ )$, there is a set consisting precisely of all objects x that satisfy $P(x)$. Indiscriminate reliance on this premise for all conceivable predicates may unfortunately lead to complications, known as "paradoxes". We may regard it as a guideline of well-formulated mathematical discourse that the set-forming notation (14.3) is to be preferred whenever it is available. We shall therefore restrict our use of (14.1) to a few indispensable occasions (and indicate its use by means of an asterisk, as in (14.6)); it is taken for granted that this exceptional use is permissible.

15. Special sets

The set

$$\emptyset :\stackrel{*}{=} \{x \mid x \neq x\}$$

has no members: indeed, for every object y, the assertion $y \neq y$ is false, and (14.2) disqualifies y for membership in $\emptyset$. By (12.1), $\emptyset$ is the only set with no members; it is called the **empty set**. $\emptyset$ is a subset of every set: for every given set S we have $\emptyset = \{x \in S \mid x \neq x\} \subset S$.

A set S is said to be **empty** if $S = \emptyset$ and **non-empty** if $S \neq \emptyset$; the terms *void* and *non-void* are sometimes encountered.

An essential peculiarity of the empty set must be noted. For every predicate $P(\ )$, the assertion $(\forall x \in \emptyset,\ P(x))$ is true and the assertion $(\exists x \in \emptyset,\ P(x))$ is false. Indeed, this is the only case where a universally quantified assertion does not imply the corresponding existentially quantified one: for every set S and predicate $P(\ )$,

$$S \neq \emptyset \quad \Leftrightarrow \quad ((\forall x \in S,\ \ P(x)) \quad \Rightarrow \quad (\exists x \in S, \quad P(x)))\,. \tag{15.1}$$

An assertion of the form $(\forall x \in S,\ P(x))$ is said to *hold vacuously* if $S = \emptyset$ and it is desired to stress that it holds for that reason. If a set S is known to be non-empty, the assertion $(\exists x \in S,\ )$ holds, and we may *choose* $s \in S$.

For every object s we define the set

$$\{s\} :\stackrel{*}{=} \{x \mid x = s\}$$

whose only member is s. This set must be carefully distinguished from the object s itself. For instance, $\emptyset$ is empty, but the collection $\{\emptyset\}$ is not. A set S is called a **singleton** if $S = \{s\}$ for some object s. The set $\{s\}$ is sometimes called the **singleton of** s. We note that (12.1) implies

$$\forall s, t, \quad s = t \quad \Leftrightarrow \quad s \in \{t\} \quad \Leftrightarrow \quad \{s\} = \{t\}\,, \tag{15.2}$$

and, for every set S,

$$\forall s, \quad s \in S \quad \Leftrightarrow \quad \{s\} \subset S. \tag{15.3}$$

15A. Proposition. *Let the set S be given. Then S is empty or a singleton if and only if*

$$\forall s, s' \in S, \quad s = s'. \tag{15.4}$$

Proof. If S is empty, (15.4) holds, since it is an abbreviation of

$$\forall s \in S, \quad (\forall s' \in S, \quad s = s'),$$

which holds vacuously. If S is a singleton, we may choose t such that $S = \{t\}$. By the definition of $\{t\}$ and by (14.2), all $s, s' \in S = \{t\}$ satisfy $s = t = s'$, and thus (15.4) holds.

Assume now that S satisfies (15.4) and is not empty. We may then choose $t \in S$. By (15.3), $\{t\} \subset S$. On the other hand, let $s \in S$ be given. By (15.4) with $s' := t$ we have $s = t$, whence $s \in \{t\}$ by (15.2); since $s \in S$ was arbitrary, we conclude that $S \subset \{t\}$. By (13.1) it follows that $S = \{t\}$, so that S is a singleton. ■

Many mathematical objects are defined according to the following scheme: "The only such that is called "; e.g., "For every positive real number s, the only positive real number r such that $r^2 = s$ is called the **(positive) square root of** s." To formulate such a definition, two steps are necessary. First, to ascertain that a given set S is a singleton — usually by proving that it satisfies (15.4) and is not empty. In the example, this set would be $\{r \in \mathbb{P} \mid r^2 = s\}$. Second, to "extract" the only, or *unique*, member from the singleton S. When the set S is known to be a singleton, we shall write

$$s :\in S$$

to mean that the object s is *defined* to be the only member of S. In the example we should therefore write $\sqrt{s} :\in \{r \in \mathbb{P} \mid r^2 = s\}$.

For given objects $s,\ t$, we define the set

$$\{s, t\} :\stackrel{*}{=} \{x \mid x = s \text{ or } x = t\}.$$

We note that

$$\forall s, t, \quad \{s, t\} = \{t, s\}. \tag{15.5}$$

$$\forall s, \quad \{s, s\} = \{s\}. \tag{15.6}$$

A set S is called a **doubleton** if $S = \{s, t\}$ for some *distinct* objects $s,\ t$.

We observe that we could have used (15.6) to *define* $\{s\}$, thus avoiding one minor instance of the undesirable use of the set-forming notation (14.1).

Quantification over singletons and doubletons takes specially simple forms. For every predicate $P(\ )$ we have

$$\forall s,\ (\forall x \in \{s\}\,, P(x)) \ \Leftrightarrow\ P(s) \ \Leftrightarrow\ (\exists x \in \{s\}\,, P(x))$$

$$\forall s,t,\ (\forall x \in \{s,t\}\,, P(x)) \ \Leftrightarrow\ (P(s) \text{ and } P(t))$$

$$\forall s,t,\ (\exists x \in \{s,t\}\,, P(x)) \ \Leftrightarrow\ (P(s) \text{ or } P(t))\,.$$

For every given set S we consider the collection consisting precisely of all subsets of S. This collection is

$$\mathfrak{P}(S) :\overset{*}{=} \{T \mid T \text{ is a set, and } T \subset S\}\,;$$

it is called the **power-set of** S. By (14.2) we have

$$\text{for all sets } S,T, \quad T \subset S \quad \Leftrightarrow \quad T \in \mathfrak{P}(S). \tag{15.7}$$

We also define $\mathfrak{P}^{\times}(S) := \{T \in \mathfrak{P}(S) \mid T \neq \emptyset\}$, the collection of all non-empty subsets of S.

16. Basic operations

Let a collection of sets $\mathcal{C}$ be given. We consider the set whose members are precisely the members of any set in $\mathcal{C}$; more precisely, the set

$$\bigcup \mathcal{C} :\overset{*}{=} \{x \mid x \in A \text{ for some } A \in \mathcal{C}\};$$

this set is called the **union of** the collection $\mathcal{C}$. We should also like to consider the set whose members are precisely those objects that are members of all sets in $\mathcal{C}$ simultaneously, i.e.,

$$\{x \mid x \in A \text{ for all } A \in \mathcal{C}\}. \tag{16.1}$$

This formula becomes questionable when $\mathcal{C}$ is the empty collection: *every* conceivable object would then qualify for membership, by (14.2), a situation to be avoided for many reasons. We shall therefore frame our definition so as to exclude this eventuality. Moreover, if $\mathcal{C}$ is not empty, then $x \in A$ for *all* $A \in \mathcal{C}$ implies that $x \in A$ for *some* $A \in \mathcal{C}$ (by (15.1)); thus every object qualifying for membership in the set proposed in (16.1) is already a member of $\bigcup \mathcal{C}$. In accordance with our guidelines for set-forming (Section 14), we shall therefore define, for every non-empty collection of sets $\mathcal{C}$, the set

$$\bigcap \mathcal{C} := \{x \in \bigcup \mathcal{C} \mid x \in A \text{ for all } A \in \mathcal{C}\} \overset{*}{=} \{x \mid x \in A \text{ for all } A \in \mathcal{C}\};$$

this set is called the **intersection of** the collection $\mathcal{C}$.

We observe that

$$\bigcup \emptyset = \emptyset \tag{16.2}$$

$$\text{for every set } A, \quad \bigcup\{A\} = \bigcap\{A\} = \bigcup \mathfrak{P}(A) = A \quad \text{and} \quad \bigcap \mathfrak{P}(A) = \emptyset. \tag{16.3}$$

For given sets A, B, we introduce the notation

$$A \cup B \;:=\; \bigcup\{A, B\} \;\overset{*}{=}\; \{x \mid x \in A \text{ or } x \in B\},$$

$$A \cap B \;:=\; \bigcap\{A, B\} \;=\; \{x \in A \mid x \in B\} \;\overset{*}{=}\; \{x \mid x \in A \text{ and } x \in B\};$$

these sets are called the **union of** A **and** B, and the **intersection of** A **and** B, respectively. The sets A and B are said to be **disjoint** if $A \cap B = \emptyset$; A and B are said to **meet**, and A is said to **meet** B, if $A \cap B \neq \emptyset$.

Let the set X and the collection $\mathcal{C}$ of subsets of X be given, so that $\mathcal{C} \subset \mathfrak{P}(X)$. We can then define the **union of $\mathcal{C}$ with respect to** X, and the **intersection of $\mathcal{C}$ with respect to** X, as

$$\bigcup{}^X\mathcal{C} \;:=\; \{x \in X \mid x \in A \text{ for some } A \in \mathcal{C}\},$$

$$\bigcap{}^X\mathcal{C} \;:=\; \{x \in X \mid x \in A \text{ for all } A \in \mathcal{C}\},$$

respectively. It is easy to verify that

$$\bigcup{}^X\mathcal{C} = \bigcup\mathcal{C}, \tag{16.4}$$

so that the notion of "union with respect to X" is in fact redundant. On the other hand, it is also easy to verify that

$$\bigcap{}^X\mathcal{C} = \bigcap(\mathcal{C} \cup \{X\}) = \begin{cases} \bigcap\mathcal{C} & \text{if } \mathcal{C} \neq \emptyset \\ X & \text{if } \mathcal{C} = \emptyset. \end{cases} \tag{16.5}$$

The operations of forming the union and intersection of collections obey certain fundamental rules that reflect the rules for logical connectives and quantifiers. We now record some of these rules.

16A. Proposition. *Let the collections of sets $\mathcal{C}$, $\mathcal{D}$, and the set X be given. Then*

$$\bigcup(\mathcal{C} \cup \mathcal{D}) \;=\; (\bigcup\mathcal{C}) \cup (\bigcup\mathcal{D}) \tag{16.6}$$

$$\bigcap(\mathcal{C} \cup \mathcal{D}) \;=\; (\bigcap\mathcal{C}) \cap (\bigcap\mathcal{D}) \quad \textit{if } \mathcal{C} \neq \emptyset,\ \mathcal{D} \neq \emptyset \tag{16.7}$$

$$\bigcap{}^X(\mathcal{C} \cup \mathcal{D}) \;=\; (\bigcap{}^X\mathcal{C}) \cap (\bigcap{}^X\mathcal{D}) \quad \textit{if } \mathcal{C}, \mathcal{D} \subset \mathfrak{P}(X) \tag{16.8}$$

$$\bigcup\mathcal{C} \;\subset\; \bigcup\mathcal{D} \quad \textit{if } \mathcal{C} \subset \mathcal{D} \tag{16.9}$$

$$\bigcap\mathcal{C} \;\supset\; \bigcap\mathcal{D} \quad \textit{if } \mathcal{C} \neq \emptyset \textit{ and } \mathcal{C} \subset \mathcal{D} \tag{16.10}$$

$$\bigcap{}^X\mathcal{C} \;\supset\; \bigcap{}^X\mathcal{D} \quad \textit{if } \mathcal{C} \subset \mathcal{D} \subset \mathfrak{P}(X). \tag{16.11}$$

16B. Proposition. *Let the sets $A,\ B,\ C$ be given. Then*

$$A \cup A = A \cap A = A \cup \emptyset = A \ \text{ and } \ A \cap \emptyset = \emptyset \tag{16.12}$$

$$A \cup B = B \cup A \ \text{ and } \ A \cap B = B \cap A \tag{16.13}$$

$$(A \cup B) \cup C = A \cup (B \cup C) \ \text{ and } \ (A \cap B) \cap C = A \cap (B \cap C) \tag{16.14}$$

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \text{ and } (A \cap B) \cup C = (A \cup C) \cap (B \cup C) \tag{16.15}$$

$$A \subset B \ \Leftrightarrow \ A \cup B = B \ \Leftrightarrow \ A \cap B = A \tag{16.16}$$

$$A \subset B \ \Leftrightarrow \ (A \cup C \subset B \cup C \ \text{ and } \ A \cap C \subset B \cap C). \tag{16.17}$$

16C. Remarks. (a): Because of rule (16.14), we may unambiguously write $A \cup B \cup C$ and $A \cap B \cap C$ without parentheses; and the same license is in effect when writing unions and intersections of more sets, such as $A \cup B \cup C \cup D$.

(b): For all $s,\ t$, it follows from the definitions that

$$\{s, t\} = \{s\} \cup \{t\}. \tag{16.18}$$

For all $s,\ t,\ u,\ v$, we may define $\{s, t, u\} := \{s\} \cup \{t\} \cup \{u\}$ and $\{s, t, u, v\} := \{s\} \cup \{t\} \cup \{u\} \cup \{v\}$. We observe that $\{s, t, t\} = \{s, t\},\ \{s, s, s\} = \{s\}$, etc.

(c): For any given sets $A,\ B,\ C$, we have $\bigcup\{A, B, C\} = A \cup B \cup C$ and $\bigcap\{A, B, C\} = A \cap B \cap C$, and similar formulas hold for sets $A,\ B,\ C,\ D$. ■

Let the sets A and B be given. We consider the set consisting of all members of A that are not in B, i.e., the set

$$A \backslash B := \{x \in A \mid x \notin B\}$$

to be read "A **without** B"; this set called the **set-difference of A and B**. When considering subsets of some set X that is fixed throughout some discussion, it is sometimes useful to call the set $X \backslash A$ the **complement of A in X** (or **with respect to** X), for every $A \in \mathfrak{P}(X)$. We record some fundamental rules connecting set-difference, union, and intersection; they reflect rules for negation and other logical connectives.

16D. Proposition. *Let the sets $A,\ B,\ C$ be given. Then*

$$A \backslash A = \emptyset \backslash A = \emptyset \ \text{ and } \ A \backslash \emptyset = A \tag{16.19}$$

(16.20) $(A \cup B) \backslash C = (A \backslash C) \cup (B \backslash C)$ *and* $(A \cap B) \backslash C = (A \backslash C) \cap B = (A \backslash C) \cap (B \backslash C)$

(16.21) $A \backslash (B \cup C) = (A \backslash B) \cap (A \backslash C) = (A \backslash B) \backslash C$ *and* $A \backslash (B \cap C) = (A \backslash B) \cup (A \backslash C)$

(16.22) $$A \backslash (B \backslash C) = (A \backslash B) \cup (A \cap C).$$

There are more general rules concerning these operations, when collections of sets and even collections of collections of sets are involved; these rules are more conveniently formulated in the language of *families*, and will be recorded later, in Sections 44 and 45.

We mention one more operation. For given sets A, B, we consider the set

$$A \triangle B := (A \backslash B) \cup (B \backslash A);$$

this set is called the **symmetric difference of A and B**.

16E. Proposition. *Let the sets A, B, C be given. Then*

(16.23) $$A \triangle B = (A \cup B) \backslash (A \cap B)$$

(16.24) $$A \triangle A = \emptyset \quad \textit{and} \quad A \triangle \emptyset = A$$

(16.25) $$A \triangle B = B \triangle A$$

(16.26) $$(A \triangle B) \triangle C = A \triangle (B \triangle C)$$

(16.27) $$(A \triangle B) \cap C = (A \cap C) \triangle (B \cap C).$$

17. Pairs; product sets

For given objects $a,\ b$, we often need a mathematical object that expresses the idea "a, and then b", and thus depends on a, on b, and on the priority of a with respect to b. We take it for granted that such an object is available; it is denoted by (a, b); and these objects, provided for all $a,\ b$, are called **pairs**. (We note that a may well be equal to b.) The essential fact about pairs is that from the pair (a, b) it is possible to retrieve a and then b. (One should not be misled by the *notation* "(a, b)", which incorporates the symbols "a" and "b"; like any mathematical object, (a, b) may be denoted by some other symbol or name, e.g., $\pi := (a, b)$.) Specifically, we take it for granted that

$$\forall a, b, c, d,\ (a, b) = (c, d) \ \Leftrightarrow\ (a = c \text{ and } b = d). \tag{17.1}$$

This property of pairs allows us to refer to a as *the* **former component of** the pair (a, b), and to b as *the* **latter component of** the pair (a, b).

▼ To be a bit more formal: an object π is a **pair** if $\pi = (a, b)$ for some $a,\ b$. If π is a pair, (17.1) implies that the sets $\{x \mid \exists y,\ (x, y) = \pi\}$ and $\{y \mid \exists x,\ (x, y) = \pi\}$ are singletons, and we may define

$$\textbf{former component of } \pi :\overset{*}{\in} \{x \mid \exists y,\ (x, y) = \pi\}$$

$$\textbf{latter component of } \pi :\overset{*}{\in} \{y \mid \exists x,\ (x, y) = \pi\}.$$

▲

Let the sets $A,\ B$ be given. It is necessary to consider the set consisting of those pairs whose former components are members of A and whose latter components are members of B. This set is

$$A \times B :\overset{*}{=} \{\pi \mid \exists a \in A,\ \exists b \in B,\ (a, b) = \pi\};$$

it is called the **product set of** A **and** B.

When defining subsets of the product set $A \times B$, it is customary to write

$$\{(x, y) \in A \times B \mid P(x, y)\}$$

as an abbreviation of $\{\pi \in A \times B \mid \exists x \in A, \exists y \in B,\ \pi = (x, y) \text{ and } P(x, y)\}$, where $P(\ ,\ )$ is a suitable two-place predicate. In particular, for every set A we define the set

$$\Delta_A := \{(x, y) \in A \times A \mid x = y\};$$

this set is called the **diagonal of** $A \times A$. This terminology is unambiguous, as will follow from Proposition 17B,(c).

17A. Remark. For given sets $A,\ B,\ C$, it is in general not the case that $(A \times B) \times C = A \times (B \times C)$. ■

17B. Proposition. (a): *Let the sets $A,\ B$ be given. Then $A \times B \neq \emptyset$ if and only if $A \neq \emptyset$ and $B \neq \emptyset$.*

(b): *Let the non-empty sets $A,\ B,\ C,\ D$ be given. Then $A \times B = C \times D$ (if and) only if $A = C$ and $B = D$. In particular, $A \times B = B \times A$ (if and) only if $A = B$.*

(c): *Let the sets $A,\ B$ be given. Then $A \times A = B \times B$ (if and) only if $A = B$.*

Proof. (a) is trivial. To prove (b), choose $a \in A,\ b \in B,\ c \in C,\ d \in D$. Let $x \in A$ be given; we have $(x, b) \in A \times B = C \times D$; hence $(x, b) = (z, u)$ for suitable $z \in C$ and $u \in D$. By (17.1) we must have $x = z \in C$. Since $x \in A$ was arbitrary, we conclude that $A \subset C$. Repeating this proof, with $A,\ B,\ C,\ D,\ b$ replaced by $C,\ D,\ A,\ B,\ d$, respectively, we also conclude that $C \subset A$, and hence $A = C$. The proof that $B = D$ is similar, and uses $a,\ c$. (c) now follows from (a) and (b). ∎

17C. Proposition. *Let the sets $A,\ B,\ C,\ D$ be given. Then*

$$(A \cup B) \times C = (A \times C) \cup (B \times C) \quad A \times (B \cup C) = (A \times B) \cup (A \times C) \tag{17.2}$$

$$(A \cap B) \times C = (A \times C) \cap (B \times C) \qquad A \times (B \cap C) = (A \times B) \cap (A \times C) \tag{17.3}$$

$$(A \backslash B) \times C = (A \times C) \backslash (B \times C) \qquad A \times (B \backslash C) = (A \times B) \backslash (A \times C) \tag{17.4}$$

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D) \tag{17.5}$$

$$\text{If } A \times B \neq \emptyset, \text{ then } (A \subset C \text{ and } B \subset D) \iff A \times B \subset C \times D. \tag{17.6}$$

There are rules more general than those recorded in Proposition 17C. They involve unions and intersections of collections of sets, and are best formulated in the language of *families*. They are special cases of the rules recorded in Section 44.

▼ **17D. Remark.** Many mathematicians have been uneasy about the idea of introducing the concept of *pair* as an undefined notion, subject to (17.1), and have looked for construction of *sets* that would serve the required purpose; they aimed at reducing the foundational complexity, at the cost of an artificial, counter-intuitive construction, containing structure that is irrelevant for the purpose at hand. It should be noted in passing that the most ingenuous idea of taking the set $\{a, b\}$ to be the pair (a, b) for all $a,\ b$ fails to satisfy (17.1) whenever $a \neq b$ and we set $c := b,\ d := a$.

The most frequently encountered successful construction of this kind defines pairs by setting $(a, b) := \{\{a\}, \{a, b\}\}$ for all $a,\ b$. Under this definition, $\bigcap(a, b) = \{a\}$ and $\bigcup(a, b) = \{a, b\}$. Thus every pair $\mathcal{P}$ is a non-empty collection of sets such that $\bigcap\mathcal{P}$ is a singleton and $(\bigcup\mathcal{P})\backslash(\bigcap\mathcal{P})$ is either a singleton or the empty set; and we may set

$$p :\in \bigcap \mathcal{P} \tag{17.7}$$

$$q \begin{cases} :\in (\bigcup \mathcal{P}) \backslash (\bigcap \mathcal{P}) & \text{if } \bigcup \mathcal{P} \neq \bigcap \mathcal{P} \\ := p & \text{if } \bigcup \mathcal{P} = \bigcap \mathcal{P}. \end{cases} \tag{17.8}$$

We then find that

$$\forall a, b, \quad \mathcal{P} = (a, b) \quad \Leftrightarrow \quad (a = p \text{ and } b = q).$$

This shows that (17.1) holds; and (17.7) and (17.8) provide the definitions of the *former* and *latter components of* the pair $\mathcal{P}$.

Let the sets A, B be given. With this definition of pairs, the *product set of* A *and* B becomes a subset of $\mathfrak{P}(\mathfrak{P}(A \cup B))$, namely

$$A \times B := \{\mathcal{P} \in \mathfrak{P}(\mathfrak{P}(A \cup B)) \mid \exists a \in A, \exists b \in B, \ \mathcal{P} = \{\{a\}, \{a, b\}\}\}.$$

▲

■

18. Partitions

A collection of sets $\mathcal{C}$ is said to be **disjoint** if distinct members of $\mathcal{C}$ are disjoint; more precisely, if

$$\forall A, B \in \mathcal{C}, \quad A \cap B \neq \emptyset \quad \Rightarrow \quad A = B.$$

Every subcollection of a disjoint collection of sets is obviously also disjoint.

Let the set S be given. A collection $\mathcal{C}$ of subsets of S is said to **cover** S, and is called a **covering of** S, if $\bigcup \mathcal{C} = S$. A disjoint collection of non-empty subsets of S that covers S is called a **partition of** S. In other words, a collection $\mathcal{P} \in \mathfrak{P}(\mathfrak{P}(S))$ is a partition of S if and only if it satisfies the following conditions:

(Part 1): $\emptyset \notin \mathcal{P}$

(Part 2): $\bigcup \mathcal{P} = S$

(Part 3): $\forall E, F \in \mathcal{P}, \quad E \cap F \neq \emptyset \;\Rightarrow\; E = F.$

(Note that (Part 1) can be omitted if the implication in (Part 3) is replaced by equivalence.)

18A. Examples. (a): Every set S has the **discrete partition** $\{E \in \mathfrak{P}(S) \mid E \text{ is a singleton}\}$. If $S \neq \emptyset$, S also has the **trivial partition** $\{S\}$ (discrete and trivial partitions coincide if and only if S is a singleton). The only partition of the empty set is the empty collection. If A is a subset of S, then $\{A, S\backslash A\}$ is a partition of S if and only if $\emptyset \subsetneqq A \subsetneqq S$.

(b)*: The collections $\{-\mathbb{P}^\times, \mathbb{P}\}, \{-\mathbb{P}^\times, \{0\}, \mathbb{P}^\times\}, \{[n, n+1[\mid n \in \mathbb{Z}\}$ are partitions of $\mathbb{R}$. ■

18B. Proposition. *Let the set S and the partitions $\mathcal{P}$ and $\mathcal{Q}$ of S be given. If $\mathcal{Q} \subset \mathcal{P}$, then $\mathcal{Q} = \mathcal{P}$.*

Proof. We assume that $\mathcal{Q} \subset \mathcal{P}$. Let $E \in \mathcal{P}$ be given. Then $E \neq \emptyset$, and we may choose $a \in E$. Since $\mathcal{Q}$ covers S, we may choose $F \in \mathcal{Q}$ such that $a \in F$. Since $a \in E \cap F$, and $\mathcal{P}$ is a disjoint collection containing E and F, we must have $E = F \in \mathcal{Q}$. Since $E \in \mathcal{P}$ was arbitrary, we conclude that $\mathcal{P} \subset \mathcal{Q}$. ■

Let the set S and the partitions $\mathcal{P}$ and $\mathcal{Q}$ of S be given. $\mathcal{Q}$ is said to be **coarser than** $\mathcal{P}$, and $\mathcal{P}$ is said to be **finer than** $\mathcal{Q}$, and one writes $\mathcal{Q} \sqsubset \mathcal{P}$ or $\mathcal{P} \sqsupset \mathcal{Q}$, if every member of $\mathcal{P}$ is included in some member of $\mathcal{Q}$, i.e., if

$$\forall E \in \mathcal{P}, \; \exists F \in \mathcal{Q}, \; E \subset F.$$

18C. Proposition. *Let the set S and the partitions $\mathcal{P}$ and $\mathcal{Q}$ of S be given. Then $\mathcal{Q}$ is coarser than $\mathcal{P}$ if and only if $\mathcal{P} \cap \mathfrak{P}(F)$ is a partition of F for every $F \in \mathcal{Q}$.*

Proof. Assume first that $\mathcal{Q}$ is coarser than $\mathcal{P}$. Let $F \in \mathcal{Q}$ be given. Since $\mathcal{P}$ is a partition, it is evident that the subcollection $\mathcal{P} \cap \mathfrak{P}(F)$ is disjoint and that $\emptyset \notin \mathcal{P} \cap \mathfrak{P}(F)$. It remains to show that $\mathcal{P} \cap \mathfrak{P}(F)$ covers F. Let $x \in F$ be given. Since $\mathcal{P}$ covers S, we may choose $E \in \mathcal{P}$ such that $x \in E$. Since $\mathcal{Q} \sqsubset \mathcal{P}$, we may further choose $G \in \mathcal{Q}$ such that $E \subset G$. Then $x \in F \cap G$; since $\mathcal{Q}$ is disjoint, we

must have $G = F$, so that $E \subset F$, and hence $E \in \mathcal{P} \cap \mathfrak{P}(F)$ and $x \in \bigcup(\mathcal{P} \cap \mathfrak{P}(F))$. Since $x \in F$ was arbitrary, we conclude that $F \subset \bigcup(\mathcal{P} \cap \mathfrak{P}(F))$. On the other hand, $\bigcup(\mathcal{P} \cap \mathfrak{P}(F)) \subset \bigcup \mathfrak{P}(F) = F$.

To prove the converse implication, assume that $\mathcal{P} \cap \mathfrak{P}(F)$ is a partition of F for every $F \in \mathcal{Q}$. Let $E \in \mathcal{P}$ be given. Since $E \neq \emptyset$, we may choose $a \in E$. Since $\mathcal{Q}$ covers S, we may choose $F \in \mathcal{Q}$ such that $a \in F$. Since $\mathcal{P} \cap \mathfrak{P}(F)$ covers F, we may choose $G \in \mathcal{P} \cap \mathfrak{P}(F)$ such that $a \in G$. Thus $a \in E \cap G$, and since $\mathcal{P}$ is disjoint we must have $E = G \subset F$. Since $E \in \mathcal{P}$ was arbitrary, we conclude that $\mathcal{Q}$ is coarser than $\mathcal{P}$. ■

18D. Proposition. *Let the set S and the partitions $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ of S be given. Then*

$$\mathcal{P} \sqsubset \mathcal{P} \tag{18.1}$$

$$(\mathcal{Q} \sqsubset \mathcal{P} \text{ and } \mathcal{P} \sqsubset \mathcal{Q}) \;\Rightarrow\; \mathcal{P} = \mathcal{Q} \tag{18.2}$$

$$(\mathcal{P} \sqsubset \mathcal{Q} \text{ and } \mathcal{Q} \sqsubset \mathcal{R}) \;\Rightarrow\; \mathcal{P} \sqsubset \mathcal{R}. \tag{18.3}$$

Proof. (18.1) and (18.3) are trivial. To prove (18.2), we assume that $\mathcal{Q} \sqsubset \mathcal{P}$ and $\mathcal{P} \sqsubset \mathcal{Q}$. Let $E \in \mathcal{P}$ be given. Since $\mathcal{Q} \sqsubset \mathcal{P}$, we may choose $F \in \mathcal{Q}$ such that $E \subset F$. Since $\mathcal{P} \sqsubset \mathcal{Q}$, we may further choose $G \in \mathcal{P}$ such that $F \subset G$. Then $E \subset G$, and hence $E \cap G = E \neq \emptyset$. Since $\mathcal{P}$ is disjoint, we must have $E = G$, and hence $E = F \in \mathcal{Q}$. Since $E \in \mathcal{P}$ was arbitrary, this implies $\mathcal{P} \subset \mathcal{Q}$. The reverse inclusion follows by the same argument with $\mathcal{P}$ and $\mathcal{Q}$ interchanged, or by using Proposition 18B. ■

Chapter 2

MAPPINGS

21. The concept of a mapping

The idea of a mapping, along with that of a set, is one of the most basic of all mathematics. Any kind of unambiguous method by which one associates with every object in some set a member of another set (possibly the same set) determines a **mapping**. Thus, in order to specify a mapping f, one first has to prescribe sets D and C, say, and then some kind of definite procedure by which one can assign to every element $x \in D$ an element $f(x) \in C$. We call $f(x)$ the **value of** f **at** x. It is important to distinguish very carefully between the mapping f itself and its values $f(x)$. Thus f is a mapping, while $f(x)$ is a member of C. In older mathematics texts the two are often not sharply distinguished, but such confusion is not permissible in contemporary mathematics. When specifying a mapping, it is also very important to make sure that the procedure under consideration can in fact be applied to *every* member of D.

The set D of objects to which the procedure determining the mapping f can be applied is called the **domain of** the mapping f and is denoted by $\mathrm{Dom}f := D$. The set C to which the values of f must belong is called the **codomain of** f and is denoted by $\mathrm{Cod}f := C$. In order to put D and C into full view, one often writes

$$f\colon D \to C \quad \text{or} \quad D \xrightarrow{f} C$$

and says that f **maps** D **to** C, or that f is a **mapping from** D **to** C. The phrase "f *is defined on* D" expresses the assertion that D is the domain of f.

Suppose that sets D and C are given. Let two procedures be given such that each assigns to every member of D a value in C. We say that both procedures determine *the same* mapping from D to C if to each member of D they both assign the same value. Thus, if f and g are mappings, we have $f = g$ if and only if

$$D := \mathrm{Dom}f = \mathrm{Dom}g \quad \text{and} \quad \mathrm{Cod}f = \mathrm{Cod}g \quad \text{and} \quad f(x) = g(x) \quad \text{for all } x \in D.$$

Terms such as *function, map, functional, transformation,* and *operator* are often used to mean the same thing as *mapping*. The term **function** is preferred when the codomain is a subset of the set of real numbers or of the set of complex numbers. A still greater variety of names is used for mappings having special properties. Moreover, in some contexts, the value of f at x is not written $f(x)$ but, among others, fx, xf, x^f, or f_x. In particular, when the domain of f is a set of pairs, one pair of parentheses is customarily omitted, so that the value of f at the pair (x, y) is then written $f(x, y)$.

In order to specify a mapping f explicitly without introducing unnecessary symbols, it is often useful to employ the notation $x \mapsto f(x)$ instead of just f. (Note that we use $\mapsto$ instead of $\to$ for this purpose.) For example, the "squaring function" $\mathrm{sq} : \mathbb{R} \to \mathbb{R}$, defined by $\mathrm{sq}(x) := x^2$ for all $x \in \mathbb{R}$, may be denoted by $(x \mapsto x^2) : \mathbb{R} \to \mathbb{R}$, and one need not waste a symbol such as "sq" to give it a name. In most contexts it is very important to make a sharp distinction between the number x^2 and the squaring function $x \mapsto x^2$.

Mathematicians have the habit of considering, so soon as a certain kind of object is defined, the set of *all* objects of that kind. Thus, given sets D and C, we can consider the set of *all* mappings from D to C. This set is denoted by $\mathrm{Map}(D, C)$, so that

$$\mathrm{Map}(D, C) :\overset{*}{=} \{f, \text{ a mapping} \mid \mathrm{Dom} f = D, \ \mathrm{Cod} f = C\}.$$

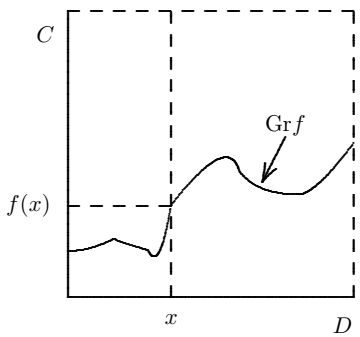
22. The graph of a mapping

Consider a mapping $f\colon D \to C$. We can associate with f a subset $\mathrm{Gr}f$, called the **graph of** f, of the product set $D \times C$ as follows:

$$\mathrm{Gr}f := \{(x, y) \in D \times C \mid y = f(x)\} \subset D \times C. \tag{22.1}$$

Thus, $\mathrm{Gr}f$ consists of all pairs of the form $(z, f(z)), z \in D$.

Not every subset of $D \times C$ is, in general, the graph of a mapping; those subsets that are graphs can be characterized rather neatly. If D and C are intervals and if we represent $D \times C$ as a rectangle in a plane, this characterization will say that the graphs of mappings from D to C are precisely those subsets of $D \times C$ that are cut by every vertical in exactly one point.



22A. Proposition. *Let the sets D and C and the subset G of $D \times C$ be given. If G is the graph of a mapping from D to C, then*

$$\{y \in C \mid (x, y) \in G\} \text{ is a singleton for every } x \in D. \tag{22.2}$$

Conversely, if G satisfies (22.2)*, then G is the graph of exactly one mapping from D to C, namely $f\colon D \to C$ defined by*

$$f(x) :\in \{y \in C \mid (x, y) \in G\} \text{ for every } x \in D. \tag{22.3}$$

Proof. Assume that $G = \mathrm{Gr}f$ for some mapping $f\colon D \to C$, and that $x \in D$ is given. If $y \in C$ is such that $(x, y) \in G = \mathrm{Gr}f$, (22.1) implies that $(x, y) = (z, f(z))$ for a suitable $z \in D$. Therefore $z = x$ and $y = f(z) = f(x)$. Thus $\{y \in C \mid (x, y) \in G\} = \{f(x)\}$ is a singleton. Since $x \in D$ was arbitrary, G satisfies (22.2).

Assume, conversely, that G satisfies (22.2); as we have just seen, if G is to be the graph of $f\colon D \to C$ we must have $f(x) \in \{y \in C \mid (x,y) \in G\}$; this shows that G is the graph of at most one $f\colon D \to C$, namely the one defined by (22.3). But with this choice of f we find that, for all $(x,y) \in D \times C$,

$$(x,y) \in G \quad \Leftrightarrow \quad y \in \{z \in C \mid (x,z) \in G\} \quad \Leftrightarrow \quad y \in \{f(x)\} \quad \Leftrightarrow \quad y = f(x);$$

and (22.1) then shows that indeed $G = \mathrm{Gr} f$. ■

There is thus a one-to-one correspondence between mappings from D to C on the one hand, and certain subsets of $D \times C$ on the other. One can use this correspondence to identify each mapping with its graph. Some mathematicians therefore say that a mapping *is* a set of pairs of objects taken from specified sets, i.e., a subset of a specified product set, and they do not distinguish between a mapping and its graph.

23. The range of a mapping; images and pre-images; the partition of a mapping

It is important to realize that not all members of the codomain C of a mapping $f\colon D \to C$ need be values of f. Those members of C that *are* values of f form a subset of C called the **range of** f and denoted by $\mathrm{Rng}f$. More precisely,

$$\mathrm{Rng}f := \{y \in C \mid y = f(x) \text{ for some } x \in D\} \subset C.$$

In order to test whether a given element y in C belongs to the range of f, one has to search for a $x \in D$ such that $y = f(x)$.

The domain and the codomain are part of the specification of a mapping and must therefore be given when a mapping is given. The range, however, is often unknown beforehand, and it may be quite difficult (and often uninteresting) to determine it. Consider, for example, the function $(x \mapsto x^4 + 6x^3 - 2x) : \mathbb{R} \to \mathbb{R}$. It is easily seen that -1000000 is not a member of the range of this function, and that therefore the range is not equal to the codomain $\mathbb{R}$; it would, however, be a major undertaking to find out what the range of this function is exactly.

Many problems in mathematics consist in asking whether a given equation has solutions. Such problems are called *existence problems*. An equation can be written in the form

$$?x \in D, \quad f(x) = c \tag{23.1}$$

where $f\colon D \to C$ is a suitably defined mapping and c a suitably prescribed member of the codomain C. We read (23.1) as "For which $x \in D$, if any, does $f(x) = c$ hold?". An element $d \in D$ such that $f(d) = c$ is called a *solution of the equation* (23.1). The existence problem for (23.1) is then equivalent to the following question: Is c a member of the range of f? If it is, then (23.1) has solutions; if it is not, then (23.1) has no solutions at all. If the range of f is known, then the existence problem can be solved for every choice of $c \in C$. Thus, to determine the range of f is equivalent to solving the existence problems for the equation (23.1) for all choices of $c \in C$.

A mapping is said to be **surjective**, and is called a **surjection**, if its range happens to coincide with its codomain. If $f\colon D \to C$ is a surjection, i.e., if $\mathrm{Rng}f = C$, we also say that f **maps** D **onto** C (rather than merely *to* C). To say that f is surjective is equivalent to saying that the equation (23.1) always has solutions, no matter how $c \in C$ is chosen.

Whether a mapping is surjective or not depends crucially on the specification of the codomain. For example, the "squaring function" $\mathrm{sq}\colon \mathbb{R} \to \mathbb{R}$ defined in Section 21 is not surjective, because -1 is not a member of the range of sq. However, if we let the codomain be the set $\mathbb{P}$ of positive numbers (i.e., numbers not less than 0), then $\overline{\mathrm{sq}} : \mathbb{R} \to \mathbb{P}$, defined by $\overline{\mathrm{sq}}(x) := x^2$ for all $x \in \mathbb{R}$, is surjective. The functions sq and $\overline{\mathrm{sq}}$ differ only in the specification of the codomain; the domain and the rule that is used to compute their values is the same. We shall discuss surjective mappings and their properties in more detail in Section 32.

Consider again a given mapping $f\colon D \to C$. If U is any subset of D, i.e., $U \in \mathfrak{P}(D)$, we define the **image of** U **under** f to be the set of all values of f at members of the domain that belong to U. We denote this image by $f_>(U)$, so that

$$f_>(U) := \{y \in C \mid y = f(x) \text{ for some } x \in U\} \subset C. \tag{23.2}$$

It is sometimes convenient to use the notation

$$\{f(x) \mid x \in U\} := f_>(U),$$

especially when no explicit name for f is available; this avoids the cumbersome notation $(x \mapsto f(x))_>(U)$. More generally, if $P(\ )$ is a given predicate, we use the notation

$$\{f(x) \mid P(x)\} := f_>(\{x \in \mathrm{Dom}f \mid P(x)\}).$$

Observe that this introduces usages of "set-forming braces" that are different from the one introduced in Section 14, and used, e.g., in (23.2); there will be no confusion, however, although the formula is read in a similar way: "The set of all ... such that ...".

The rule (23.2) defines a new mapping

$$f_> : \mathfrak{P}(D) \to \mathfrak{P}(C),$$

which is called the **image mapping induced by** the mapping f. The value of $f_>$ at $D \in \mathfrak{P}(D)$ is the range of f, i.e., $f_>(D) = \mathrm{Rng}f$. Therefore f is surjective if and only if $f_>(\mathrm{Dom}f) = \mathrm{Cod}f$. Note that the image under f of the empty set is always the empty set again, no matter what f is: $f_>(\emptyset) = \emptyset$. The image of a singleton is a singleton: $f_>(\{x\}) = \{f(x)\}$ for all $x \in D$.

In most contexts, D and $\mathfrak{P}(D)$ have nothing in common. When this is the case, no confusion can arise when one writes $f(U)$ instead of $f_>(U)$ for each $U \in \mathfrak{P}(D)$, and many mathematicians do so. We usually do not.

Let V be any subset of the codomain C of the given mapping $f\colon D \to C$. The **pre-image of** V **under** f is defined to be the set of all members of D at which the values of f belong to V. We denote this pre-image by $f^<(V)$, so that

$$f^<(V) := \{x \in D \mid f(x) \in V\} \subset D. \tag{23.3}$$

The rule (23.3) defines a new mapping,

$$f^< : \mathfrak{P}(C) \to \mathfrak{P}(D),$$

which is called the **pre-image mapping induced by** the mapping f. Note that the pre-image under f of the empty set is always the empty set, while the domain of f is the pre-image both of the codomain and of the range of f: thus $f^<(\emptyset) = \emptyset$, $f^<(C) = f^<(\mathrm{Rng}f) = D$.

Many mathematicians use the symbol f^{-1} instead of $f^<$ for the induced pre-image mapping. We shall never do so, because of the danger of confusion with various other objects.

Applying the preceding definitions to the mappings $f_>$ and $f^<$ instead of f itself, we can construct the mappings $(f_>)_>$ and $(f^<)^<$ from $\mathfrak{P}(\mathfrak{P}(D))$ to $\mathfrak{P}(\mathfrak{P}(C))$ and the mappings $(f_>)^<$ and $(f^<)_>$ from $\mathfrak{P}(\mathfrak{P}(C))$ to $\mathfrak{P}(\mathfrak{P}(D))$. In the following proposition we collect some elementary rules satisfied by the image and pre-image mappings induced by a given mapping. The proofs are left to the reader.

23A. Proposition. *The mappings $f_>$ and $f^<$ induced by a given mapping $f\colon D \to C$ satisfy the following rules for all subsets U, U', U'' of D, all subsets V, V', V'' of C, all subcollections $\mathcal{U}$ of $\mathfrak{P}(D)$, and all subcollections $\mathcal{V}$ of $\mathfrak{P}(C)$:*

$$U' \subset U'' \;\Rightarrow\; f_>(U') \subset f_>(U'') \tag{23.4}$$

$$V' \subset V'' \;\Rightarrow\; f^<(V') \subset f^<(V'') \tag{23.5}$$

$$f_>(U) \subset V \quad\Leftrightarrow\quad U \subset f^<(V) \tag{23.6}$$

$$f^<(f_>(U)) \supset U \tag{23.7}$$

$$f_>(f^<(V)) = V \cap \mathrm{Rng} f \subset V \tag{23.8}$$

$$f_>\Big(\bigcup \mathcal{U}\Big) = \bigcup (f_>)_>(\mathcal{U}) = \bigcup \{f_>(S) \mid S \in \mathcal{U}\}$$
$$\text{and in particular } f_>(U' \cup U'') = f_>(U') \cup f_>(U'') \tag{23.9}$$

$$f_>\Big(\bigcap{}^D \mathcal{U}\Big) \subset \bigcap{}^C (f_>)_>(\mathcal{U}) = \bigcap{}^C \{f_>(S) \mid S \in \mathcal{U}\}$$
$$\text{and in particular } f_>(U' \cap U'') \subset f_>(U') \cap f_>(U'') \tag{23.10}$$

$$f_>(U \cap f^<(V)) = f_>(U) \cap f_>(f^<(V)) = f_>(U) \cap V \tag{23.11}$$

$$f^<\Big(\bigcup \mathcal{V}\Big) = \bigcup (f^<)_>(\mathcal{V}) = \bigcup \{f^<(T) \mid T \in \mathcal{V}\}$$
$$\text{and in particular } f^<(V' \cup V'') = f^<(V') \cup f^<(V'') \tag{23.12}$$

$$f^<\Big(\bigcap{}^C \mathcal{V}\Big) = \bigcap{}^D (f^<)_>(\mathcal{V}) = \bigcap{}^D \{f^<(T) \mid T \in \mathcal{V}\}$$
$$\text{and in particular } f^<(V' \cap V'') = f^<(V') \cap f^<(V'') \tag{23.13}$$

$$f^<(C \backslash V) = D \backslash f^<(V). \tag{23.14}$$

Let a mapping $f\colon D \to C$ again be given. The pre-images under f of singletons are particularly interesting subsets of the domain D. We note that for every $y \in C$ we

have $f^<(\{y\}) \neq \emptyset$ if and only if $y \in \mathrm{Rng} f$. These non-empty pre-images of singletons actually constitute a partition of D, called the **partition of** f; it is denoted by $\mathrm{Part} f$, and is given by

$$\mathrm{Part} f := \{f^<(\{y\}) \mid y \in \mathrm{Rng} f\} \subset \mathfrak{P}(D).$$

We verify that this is indeed a partition of D; its members are non-empty; for every $x \in D$ we have $f(x) \in \mathrm{Rng} f$ and hence $x \in f^<(\{f(x)\}) \in \mathrm{Part} f$, so that $D = \bigcup \mathrm{Part} f$; and for all $y, z \in \mathrm{Rng} f$, (23.13) shows that $f^<(\{y\}) \cap f^<(\{z\}) = f^<(\{y\} \cap \{z\})$, so that the former intersection is not empty if and only if $y = z$ and hence $f^<(\{y\}) = f^<(\{z\})$.

24. Inclusion, identity, and partition mappings

We introduce some simple mappings that are helpful in keeping accounts of sets and mappings straight.

Let S be a set, and U a subset of S. We define the mapping $1_{U\subset S} : U \to S$ by the rule

$$1_{U\subset S}(x) := x \quad \text{for all } x \in U;$$

this mapping is called the **inclusion mapping of** U **into** S. It is obvious that $\mathrm{Rng} 1_{U\subset S} = U$. We note the formulas

$$(1_{U\subset S})_{>} = 1_{\mathfrak{P}(U)\subset\mathfrak{P}(S)}, \quad (1_{U\subset S})^{<}(T) = T \cap U \quad \text{for all } T \in \mathfrak{P}(S).$$

It may be useful to point out that the definition of $1_{U\subset S}$ remains meaningful (although vacuously so) if $U = \emptyset$. Indeed, $1_{\emptyset\subset S}$ is the *only* mapping from $\emptyset$ to S (cf. Section 21 or Proposition 22A). On the other hand, there is obviously no mapping from a non-empty set to the empty set.

For each set S we have the mapping $1_S := 1_{S\subset S}$; it is called the **identity mapping of** S. The mapping 1_S is obviously surjective.

Let $\mathcal{P}$ be a partition of a set S. For every $x \in S$, the collection $\{E \in \mathcal{P} \mid x \in E\}$ is a singleton: it is not empty, since $x \in S = \bigcup\mathcal{P}$ and if $x \in E \in \mathcal{P}$ and $x \in F \in \mathcal{P}$, then $E \cap F \neq \emptyset$ and therefore $E = F$. We may therefore define a mapping $\Omega_{\mathcal{P}} : S \to \mathcal{P}$ by the rule

$$\Omega_{\mathcal{P}}(x) :\in \{E \in \mathcal{P} \mid x \in E\} \quad \text{for all } x \in S.$$

This mapping is called the **partition mapping of** $\mathcal{P}$. We find that $x \in \Omega_{\mathcal{P}}(x)$ for all $x \in S$, and $\Omega_{\mathcal{P}}^{<}(\{E\}) = E$ for every $E \in \mathcal{P}$, so that $\mathrm{Rng}\Omega_{\mathcal{P}} = \mathrm{Part}\Omega_{\mathcal{P}} = \mathcal{P}$. We further note that $\Omega_{\mathcal{P}}^{<}(\mathcal{A}) = \bigcup\mathcal{A}$ for every subcollection $\mathcal{A}$ of $\mathcal{P}$.

For every mapping $f\colon D \to C$, we have $x \in f^{<}(\{f(x)\})$, and therefore

$$\Omega_{\mathrm{Part}f}(x) = f^{<}(\{f(x)\}) \quad \text{for all } x \in D.$$

25. Composition of mappings; diagrams; restrictions and adjustments

Assume that a mapping $f\colon D \to C$ and a mapping $g\colon C \to B$ are given, so that the codomain of the former is the domain of the latter. A new mapping $g \circ f\colon D \to B$ is then defined by the rule

$$(g \circ f)(x) \;:=\; g(f(x)) \quad \text{for all } x \in D.$$

This mapping is called the **composite of f and g**, and is said to be obtained by **composition of f with g**. We read $g \circ f$ as "f **composed with** g" (note the reversal of priority). Thus the values of $g \circ f$ are obtained by first operating with f and then on the resulting values with g. The equality $\mathrm{Cod}f = \mathrm{Dom}g$ is essential to the definition of $g \circ f$; if this condition is satisfied, then $\mathrm{Dom}(g \circ f) = \mathrm{Dom}f$ and $\mathrm{Cod}(g \circ f) = \mathrm{Cod}g$.

Composition of mappings obeys the associative law in the following sense: If $f,\ g,\ h$ are mappings with $\mathrm{Dom}g = \mathrm{Cod}f$ and $\mathrm{Dom}h = \mathrm{Cod}g$, then

$$(h \circ g) \circ f = h \circ (g \circ f).$$

This rule is an almost immediate consequence of the definition of composition. Because of this rule, we may unambiguously write $h \circ g \circ f$ without parentheses; and a similar license is in effect for denoting composites of more mappings, provided that the domains and codomains are appropriately matched.

We note that $f \circ 1_{\mathrm{Dom}f} = 1_{\mathrm{Cod}f} \circ f = f$ for every mapping f.

Suppose that $f,\ g$ are mappings such that $\mathrm{Dom}g = \mathrm{Cod}f$. Then $\mathrm{Dom}(g_>) = \mathfrak{P}(\mathrm{Dom}g) = \mathfrak{P}(\mathrm{Cod}f) = \mathrm{Cod}(f_>)$ and similarly $\mathrm{Dom}(f^<) = \mathrm{Cod}(g^<)$, and the following rules hold:

$$(g \circ f)_> = g_> \circ f_> \tag{25.1}$$

$$(g \circ f)^< = f^< \circ g^< \tag{25.2}$$

$$\mathrm{Rng}(g \circ f) = g_>(\mathrm{Rng}f) \subset \mathrm{Rng}g \tag{25.3}$$

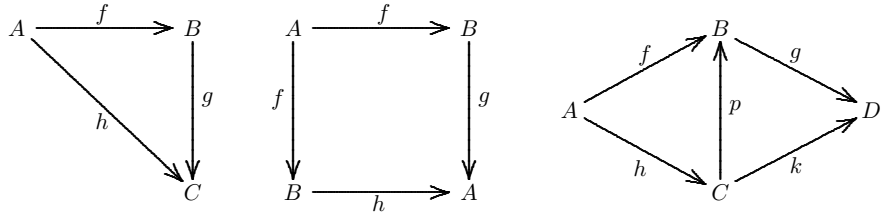
$$\mathrm{Part}(g \circ f) \sqsubset \mathrm{Part}f. \tag{25.4}$$

All these rules, except perhaps (25.4), follow trivially from the definitions. To prove (25.4), we let $y \in \mathrm{Rng}f$ be given. Then $g(y) \in \mathrm{Rng}(g \circ f)$, and $\{y\} \subset g^<(\{g(y)\})$, so that, using (23.5) and (25.2),

$$f^<(\{y\}) \subset f^<(g^<(\{g(y)\}) = (g \circ f)^<(\{g(y)\}) \in \mathrm{Part}(g \circ f).$$

Since $y \in \mathrm{Rng}f$ was arbitrary, the definition of $\mathrm{Part}f$ in Section 23 shows that (25.4) holds.

A complicated situation involving several mappings can often be made clearer by means of an informal graphic device called a *diagram*. Simple examples of diagrams are



A diagram consists of *places*, each labelled with (the name of) a set, and *arrows*, each labelled with (the name of) a mapping. The sets in the places at the head and the tail of an arrow labelled f are the domain and the codomain of f, respectively. In our first example the mappings are $f\colon A \to B$, $g\colon B \to C$, $h\colon A \to C$. Our second example illustrates the fact that the same set may appear in more than one place, and the same mapping on more than one arrow. Mappings that are labels of consecutive (i.e., head-to-tail) arrows can always be composed.

If, in our first example, $h = g \circ f$, this diagram is said to be commutative; the second diagram is said to be commutative if $g \circ f = h \circ f$; the third, if $f = p \circ h$ and $k = g \circ p$ (it then follows that $g \circ f = k \circ h$, etc.). More generally, a diagram is said to be *commutative* if, for any two chains of consecutive arrows that go from the same initial place to the same final place, the corresponding composites of mappings are equal. Commutativity of a diagram is often indicated by the symbol ⁜.

Let the mapping f and the sets A and B be given. If $A \subset \mathrm{Dom} f$, we define a new mapping $f|_A : A \to \mathrm{Cod} f$ by $f|_A(x) := f(x)$ for all $x \in A$. We call $f|_A$ the **restriction of** f **to** A. More generally, if $A \subset \mathrm{Dom} f$ and $B \supset f_>(A)$, we define $f|_A^B : A \to B$ by $f|_A^B(x) := f(x)$ for all $x \in A$. We say that $f|_A^B$ is obtained from f by **adjustment**; and we say that f **induces** a given mapping g if g is obtained from f by adjustment. Of course $f|_A^{\mathrm{Cod} f} = f_A$. For arbitrary B we note that, by (23.8), $f_>(f^<(B \cap \mathrm{Cod} f)) = f_>(f^<(B \cap \mathrm{Rng} f)) = B \cap \mathrm{Rng} f \supset B$, and we define $f|^B := f|^B_{f^<(B \cap \mathrm{Cod} f)} = f|^B_{f^<(B \cap \mathrm{Rng} f)}$. We have $\mathrm{Dom}(f|^B) = \mathrm{Dom} f$ if and only if $B \supset \mathrm{Rng} f$. We say that $f|^B$ is obtained from f by **adjusting the codomain to** B. The **surjective reduction** $f|^{\mathrm{Rng}} : \mathrm{Dom} f \to \mathrm{Rng} f$ **of** f is defined by adjusting the codomain to the range, i.e., by $f|^{\mathrm{Rng}} := f|^{\mathrm{Rng} f}$. We note that, if $A \subset \mathrm{Dom} f$ and $B \supset \mathrm{Rng} f$, then

$$f|_A = f \circ 1_{A \subset \mathrm{Dom} f} \qquad f|^B = 1_{\mathrm{Rng} f \subset B} \circ f|^{\mathrm{Rng}}.$$

Let the mappings $f\colon D \to C$ and $f'\colon D' \to C'$ be given. For every subset S of $D \cap D'$, f and f' are said to **agree on** S, and f is said to **agree with** f' **on** S, if $f(s) = f'(s)$ for all $s \in S$, i.e., if $f|_S^{C \cup C'} = f'|_S^{C \cup C'}$.

Let the sets A and B be given. For every $a \in A$ and $b \in B$ we define the mappings

$(\cdot, b) : A \to A \times B$ and $(a, \cdot) : B \to A \times B$ by the rules

$$(\cdot, b)(x) := (x, b) \text{ for all } x \in A \qquad (a, \cdot)(y) := (a, y) \text{ for all } y \in B.$$

These notations are used in contexts in which the sets A and B are understood, and most frequently in conjunction with a mapping f with $\mathrm{Dom} f \subset A \times B$. It is then customary to write $f(\cdot, b) := f \circ (\cdot, b)|^{\mathrm{Dom} f}$ and $f(a, \cdot) := f \circ (a, \cdot)|^{\mathrm{Dom} f}$, so that

$$(f(\cdot, b))(a) = f(a, b) = (f(a, \cdot))(b) \text{ for all} (a, b) \in \mathrm{Dom} f. \tag{25.5}$$

For every $D \in \mathfrak{P}(A \times B)$ we define

$$D^\top := \{(y, x) | (x, y) \in D\} \in \mathfrak{P}(B \times A)$$

and the mapping $\top_D : D \to D^\top$ by

$$\top_D(x, y) := (y, x) \quad \text{for all } (x, y) \in D.$$

We note that $(A \times B)^\top = B \times A$, $D^\top = (\top_{A \times B})_>(D)$, $\top_D = (\top_{A \times B})|_D^{D^\top}$, $D^{\top\top} = D$, and $\top_{D^\top} \circ \top_D = 1_D$.

For every mapping f with $\mathrm{Dom}\ f = D \in \mathfrak{P}(A \times B)$ we define the mapping $f^\top$ with domain $D^\top$ and $\mathrm{Cod} f^\top := \mathrm{Cod} f$ by $f^\top := f \circ \top_{D^\top}$, so that

$$f^\top(y, x) = f(x, y) \quad \text{for all } (y, x) \in D^\top.$$

26. Mappings from a set to itself

There are some useful notions that pertain only to mappings from a set to itself. Let the set D be given. The mapping

$$((f,g)\mapsto g\circ f) : \mathrm{Map}(D,D)\times\mathrm{Map}(D,D)\to\mathrm{Map}(D,D),$$

called **composition in** D, of course satisfies the associative law

$$(h\circ g)\circ f = h\circ(g\circ f) \quad \text{for all} \quad f,g,h\in\mathrm{Map}(D,D)$$

as well as

$$f\circ 1_D = f = 1_D\circ f \quad \text{for all } f\in\mathrm{Map}(D,D),$$

but in general not the commutative law: $g\circ f\neq f\circ g$ for some $f,g\in\mathrm{Map}(D,D)$. (The only exceptions occur when D is empty or a singleton.) For given $f,g\in\mathrm{Map}(D,D)$, we say that f and g **commute**, or that f **commutes with** g, if $g\circ f = f\circ g$. A subset F of $\mathrm{Map}(D,D)$ is said to be **commutative** if f and g commute for all $f,g\in F$.

A mapping $f : D\to D$ is said to be **idempotent** if $f\circ f = f$; it is said to be **involutory**, and is called an **involution**, if $f\circ f = 1_D$.

Many questions in mathematics can be reduced to solving a problem of the form

$$?x\in D, \quad f(x)=x$$

where $f : D\to D$ is a suitable mapping. A solution to this problem, i.e., a member d of D such that $f(d)=d$, is called a **fixed point of** f. We shall encounter later several important cases in which the existence of a fixed point can be established. We define

$$\mathrm{Fix}f := \{x\in D \mid f(x)=x\},$$

the set of all fixed points of f.

Let the set D and the mapping $f\colon D\to D$ be given. A subset S of D is said to be **stable under** f, or f-**stable** for short, if $f_>(S)\subset S$. We note that $d\in D$ is a fixed point of f if and only if $\{d\}$ is stable under f; this observation is used to derive Part (b) from Part (a) in the following proposition.

26A. Proposition. *Let the set D and the mappings $f,g\in\mathrm{Map}(D,D)$ be given, and assume that f and g commute.*

(a): *If the subset S of D is stable under f, then $g_>(S)$ is also stable under f.*

(b): *If $x\in D$ is a fixed point of f, then $g(x)$ is also a fixed point of f. Consequently, if $x\in D$ is the only fixed point of f, then x is also a fixed point of g.*

26B. Example. Let the set D and the mapping $f\colon D\to D$ be given. Every fixed point of f is plainly also a fixed point of $f\circ f\circ f\circ f$. Assume, on the other hand, that $x\in D$ is the only fixed point of $f\circ f\circ f\circ f$. Since $f\circ f\circ f\circ f$ and f commute, Proposition 26A shows that x is also a fixed point of f; and it is then the only one. (The same argument is valid with $f\circ f$ or $f\circ f\circ f$ instead of $f\circ f\circ f\circ f$.) ■

26C. Proposition. *Let the set D and the mapping $f\colon D \to D$ be given. Then f is idempotent if and only if the range of f is the set of fixed points of f, i.e.,*

$$\mathrm{Rng} f = \mathrm{Fix} f.$$

Proof. Obviously, $\mathrm{Fix} f \subset \mathrm{Rng} f$.

Assume first that $\mathrm{Rng} f \subset \mathrm{Fix} f$, and let $z \in D$ be given. Then $f(z) \in \mathrm{Rng} f$, and hence $f(f(z)) = f(z)$. Since $z \in D$ was arbitrary, we conclude that f is indempotent.

Assume, conversely, that f is idempotent, and let $y \in \mathrm{Rng} f$ be given. We may choose $z \in D$ such that $f(z) = y$, and find $f(y) = f(f(z)) = f(z) = y$. Since $y \in \mathrm{Rng} f$ was arbitrary, we have $\mathrm{Rng} f \subset \mathrm{Fix} f$. ■

Chapter 3

PROPERTIES OF MAPPINGS

31. Constants

Mappings of a very simple kind are the constants, that is, mappings that have the same value at all members of the domain. More formally, a mapping $f\colon D \to C$ is said to be **constant**, and is called a **constant (mapping)**, if

(Const): $\forall x, x' \in D, \quad f(x) = f(x')$.

Obvious equivalent variants of this definiens are:

(Const$_1$): Part f *is* $\{D\}$ *(when* $D \neq \emptyset$*) or* $\emptyset$ *(when* $D = \emptyset$*).*

(Const$_2$): Rng$f$ *is a singleton (when* $D \neq \emptyset$*) or* $\emptyset$ *(when* $D = \emptyset$*).*

Given the sets D and C and $c \in C$, the mapping $f\colon D \to C$ defined by the rule $f(x) := c$ for all $x \in D$ is a constant, since $f^<(\{c\}) = D$; it is denoted by $c_{D\to C}$. Every constant other than $1_\emptyset$ is of this form. In many contexts it would be pedantic to use different symbols for a constant and its only value; but the conceptual distinction must always be kept in mind; thus $c_{D\to C}$ is often abbreviated to c if confusion is unlikely. We make a trivial remark: If $D \neq \emptyset$ and $f\colon D \to C$ is constant, then $f = f(d)_{D\to C}$ for each $d \in D$; and if $C \neq \emptyset$, then $1_{\emptyset\subset C} = c_{\emptyset\to C}$ for every $c \in C$. A consequence of this remark is that, for all sets D and C, $\mathrm{Map}(D, C) = \emptyset$ if and only if $D \neq \emptyset$ but $C = \emptyset$.

31A. Proposition. *Let mappings f and g be given with* Dom$g$ = Cod$f$. *Then $g \circ f$ is constant if f is constant or if g is constant.*

31B. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(Const): *f is constant.*

(Const$_3$): *There are mappings $g\colon D \to S$ and $h\colon S \to C$ such that $f = h \circ g$ and S is empty or a singleton.*

32. Injective, surjective, and bijective mappings

Consider a given mapping $f\colon D \to C$. In Section 23, we discussed the notion of *solution* of equation (23.1), i.e.,

$$?x \in D, \qquad f(x) = c \tag{32.1}$$

where $c \in C$ is given. We observe that the set of all solutions of (32.1) is precisely $f^<(\{c\})$. If equation (32.1) has solutions, i.e., if $f^<(\{c\}) \neq \emptyset$, or, equivalently $c \in \mathrm{Rng} f$, it is usual to say that (32.1) **has at least one solution**; the problem of ascertaining whether this is the case was called an *existence problem* in Section 23. If, on the other hand, the set $f^<(\{c\})$ of solutions of (32.1) is either empty or a singleton, it is usual to say that (32.1) **has at most one solution**; and the problem of ascertaining whether this is the case is called a *uniqueness problem*. If the set $f^<(\{c\})$ of solutions of (32.1) is actually a singleton, one says that (32.1) **has exactly one solution**, or, less aptly, that (32.1) *has a unique solution*.

In this section we discuss properties of a mapping f that ensure an affirmative answer to the existence problem, to the uniqueness problem, or to both, for equation (32.1) no matter how $c \in C$ is chosen. One of these properties, namely surjectivity, was already introduced in Section 23.

In this and subsequent sections, many propositions come in pairs, or even triples, of analogues; the propositions in the same pair or triple are given the same designation with a distinguishing addition, namely L ("left"), R ("right"), and, if necessary, B ("bilateral").

A mapping $f\colon D \to C$ is said to be **injective**, and is called an **injection**, if it satisfies the following condition:

(Inj): $\forall x, x' \in D,\ f(x) = f(x') \ \Rightarrow\ x = x'$.

Obvious equivalent variants of this definiens are:

(Inj$_1$): $\forall y \in C,\ f^<(\{y\})$ *is empty or a singleton.*

(Inj$_2$): *For every* $c \in C$, *the equation* $?x \in D,\ f(x) = c$ *has at most one solution.*

(Inj$_3$): $\mathrm{Part} f = \{\{x\} \mid x \in D\}$, *the discrete partition of* D.

A mapping $f\colon D \to C$ is said to be **surjective**, and is called a **surjection** (cf. Section 23), if it satisfies the following condition:

(Surj): $\forall y \in C, \exists x \in D,\ f(x) = y$.

Obvious equivalent variants of this definiens are:

(Surj$_1$): $\forall y \in C,\ f^<(\{y\}) \neq \emptyset$.

(Surj$_2$): *For every* $c \in C$, *the equation* $?x \in D, f(x) = c$ *has at least one solution.*

(Surj$_3$): $\mathrm{Rng} f = C$.

A mapping $f\colon D \to C$ is said to be **bijective**, and is called a **bijection**, if it satisfies the following condition:

(Bij): f *is both injective and surjective.*

Obvious equivalent variants of this definiens are:

(Bij$_1$): $\forall y \in C,\ f^<(\{y\})$ *is a singleton.*

(Bij$_2$): *For every* $c \in C$, *the equation* $?x \in D, f(x) = c$ *has exactly one solution.*

(Bij$_3$): $\mathrm{Part}f = \{\{x\} \mid x \in D\}$ *and* $\mathrm{Rng}f = C$.

We note that every inclusion mapping is injective, every partition mapping is surjective (and bijective if and only if the partition is the discrete partition of the domain), and every identity mapping is bijective.

32A. Proposition. *For every pair* (S,T) *of sets, the mapping* $((x,y) \mapsto y) : S \times T \to T$ *is surjective unless* $S = \emptyset$ *and* $T \neq \emptyset$.

32B.L. Proposition. *Let the mappings* $f\colon D \to C$ and $g\colon C \to B$ *be given.*

(a): *If* f *and* g *are injective, then* $g \circ f$ *is injective.*

(b): *If* $g \circ f$ *is injective, then* f *is injective; if, moreover,* f *is surjective, then* g *is also injective.*

(c): *If* $g \circ f$ *is constant and* g *is injective, then* f *is constant.*

32B.R. Proposition. *Let the mappings* $f\colon D \to C$ *and* $g\colon C \to B$ *be given.*

(a): *If* f *and* g *are surjective, then* $g \circ f$ *is surjective.*

(b): *If* $g \circ f$ *is surjective, then* g *is surjective; if, moreover,* g *is injective, then* f *is also surjective.*

(c): *If* $g \circ f$ *is constant and* f *is surjective, then* g *is constant.*

32C.L. Proposition. *Let the mappings* $f\colon D \to C$ *and* $g\colon C \to B$ *be given. If* g *is injective, then* $\mathrm{Part}(g \circ f) = \mathrm{Part}f$.

Proof. For every $z \in \mathrm{Rng}(g \circ f) \subset \mathrm{Rng}g$, the set $g^<(\{z\})$ is a singleton, hence $(g \circ f)^<(\{z\}) = f^<(g^<(\{z\})) \in \mathrm{Part}f$. We have shown that $\mathrm{Part}(g \circ f) \subset \mathrm{Part}f$. Since both collections are partitions of D, they are equal (Proposition 18B). ■

32C.R. Proposition. *Let the mappings* $f\colon D \to C$ *and* $g\colon C \to B$ *be given. If* f *is surjective, then* $\mathrm{Rng}(g \circ f) = \mathrm{Rng}g$.

Proof. For every $y \in C$, we have $y \in f_>(f^<(\{y\}))$; hence $g(y) \in g_>(f_>(f^<(\{y\}))) = (g \circ f)_>(f^<(\{y\})) \subset \mathrm{Rng}(g \circ f)$. Since $y \in C$ was arbitrary, we conclude that $\mathrm{Rng}g \subset \mathrm{Rng}(g \circ f)$. The reverse inclusion is valid by (25.3). ■

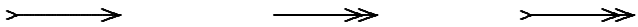
32D. Proposition. *Given a mapping* $f\colon D \to C$, *there are mappings* $g\colon D \to S$ *and* $h\colon S \to C$ *such that* $f = h \circ g$ *and* g *is surjective and* h *is injective.*

Proof. Set $S := \mathrm{Rng}f,\ g := f|^{\mathrm{Rng}},\ h := 1_{\mathrm{Rng}f \subset C}$. ■

32E. Proposition. *A mapping from a set to its power-set cannot be surjective.*

Proof. Let the set D and the mapping $F\colon D \to \mathfrak{P}(D)$ be given. Consider the set $K := \{x \in D \mid x \notin F(x)\}$. Let $z \in D$ be given. If $z \in K$, then $z \notin F(z)$, and therefore $F(z) \neq K$; if $z \notin K$, then $z \in F(z)$, and therefore $F(z) \neq K$. We conclude that $K \notin \mathrm{Rng}F$, so that F is not surjective. ■

In drawing a diagram (see Section 25) it is often convenient to record that a specific mapping is injective, surjective, or bijective. This may be done by marking the arrows labelled with that mapping as follows:



(injective) (surjective) (bijective)

When the mapping is an inclusion mapping, it is customary to omit the label and use a "hooked arrow": $\hookrightarrow$.

33. Inverses and invertibility

Let the mapping $f\colon D \to C$ be given. A mapping $g\colon C \to D$ is called a **left-inverse of** f if $g \circ f = 1_D$, a **right-inverse of** f if $f \circ g = 1_C$, and an **inverse of** f if g is both a left-inverse and a right-inverse of f. We observe that g is a left-inverse of f if and only if f is a right-inverse of g. We note that g is an inverse of f if and only if

$$\forall (x,y) \in D \times C, \qquad y = f(x) \qquad \Leftrightarrow \qquad x = g(y). \tag{33.1}$$

33A. Proposition. (a): *If $f\colon D \to C$ and $g\colon C \to B$ are given mappings, and h and k are left-inverses of f and g, respectively, then $h \circ k$ is a left-inverse of $g \circ f$. The assertion remains valid if "left-inverse" is replaced everywhere by "right-inverse" or by "inverse".*

(b): *If f is a given mapping, and g is a left-inverse of f and h is a right-inverse of f, then $g = h$, and g is an inverse of f.*

Proof of (b). $g = g \circ 1_C = g \circ (f \circ h) = (g \circ f) \circ h = 1_D \circ h = h$. ■

A mapping f is said to be **left-invertible** if there is a left-inverse of f, and **right-invertible** if there is a right-inverse of f; f is said to be **invertible** (and is called a **set-isomorphism**) if there is an inverse of f.

33B. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(LRInv): *f is both left-invertible and right-invertible.*

(Inv): *f is invertible.*

(UInv): *There is a mapping $g\colon C \to D$ that is the unique left-inverse of f and the unique right-inverse of f, as well as the unique inverse of f.*

Proof. Proposition 33A,(b). ■

It follows from Proposition 33B that if f is invertible there is exactly one inverse of f; this unique inverse will be denoted by $f^{\leftarrow}$. (Although a more usual notation is f^{-1} we shall avoid it, except in some special cases in which it will be particularly appropriate.) It is important to distinguish carefully the inverse $f^{\leftarrow} : C \to D$ from the pre-image mapping $f^{<} : \mathfrak{P}(C) \to \mathfrak{P}(D)$; the latter is defined for every mapping f, but the former only when f is invertible.

33C. Proposition. (a): *If f is an invertible mapping, then $f^{\leftarrow}$ is invertible, and $f^{\leftarrow\leftarrow} = f$.*

(b): *If f is an invertible mapping, then $(f^{\leftarrow})_{>} = f^{<}$ and $(f^{\leftarrow})^{<} = f_{>}$.*

(c): *If f and g are invertible mappings and* $\mathrm{Dom} g = \mathrm{Cod} f$*, then $g \circ f$ is invertible, and $(g \circ f)^{\leftarrow} = f^{\leftarrow} \circ g^{\leftarrow}$.*

(d): *If f and g satisfy* $\mathrm{Dom} g = \mathrm{Cod} f$*, and $g \circ f$ is invertible, then f is invertible if and only if g is invertible.*

Proof of (d). Suppose that f is invertible, and set $h := f \circ (g \circ f)^{\leftarrow}$. Then $g \circ h = (g \circ f) \circ (g \circ f)^{\leftarrow} = 1_{\mathrm{Cod} g}$ and $h \circ g = h \circ g \circ f \circ f^{\leftarrow} = f \circ (g \circ f)^{\leftarrow} \circ (g \circ f) \circ f^{\leftarrow} = f \circ 1_{\mathrm{Dom} f} \circ f^{\leftarrow} = 1_{\mathrm{Cod} f} = 1_{\mathrm{Dom} g}$. Thus g is invertible, and $g^{\leftarrow} = h$. The converse implication is proved in an analogous manner. ■

▼ **33D. Proposition**. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(Inv): *f is invertible.*

(ULInv^1): *There is exactly one left-inverse of f, and if D is a singleton then C is a singleton.*

(URInv): *There is exactly one right-inverse of f.*

Proof. (Inv) $\Rightarrow$ (ULInv^1). By Proposition 33B, (Inv) $\Rightarrow$ (UInv). If f is invertible, it therefore has exactly one left-inverse, namely $f^{\leftarrow}$. If $D = \{d\}$, then $f(d) \in C$, and for each $y \in C$, $y = f(f^{\leftarrow}(y)) = f(d)$, so that $C = \{f(d)\}$.

(Inv) $\Rightarrow$ (URInv). By Proposition 33B, (Inv) $\Rightarrow$ (UInv). If f is invertible, it therefore has exactly one right-inverse, namely $f^{\leftarrow}$.

(ULInv^1) $\Rightarrow$ (Inv). Suppose that f satisfies (ULInv^1), and let g be the unique left-inverse of f. For each $d \in D$ we define $g_d\colon C \to D$ by the rule

$$g_d(y) := \begin{cases} g(y) & \text{if } f(g(y)) = y \\ \\ d & \text{if } f(g(y)) \neq y \end{cases} \qquad y \in C.$$

For each $x \in D$ we have $f(g(f(x))) = f(x)$, and therefore $g_d(f(x)) = x$; hence $g_d \circ f = 1_D$; but g is the unique left-inverse of f, so that $g_d = g$.

Now let $y \in C$ be given, and suppose that $f(g(y)) \neq y$; in particular, C is then not a singleton. For each $d \in D$ we have $d = g_d(y) = g(y)$, so that $D = \{g(y)\}$ is a singleton, contradicting the assumption. Consequently, $f(g(y)) = y$ for all $y \in C$, so that $f \circ g = 1_C$. Hence g is also a right-inverse of f, and hence an inverse of f.

(URInv) $\Rightarrow$ (Inv). Suppose that f satisfies (URInv), and let g be the unique right-inverse of f. For each $x \in D$, we define $g_x : C \to D$ by the rule

$$g_x(y) := \begin{cases} g(y) & \text{if } y \neq f(x) \\ \\ x & \text{if } y = f(x) \end{cases} \qquad \text{for all } y \ \in C.$$

For each $y \in C$ we then have $f(g_x(y)) = f(g(y)) = y$ if $y \neq f(x)$, and $f(g_x(y)) = f(x) = y$ if $y = f(x)$, so that $f \circ g_x = 1_C$. But g is the unique right-inverse of f, so that $g_x = g$; and therefore $g(f(x)) = g_x(f(x)) = x$. Since this holds for all $x \in D$, we have $g \circ f = 1_D$. Thus g is also a left-inverse of f, and hence an inverse of f. ■

Remark. If $D := \{d\}$ is a singleton and C is neither empty nor a singleton, each mapping $f\colon D \to C$ satisfies $f = f(d)_{D\to C}$ and has the unique left-inverse $g := d_{C\to D}$;
▲ but none is invertible, since $f \circ g = f(d)_{C\to C} \neq 1_C$.

There is a close connection, which we now explore, between the various invertibility properties of a mapping on the one hand, and the properties of injectivity, surjectivity, and bijectivity of the mapping on the other.

33E.L. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent*:

(LInv): *f is left-invertible.*

(Inj^0): *f is injective, and if $D = \emptyset$ then $C = \emptyset$.*

Proof. (LInv) $\Rightarrow$ (Inj^0). Choose a left-inverse g of f; then $g \circ f = 1_D$, which is bijective. Hence f is injective (Proposition 32.B.L, (b)). The set $\mathrm{Map}(C, D)$ is not empty, since it contains g; therefore $D \neq \emptyset$ or $C = \emptyset$.

(Inj^0) $\Rightarrow$ (LInv). Assume that f satisfies (Inj^0). If $D = C = \emptyset$, then $f = 1_\emptyset$, and $1_\emptyset$ is a left-inverse of f. Assume now that $D \neq \emptyset$, and choose $d \in D$. For each $y \in C$, the set $f^<(\{y\})$ is a singleton or empty, according as $y \in \mathrm{Rng} f$ or $y \in C \backslash \mathrm{Rng} f$. Define $g\colon C \to D$ by the following rule: $g(y) :\in f^<(\{y\})$ if $y \in \mathrm{Rng} f$, and $g(y) := d$ if $y \in C \backslash \mathrm{Rng} f$. For every $x \in D$ we then find $x \in f^<(\{f(x)\}) = \{g(f(x))\}$, so that $x = g(f(x))$. Hence $g \circ f = 1_D$, and g is a left-inverse of f. ■

If $D = \emptyset \neq C$, then $f := 1_{\emptyset \subset C}$ is injective, but not left-invertible, since $\mathrm{Map}(C, D) = \emptyset$.

•33E.R. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(RInv): *f is right-invertible.*

(Surj): *f is surjective.*

Proof. (RInv) $\Rightarrow$ (Surj). Choose a right-inverse g of f; then $f \circ g = 1_C$, which is bijective. Hence f is surjective (Proposition 32B.R, (b)).

•(Surj) $\Rightarrow$ (RInv). Assume that f is surjective. For each $y \in C$ we have $f^<(\{y\}) \neq \emptyset$. We may •therefore choose a mapping $g\colon C \to D$ such that $g(y) \in f^<(\{y\})$ for each $y \in C$. We then have $f(g(y)) = y$ for all $y \in C$, so that $f \circ g = 1_C$. Thus g is a right-inverse of f. ■

33E.B. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(Inv): *f is invertible.*

(Bij): *f is bijective.*

Proof. (Inv) $\Rightarrow$ (Bij). If f is invertible, it is both left-invertible and right-invertible, and hence both injective and surjective (Propositions 33E.L and 33E.R).

(Bij) $\Rightarrow$ (Inv). Assume that f is bijective. For each $y \in C$, $f^<(\{y\})$ is a singleton. Define $g\colon C \to D$ by the rule $g(y) :\in f^<(\{y\})$ for each $y \in C$. For every $x \in D$ we then have $x \in f^<(\{f(x)\}) = \{g(f(x))\}$, so that $x = g(f(x))$; thus $g \circ f = 1_D$. For every $y \in C$, $f(g(y)) \in f_>(\{g(y)\}) = f_>(f^<(\{y\})) \subset \{y\}$, so that $f(g(y)) = y$; thus $f \circ g = 1_C$. We conclude that g is an inverse of f. ■

For given sets D, C we shall sometimes consider the set of all invertible mappings from D to C, i.e., $\mathrm{Inv}(D, C) := \{f \in \mathrm{Map}(D, C) \mid f \text{ invertible}\}$. D and C are said to be **equinumerous**, and D is said to be **equinumerous to** C, if there exists an invertible mapping from D to C, i.e., if $\mathrm{Inv}(D, C) \neq \emptyset$. We note that in every collection of sets the relation "is equinumerous to" is an equivalence relation.

Let the set D be given. A mapping from D to D is called a **permutation of** D if it is invertible. The set of all permutations of D is denoted by $\mathrm{Perm}(D)$. Thus $\mathrm{Perm}(D) := \mathrm{Inv}(D, D)$.

33F. Proposition. *Let the set D and the mappings $f, g \in \mathrm{Map}(D, D)$ be given. If g is invertible (i.e., a permutation of D) and f and g commute, then f and $g^{\leftarrow}$ commute.*

34. Injectivity, surjectivity, and bijectivity: The induced mappings

34A.L. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(Inj): *f is injective.*

(Inj$_4$): $f^< \circ f_> = 1_{\mathfrak{P}(D)}$.

(Inj$_5$): *$f_>$ is injective.*

(Inj$_6$): *$f^<$ is surjective.*

(Inj$_7$): $\forall U \in \mathfrak{P}(D),\ f_>(D\backslash U) \subset C\backslash f_>(U)$.

Proof. (Inj) $\Rightarrow$ (Inj$_4$). Assume that f is injective. Let $U \in \mathfrak{P}(D)$ and $x \in f^<(f_>(U))$ be given. Then $f(x) \in f_>(U)$; hence we may choose $z \in U$ such that $f(x) = f(z)$. Since f is injective, we conclude that $x = z \in U$. Since $x \in f^<(f_>(U))$ was arbitrary, we find that $f^<(f_>(U)) \subset U$. Since the reverse inclusion holds by Proposition 23A, we have $f^<(f_>(U)) = U$. Since $U \in \mathfrak{P}(D)$ was arbitrary, (Inj$_4$) is verified.

(Inj$_4$) $\Rightarrow$ (Inj$_5$) *and* (Inj$_4$) $\Rightarrow$ (Inj$_6$). These implications follow at once from Propositions 32B.L,(b) and 32B.R,(b).

(Inj$_5$) $\Rightarrow$ (Inj). Assume that (Inj$_5$) holds. Given $x, x' \in D$ such that $f(x) = f(x')$, we have $f_>(\{x\}) = \{f(x)\} = \{f(x')\} = f_>(\{x'\})$. Since $f_>$ is injective, we conclude that $\{x\} = \{x'\}$, and hence $x = x'$.

(Inj$_6$) $\rightarrow$ (Inj). Assume that (Inj$_6$) holds. Given $x, x' \in D$ such that $f(x) = f(x')$, we may choose $V \in \mathfrak{P}(C)$ such that $\{x\} = f^<(V)$. Then $f(x') = f(x) \in V$. Therefore $x' \in f^<(\{f(x')\}) \subset f^<(V) = \{x\}$, so that $x' = x$.

(Inj$_4$) $\Rightarrow$ (Inj$_7$). Assume that (Inj$_4$) holds. For every $U \in \mathfrak{P}(D)$ we have $f^<(C\backslash f_>(U)) = D\backslash f^<(f_>(U)) = D\backslash U$, and hence $f_>(D\backslash U) = f_>(f^<(C\backslash f_>(U))) \subset C\backslash f_>(U)$.

(Inj$_7$) $\Rightarrow$ (Inj$_4$). Assume that (Inj$_7$) holds. For every $U \in \mathfrak{P}(D)$ we have $f_>(U) = f_>(D\backslash(D\backslash U)) \subset C\backslash f_>(D\backslash U)$, and hence

$$U \subset f^<(f_>(U)) \subset f^<(C\backslash f_>(D\backslash U)) = D\backslash f^<(f_>(D\backslash U)) \subset D\backslash(D\backslash U) = U;$$

we conclude that equality must hold, and therefore $f^< \circ f_> = 1_{\mathfrak{P}(D)}$. ■

34A.R. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(Surj): *f is surjective.*

(Surj$_4$): $f_> \circ f^< = 1_{\mathfrak{P}(C)}$.

(Surj$_5$): *$f_>$ is surjective.*

(Surj$_6$): *$f^<$ is injective.*

(Surj$_7$): $\forall U \in \mathfrak{P}(D),\ \ f_>(D\backslash U) \supset C\backslash f_>(U)$.

Proof. (Surj) $\Rightarrow$ (Surj$_4$). Assume that f is surjective. For every $V \in \mathfrak{P}(C)$, $f_>(f^<(V)) = V \cap \mathrm{Rng} f = V \cap C = V$. Thus $f_> \circ f^< = 1_{\mathfrak{P}(C)}$.

$(\mathrm{Surj}_4) \Rightarrow (\mathrm{Surj}_5)$ *and* $(\mathrm{Surj}_4) \Rightarrow (\mathrm{Surj}_6)$. These implications follow at once form Proposition 32B.L,(b) and 32B.R,(b).

$(\mathrm{Surj}_5) \Rightarrow (\mathrm{Surj})$. If $f_>$ is surjective, we may choose $U \in \mathfrak{P}(D)$ such that $f_>(U) = C$. Then $C = f_>(U) \subset f_>(D) \subset C$, and hence $\mathrm{Rng}f = f_>(D) = C$.

$(\mathrm{Surj}_6) \Rightarrow (\mathrm{Surj})$. $f^<(\mathrm{Rng}f) = D = f^<(C)$. If $f^<$ is injective, $\mathrm{Rng}f = C$.

$(\mathrm{Surj}_4) \Rightarrow (\mathrm{Surj}_7)$. Assume that (Surj_4) holds. For every $U \in \mathfrak{P}(D)$ we have $f^<(C\backslash f_>(U)) = D\backslash f^<(f_>(U)) \subset D\backslash U$ and hence $f_>(D\backslash U) \supset f_>(f^<(C\backslash f_>(U))) = C\backslash f_>(U)$.

$(\mathrm{Surj}_7) \Rightarrow (\mathrm{Surj})$. Assume that (Surj_7) holds. Then $C \supset f_>(D) = f_>(D\backslash\emptyset) \supset C\backslash f_>(\emptyset) = C\backslash\emptyset = C$. Hence $\mathrm{Rng}f = f_>(D) = C$. ■

34A.B. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(Bij): *f is bijective.*

(Bij_4): *$f_>$ and $f^<$ are invertible, and each is the inverse of the other.*

(Bij_5): *$f_>$ is bijective.*

(Bij_6): *$f^<$ is bijective.*

(Bij_7): $\forall U \subset \mathfrak{P}(D), \quad f_>(D\backslash U) = C\backslash f_>(U)$.

34B.L. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(Inj): *f is injective.*

(Inj_8): *for every non-empty subcollection $\mathcal{U}$ of $\mathfrak{P}(D)$,* $f_>(\bigcap \mathcal{U}) = \bigcap (f_>)_>(\mathcal{U})$.

(Inj_9): $\forall U', U'' \in \mathfrak{P}(D), \quad U' \cap U'' = \emptyset \quad \Rightarrow \quad f_>(U') \cap f_>(U'') = \emptyset$.

Proof. $(\mathrm{Inj}) \Rightarrow (\mathrm{Inj}_8)$. Although a direct proof is not complicated, we use Proposition 34A.L and assume that $f^< \circ f_> = 1_{\mathfrak{P}(D)}$. It follows that $(f^<)_> \circ (f_>)_> = (1_{\mathfrak{P}(D)})_> = 1_{\mathfrak{P}(\mathfrak{P}(D))}$. Let a non-empty subcollection $\mathcal{U}$ of $\mathfrak{P}(D)$ be given. Choose $U \in \mathcal{U}$; then $f_>(U) \in (f_>)_>(\mathcal{U})$, whence $\bigcap (f_>)_>(\mathcal{U}) \subset f_>(U) \subset \mathrm{Rng}f$. On the other hand, Proposition 23A implies

$$f^<(\bigcap (f_>)_>(\mathcal{U})) = \bigcap (f^<)_>((f_>)_>(\mathcal{U})) = \bigcap \mathcal{U};$$

using Proposition 23A again, and applying $f_>$ to the ends of this chain of equalities, we find

$$\bigcap (f_>)_>(\mathcal{U}) = (\bigcap (f_>)_>(\mathcal{U})) \cap \mathrm{Rng}f = f_>(f^<(\bigcap (f_>)_>(\mathcal{U}))) = f_>(\bigcap \mathcal{U}).$$

$(\mathrm{Inj}_8) \Rightarrow (\mathrm{Inj}_9)$. This is trivial: set $\mathcal{U} := \{U', U''\}$.

$(\mathrm{Inj}_9) \Rightarrow (\mathrm{Inj})$. Assume that (Inj_9) holds. Let $x, x' \in D$ be given. If $x \neq x'$, then $\{x\} \cap \{x'\} = \emptyset$, and therefore $\{f(x)\} \cap \{f(x')\} = f_>(\{x\}) \cap f_>(\{x'\}) = \emptyset$, so that $f(x) \neq (x')$. ■

34B.B. Proposition. *Let the mapping $f: D \to C$ be given. The following statements are equivalent:*

(Bij): *f is bijective.*

(Bij$_8$): *for every subcollection $\mathcal{U}$ of $\mathfrak{P}(D)$,* $\quad f_>(\bigcap^D \mathcal{U}) = \bigcap^C (f_>)_>(\mathcal{U})$.

Proof. (Bij) $\Rightarrow$ (Bij$_8$). Assume that f is bijective. Let the subcollection $\mathcal{U}$ of $\mathfrak{P}(D)$ be given. If $\mathcal{U} \neq \emptyset$, Proposition 34B.L shows that $f_>(\bigcap^D \mathcal{U}) = \bigcap^C (f_>)_>(\mathcal{U})$. On the other hand, f is surjective, and hence $f_>(\bigcap^D \emptyset) = f_>(D) = \mathrm{Rng} f = C = \bigcap^C \emptyset = \bigcap^C (f_>)_>(\emptyset)$.

(Bij$_8$) $\Rightarrow$ (Bij). By Proposition 34B.L, f is injective; and $\mathrm{Rng} f = f_>(D) = f_>(\bigcap^D \emptyset) = \bigcap^C (f_>)_>(\emptyset) = \bigcap^C \emptyset = C$. so that f is surjective. ■

34C. Remark. The proof of Proposition 34B.B includes the proof of a vestigial Proposition 34B.R: *The mapping $f: D \to C$ is surjective if and only if $f_>(\bigcap^D \emptyset) = \bigcap^C (f_>)_>(\emptyset)$.* ■

35. Cancellability

Let $f\colon D \to C$ be a mapping and S a set. Then f is said to be **left-cancellable with respect to** S if

$$\forall g, h \in \ \mathrm{Map}(S, D), \quad f \circ g = f \circ h \ \Rightarrow \ g = h;$$

and f is said to be **right-cancellable with respect to** S if

$$\forall g, h \in \mathrm{Map}(C, S), \quad g \circ f = h \circ f \ \Rightarrow \ g = h.$$

A mapping f is said to be **left-cancellable** (and is called a **set-monomorphism**) if f is left-cancellable with respect to S for every set S; and f is said to be **right-cancellable** (and is called a **set-epimorphism**) if f is right-cancellable with respect to S for every set S.

35A.L. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(Inj): *f is injective.*

(Mono): *f is left-cancellable.*

(Mono_1): *f is left-cancellable with respect to some singleton.*

(Mono_2): *f is left-cancellable with respect to some non-empty set.*

Proof. (Inj) $\Rightarrow$ (Mono). Assume that f is injective. Let the set S and the mappings $g, h \in \mathrm{Map}(S, D)$ be such that $f \circ g = f \circ h$. For each $s \in S$ we have $f(g(s)) = f(h(s))$, and hence $g(s) = h(s)$. Therefore $g = h$.

(Mono) $\Rightarrow$ (Mono_1) *and* (Mono_1) $\Rightarrow$ (Mono_2). These implications are trivial.

(Mono_2) $\Rightarrow$ (Inj). Assume that (Mono_2) holds, and choose a non-empty set S such that f is left-cancellable with respect to S. If $x, x' \in D$ are such that $f(x) = f(x')$, we have $f \circ x_{S\to D} = f \circ x'_{S\to D}$, and hence $x_{S\to D} = x'_{S\to D}$. Since $S \neq \emptyset$, it follows that $x = x'$. ■

Remark. Every mapping $f\colon D \to C$ is left-cancellable with respect to $\emptyset$, since $\mathrm{Map}(\emptyset, D)$ is the singleton $\{1_{\emptyset\subset D}\}$.

35A.R. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(Surj): *f is surjective.*

(Epi): *f is right-cancellable.*

(Epi_1): *f is right-cancellable with respect to some doubleton.*

(Epi_2): *f is right-cancellable with respect to some set that is neither empty nor a singleton.*

Proof. (Surj) $\Rightarrow$ (Epi). Assume that f is surjective. Let the set S and the mappings $g, h \in \mathrm{Map}(C, S)$ be such that $g \circ f = h \circ f$. Given $y \in C$, choose $x \in f^{<}(\{y\})$; then $g(y) = g(f(x)) = h(f(x)) = h(y)$. Thus $g(y) = h(y)$ for all $y \in C$, and hence $g = h$.

(Epi) $\Rightarrow$ (Epi_1) *and* (Epi_1) $\Rightarrow$ (Epi_2). These implications are trivial.

(Epi$_2$) $\Rightarrow$ (Surj). Assume that (Epi$_2$) holds, and choose a set S, neither empty nor a singleton, such that f is right-cancellable with respect to S. Choose $s, s' \in S$ such that $s \neq s'$. Define $g, h \in \mathrm{Map}(C, S)$ by $g := s_{C \to S}$ and

$$h(y) := \begin{cases} s & \text{if } y \in \mathrm{Rng} f \\ \\ s' & \text{if } y \in C \backslash \mathrm{Rng} f. \end{cases}$$

For every $x \in D$, $f(x) \in \mathrm{Rng} f$ and hence $g(f(x)) = s = h(f(x))$, so that $g \circ f = s_{D \to S} = h \circ f$. By the assumption, $g = h$. For every $y \in C \backslash \mathrm{Rng} f$ this implies $s = g(y) = h(y) = s'$, contradicting the assumption that $s \neq s'$. Therefore $C \backslash \mathrm{Rng} f = \emptyset$, so that f is surjective. ■

Remark. Every mapping $f\colon D \to C$ is right-cancellable with respect to $\emptyset$ as well as with respect to every singleton $\{s\}$, since $\mathrm{Map}(C, \emptyset)$ is empty if $C \neq \emptyset$ and is the singleton $\{1_\emptyset\}$ when $C = \emptyset$, while $\mathrm{Map}(C, \{s\})$ is the singleton $\{s_{C \to \{s\}}\}$.

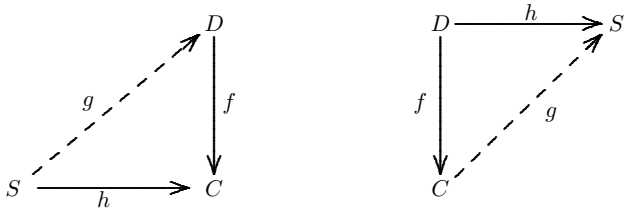
36. Factorization

We next discuss questions of the following kind: Given mappings $f\colon D \to C$ and $h\colon S \to C$, is there a mapping $g\colon S \to D$ such that $h = f \circ g$? Given mappings $f\colon D \to C$ and $h\colon D \to S$, is there a mapping $g\colon C \to S$ such that $h = g \circ f$? If there is such a mapping, is there exactly one? (The question of the left- or right-invertibility of f is a special case of this, in which h is an identity mapping.) Some *necessary* conditions for the *existence* of g are known from (25.3) and (25.4): for every mapping $f\colon D \to C$ and set S,

$$\text{(36.1)} \qquad \mathrm{Rng}(f \circ g) \subset \mathrm{Rng} f \qquad \text{for all } g \in \mathrm{Map}(S, D)$$

$$\text{(36.2)} \qquad \mathrm{Part}(g \circ f) \sqsubset \mathrm{Part} f \qquad \text{for all } g \in \mathrm{Map}(C, S).$$

A mapping $f\colon D \to C$ is called a **(set-)embedding** if for every mapping $h\colon S \to C$ with $\mathrm{Rng} h \subset \mathrm{Rng} f$ there is exactly one mapping $g\colon S \to D$ such that $h = f \circ g$. A mapping $f\colon D \to C$ is called a **(set-)quotient-mapping** if for every mapping $h\colon D \to S$ with $\mathrm{Part} h \sqsubset \mathrm{Part} f$ there is exactly one mapping $g\colon C \to S$ such that $h = g \circ f$.



36A.L. Proposition. *If $f : D \to C$ and $f' : D' \to C$ are embeddings with $\mathrm{Rng} f = \mathrm{Rng} f'$, then the unique mappings $g\colon D' \to D$ and $g' : D \to D'$ that satisfy $f \circ g = f'$, $f' \circ g' = f$ are invertible, and each is the inverse of the other.*

Proof. We have $f \circ g \circ g' = f' \circ g' = f = f \circ 1_D$; by the uniqueness condition in the definition of embedding, $g \circ g' = 1_D$. Similarly, $f' \circ g' \circ g = f' \circ 1_{D'}$, whence $g' \circ g = 1_{D'}$. ■

36A.R. Proposition. *If $f\colon D \to C$ and $f' : D \to C'$ are quotient-mappings with $\mathrm{Part} f = \mathrm{Part} f'$, then the unique mappings $g\colon C \to C'$ and $g' : C' \to C$ that satisfy $g \circ f = f'$, $g' \circ f' = f$ are invertible, and each is the inverse of the other.*

Proof. The same as the proof of Proposition 36A.L, with the obvious modifications. ■

We now show that set-embeddings are precisely the injections, and set-quotient-mappings are precisely the surjections.

36B.L. Proposition. *Let the mapping $f : D \to C$ be given. The following statements are equivalent:*

(Emb): *f is an embedding.*

(Mono): *f is left-cancellable.*

(Inj): *f is injective.*

Proof. (Emb) ⇒ (Mono). Assume that f is an embedding. Let the set S be given, and let the mappings $g, h \in \mathrm{Map}(S, D)$ satisfy $f \circ g = f \circ h$. Then $\mathrm{Rng}(f \circ g) = \mathrm{Rng}(f \circ h) \subset \mathrm{Rng} f$. Since f is an embedding, we must have $g = h$.

(Mono) ⇒ (Inj). Proposition 35A.L.

(Inj) ⇒ (Emb). Assume that f is injective and let the mapping $h\colon S \to C$ with $\mathrm{Rng} h \subset \mathrm{Rng} f$ be given. A mapping $g\colon S \to D$ satisfies $h = f \circ g$ if and only if $f(g(s)) = h(s)$ or, equivalently, $g(s) \in f^<(\{h(s)\})$, for each $s \in S$; but $f^<(\{h(s)\})$ is a singleton for each $s \in S$, since f is injective and $h(s) \in \mathrm{Rng} h \subset \mathrm{Rng} f$. Therefore there is indeed exactly one $g\colon S \to D$ such that $h = f \circ g$: it is defined by the rule $g(s) :\in f^<(\{h(s)\})$ for all $s \in S$. ■

36B.R. Proposition. *Let the mapping $f\colon D \to C$ be given. The following statements are equivalent:*

(Quot): *f is a quotient-mapping.*

(Epi): *f is right-cancellable.*

(Surj): *f is surjective.*

Proof. (Quot) ⇒ (Epi). Assume that f is a quotient-mapping. Let the set S be given, and let the mappings $g, h \in \mathrm{Map}(C, S)$ satisfy $g \circ f = h \circ f$. Then $\mathrm{Part}(g \circ f) = \mathrm{Part}(h \circ f) \sqsubset \mathrm{Part} f$. Since f is a quotient-mapping, we must have $g = h$.

(Epi) ⇒ (Surj). Proposition 35A.R.

(Surj) ⇒ (Quot). Assume that f is surjective and let the mapping $h\colon D \to S$ with $\mathrm{Part} h \sqsubset \mathrm{Part} f$ be given. Since f is surjective, $f^<(\{y\}) \in \mathrm{Part} f$ for each $y \in C$, and since $\mathrm{Part} h \sqsubset \mathrm{Part} f$ it follows that $h_>(f^<(\{y\}))$ is a singleton for every $y \in C$. If a mapping $g\colon C \to S$ satisfies $h = g \circ f$, then

$$h_>(f^<(\{y\})) = g_>(f_>(f^<(\{y\}))) \subset g_>(\{y\}) = \{g(y)\}$$

for all $y \in C$, so that

$$g(y) \in h_>(f^<(\{y\})) \qquad \text{for all } y \in C. \tag{36.3}$$

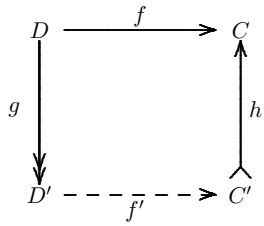
Conversely, if g satisfies (36.3), then $h(x) \in h_>(f^<(\{f(x)\})) = \{g(f(x))\}$ for all $x \in D$, so that $h = g \circ f$. We conclude that there is exactly one mapping $g\colon C \to S$ such that $h = g \circ f$, namely the one defined by the rule $g(y) :\in h_>(f^<(\{y\}))$ for all $y \in C$. ■

36C. Theorem. *Let the mappings $g\colon D \to D'$ and $h\colon C' \to C$ be given, and assume that g is surjective and h is injective. Then:*

(a): *for a given mapping $f \in \mathrm{Map}(D, C)$ there is at most one $f' \in \mathrm{Map}(D', C')$ such that $f = h \circ f' \circ g$; such a mapping f' exists if and only if* $\mathrm{Part} f \sqsubset \mathrm{Part} g$ *and* $\mathrm{Rng} f \subset \mathrm{Rng} h$.

(b): *this mapping f' is injective if and only if* $\mathrm{Part} f = \mathrm{Part} g$, *and is surjective if*

and only if $\mathrm{Rng}f = \mathrm{Rng}h$.

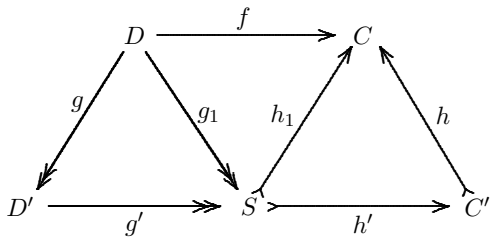


Proof. The parts of the proof that pertain to Part (b) of the statement will be given in square brackets.

Suppose that $f', f'' \in \mathrm{Map}(D', C')$ satisfy $h \circ f' \circ g = f = h \circ f'' \circ g$. Since h is left-cancellable and g is right-cancellable (Propositions 35A.L and 35 A.R), we must have $f' = f''$. Hence there is at most one $f' \in \mathrm{Map}(D', C')$ such that $h \circ f' \circ g = f$.

If $f' \in \mathrm{Map}(D', C')$ satisfies $h \circ f' \circ g = f$, then (36.2) and (36.1) imply $\mathrm{Part}f = \mathrm{Part}((h \circ f') \circ g) \sqsubset \mathrm{Part}g$ and $\mathrm{Rng}f = \mathrm{Rng}(h \circ (f' \circ g)) \subset \mathrm{Rng}h$. [If f' is injective, so is $h \circ f'$, and $\mathrm{Part}f = \mathrm{Part}g$ (Propositions 32B.L,(a) and 32C.L). If f' is surjective, so is $f' \circ g$, and $\mathrm{Rng}f = \mathrm{Rng}h$ (Propositions 32B.R,(a) and 32C.R).]

Conversely, assume that f satisfies $\mathrm{Part}f \sqsubset \mathrm{Part}g$ and $\mathrm{Rng}f \subset \mathrm{Rng}h$.



By Proposition 32D we may choose a surjective mapping $g_1 : D \to S$ and an injective mapping $h_1 : S \to C$ such that $f = h_1 \circ g_1$. By Propositions 32C.L and 32C.R, $\mathrm{Part}g_1 = \mathrm{Part}f$ and $\mathrm{Rng}h_1 = \mathrm{Rng}f$.

Since $\mathrm{Part}g_1 = \mathrm{Part}f \sqsubset \mathrm{Part}g$, and g is a quotient-mapping (Proposition 36B.R), there is exactly one mapping $g' \colon D' \to S$ such that $g_1 = g' \circ g$; this mapping is surjective (Proposition 32B.R,(b)).

Similarly, since $\mathrm{Rng}h_1 = \mathrm{Rng}f \subset \mathrm{Rng}h$, and h is an embedding (Proposition 36B.L), there is exactly one mapping $h' : S \to C'$ such that $h_1 = h \circ h'$; this mapping is injective (Proposition 32B.L,(b)).

We conclude that $f = h_1 \circ g_1 = h \circ h' \circ g' \circ g = h \circ f' \circ g$, where $f' := h' \circ g'$.

[If $\mathrm{Part}f = \mathrm{Part}g$, then $\mathrm{Part}g_1 = \mathrm{Part}g$, and g' is invertible, by Proposition 36A.R; hence both g' and h' are injective, and so is their composite f'. If, on the other hand, $\mathrm{Rng}f = \mathrm{Rng}h$, then $\mathrm{Rng}h_1 = \mathrm{Rng}h$, and h' is invertible, by Proposition 36A.L; hence both g' and h' are surjective, and so is their composite f'.] ■

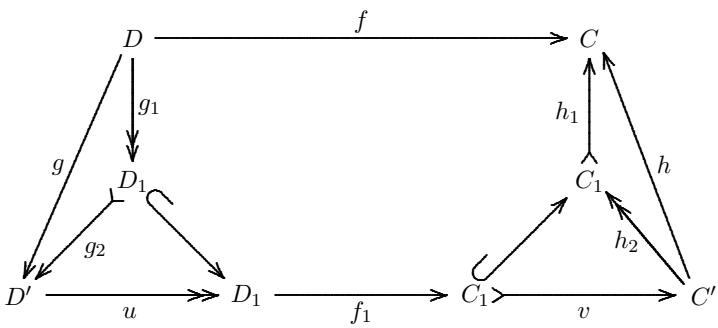
36D. Corollary. *Let the mapping $f \colon D \to C$ be given. Then there is a unique mapping $f' : \mathrm{Part}f \to \mathrm{Rng}f$ such that $f = 1_{\mathrm{Rngf}\subset\mathrm{C}} \circ f' \circ \Omega_{\mathrm{Part}f}$; this mapping f' is bijective.*

▼ For the sake of completeness, we record a factorization theorem that is more general than Theorem 36C in that it makes no assumptions on g and h.

•36E. Theorem. *Let the mappings $g\colon D \to D'$ and $h\colon C' \to C$ be given. For every $f \in \mathrm{Map}(D,C)$ there is $f' \in \mathrm{Map}(D',C')$ with $f = h \circ f' \circ g$ if and only if $\mathrm{Part}f \sqsubset \mathrm{Part}g$ and $\mathrm{Rng}f \subset \mathrm{Rng}h$, and $D \neq \emptyset$ or $C' \neq \emptyset$ or $D' = \emptyset$. There is at most one such f' for every $f \in \mathrm{Map}(D,C)$ if and only if either g is surjective and h is injective, or $D' = \emptyset$, or C' is empty or a singleton.*

Proof. If $f' \in \mathrm{Map}(D',C')$ satisfies $f = h \circ f' \circ g$, then $\mathrm{Part}f = \mathrm{Part}((h \circ f') \circ g) \sqsubset \mathrm{Part}g$ and $\mathrm{Rng}f = \mathrm{Rng}(h \circ (f' \circ g)) \subset \mathrm{Rng}h$. If $D = \emptyset$ and $C' = \emptyset$ we have $\mathrm{Map}(D,C) = \{1_{\emptyset \subset C}\} \neq \emptyset$; if there then exists $f' \in \mathrm{Map}(D',C')$ we must have $\mathrm{Map}(D',\emptyset) \neq \emptyset$, whence $D' = \emptyset$.

Assume, conversely, that $D \neq \emptyset$ or $C' \neq \emptyset$ or $D' = \emptyset$, and that $f \in \mathrm{Map}(D,C)$ is given and satisfies $\mathrm{Part}f \sqsubset \mathrm{Part}g$ and $\mathrm{Rng}f \subset \mathrm{Rng}h$. If $D = \emptyset$ and $D' \neq \emptyset$, this implies $f = 1_{\emptyset \subset C}$ and $g = 1_{\emptyset \subset D'}$ and $C' \neq \emptyset$, which in turn implies $\mathrm{Map}(D',C') \neq \emptyset$; choose any $f' \in \mathrm{Map}(D',C')$; then $h \circ f' \circ g = h \circ f' \circ 1_{\emptyset \subset D'} = 1_{\emptyset \subset C} = f$.



We shall therefore assume from now on that $D \neq \emptyset$ or $D' = \emptyset$. By Proposition 32D we may choose mappings $g_1 : D \to D_1$, $g_2 : D_1 \to D'$ and $h_1 : C_1 \to C, h_2 : C' \to C_1$ such that g_2 and h_1 are injective, g_1 and h_2 are surjective, and $g = g_2 \circ g_1$ and $h = h_1 \circ h_2$. By Theorem 36C there is exactly one $f_1 : D_1 \to C_1$ such that $f = h_1 \circ f_1 \circ g_1$. By Proposition 33E.L, we may choose a left-inverse of g_2, say $u\colon D' \to D_1$; by •Proposition 33E.R, we may choose a right-inverse of h_2, say $v\colon C_1 \to C'$. Then $f = h_1 \circ 1_{C_1} \circ f_1 \circ 1_{D_1} \circ g_1 = h_1 \circ h_2 \circ v \circ f_1 \circ u \circ g_2 \circ g_1 = h \circ f' \circ g$ with $f' := v \circ f_1 \circ u$.

We now examine the uniqueness question. Suppose that $f', f'' \in \mathrm{Map}(D',C')$ satisfy $h \circ f' \circ g = h \circ f'' \circ g$. If g is surjective and h is injective, then g is right-cancellable and h is left-cancellable, and it follows that $f' = f''$; if, on the other hand, $D' = \emptyset$ or C' is empty or a singleton, then $\mathrm{Map}(D',C')$ is empty or a singleton, and hence again $f' = f''$. Conversely, if

$$\forall f', f'' \in \mathrm{Map}(D',C'), \qquad h \circ f' \circ g = h \circ f'' \circ g \quad \Rightarrow \quad f' = f'',$$

it follows that g is right-cancellable with respect to C' and h is left-cancellable with respect to D'; by Propositions 35A.L and 35A.R, this implies that g is surjective or ▲ C' is empty or a singleton, and that h is injective or $D' = \emptyset$. ■

This page intentionally left blank

Chapter 4

FAMILIES

41. The concept of a family

We are well acquainted with the notion of a sequence: a sequence of numbers, say, is given by some rule that determines what number comes in the nth place, for each $n \in \mathbb{N}$ (or $n \in \mathbb{N}^{\times}$, for those who prefer to count from 1). The natural number n is thought of as a label, an *index*, and the number in the nth place as the nth *term* of the sequence. We must resist the urge to consider a sequence as some kind of infinite set, or as an "ordered set": apart from the basic conceptual differences involved here, the mere fact that one and the same number may appear as the nth term of the sequence for more than one $n \in \mathbb{N}$ (indeed, perhaps for *all* $n \in \mathbb{N}$) should prevent such a confusion. There is much more to be said for the view that a sequence of numbers is a special kind of mapping: one whose domain is $\mathbb{N}$. One must merely make the right translation: the *index set* $\mathbb{N}$ becomes the domain, the nth term becomes the value at n.

More generally, any unambiguous method that allows us to associate with each member i of a given set I an object a_i determines a **family** a **(defined) on** the set I, called the **index set**; for each $i \in I$, the object a_i is called the **term of index** i, or the i**th term, of** the family a. In a manner similar to that used for mappings, the families a and b, with respective index sets I and J, are the same family if and only if $I = J$ and $a_i = b_i$ for all $i \in I$. In some contexts, notations such as $a(i), a^i$, etc., are used instead of a_i. A family is said to be **empty** or **non-empty** according as its index set is empty or non-empty.

One often uses the notation $(a_i \mid i \in I)$ for a family, especially when no explicit name for the family itself is available; thus, for instance, $(n^3 \mid n \in \mathbb{N})$ denotes a family. (Note the use of parentheses, not braces.)

The set of all the terms of a family a is called the **range of** a, and is denoted by $\mathrm{Rng}a$ or by $\{a_i \mid i \in I\}$, where I is the index set:

$$\mathrm{Rng}a := \{a_i \mid i \in I\} \stackrel{*}{:=} \{x \mid x = a_i \text{ for some } i \in I\}.$$

If a is a family and A is a set such that $\mathrm{Rng}a \subset A$, one says that a is a **family in** A, or a **family of** members of A.

We observe that the concept of a family is almost identical to that of a mapping, with the following short glossary:

index set	domain
ith term, term of index i	value at i
$(a_i \mid i \in I)$	$(i \mapsto a_i) : I \to$

Except for the matter of the presence or absence of a codomain, to be discussed a little later, there is no formal need to maintain a separate concept of family and the associated terminology. It must be said, however, that the two "languages", of mappings and families, "sound" different to the mathematician, suggesting different associations of ideas, and that both are used profusely; it is therefore necessary to become fluent in both.

We note that in our description of the concept of family there was nothing corresponding to the codomain of a mapping. It is most nearly consistent with the common usage concerning families to regard them as unprovided with a codomain (or, in some intuitive sense, as having "everything" as the codomain); not all mathematicians would agree. We should like, however, to make use of the terminology and notations introduced for mappings when dealing with families; as well as to pay some respect to Occam's Razor (*entia non sunt multiplicanda praeter necessitatem*; roughly, Concepts should not be multiplied beyond need). For formal purposes, we may therefore make a family into a mapping by providing it with a "formal" codomain: for this role we select, for lack of a better choice, the *range* of the family. We thus have the following definition, which supersedes, in a purely formal sense, the preceding discussion: A **family** is defined to be a surjective mapping. The terminology and notation not covered by the preceding glossary will be accounted for by this definition, with the following amendments: If $a := (a_i \mid i \in I)$ is a family and A is a set, we define the **pre-image of** A under a even when A is not included in $\mathrm{Rng}a$:

$$a^<(A) := \{i \in I \mid a_i \in A\} = a^<(A \cap \mathrm{Rng}a);$$

and in any composition involving a family — at most one will appear, the other ingredients being mappings — its codomain is adjusted as needed to any set that includes its range, while the composite is again regarded as a family. As a case in point, if $J \subset I$, the restriction to J of the family $(a_i \mid i \in I)$ is the family $(a_i \mid i \in J)$.

42. Special families

For some kinds of index set, special terminology is used. If $n \in \mathbb{N}$, a family defined on $n^{\sqsubset} := \{0, \dots, n-1\}$ or on $n^{\sqsupset} := \{1, \dots, n\}$ (formal definitions of these sets will be given later) is called a **list of length** n. (A common barbarism for this is "n-tuple"; a term often used is *(finite) sequence of length* n.) If the length is small, a list a may be denoted by $(\ )$, (a_1), (a_1, a_2), (a_1, a_2, a_3), and so on, according to the length. A family defined on $\mathbb{N}$ or on $\mathbb{N}^{\times} := \mathbb{N}\backslash\{0\}$ is called a **sequence**.

If I, J are sets, a family defined on $I \times J$ is called an $I \times J$-**matrix**. The term of index (i, j) of an $I \times J$-matrix M is called its (i, j)-**entry**, and is usually written $M_{i,j}$ or even M_{ij}, instead of $M_{(i,j)}$. For each $i \in I$, the family $(M_{i,j} \mid j \in J)$ is called the i**th row of** the matrix M, and for each $j \in J$ the family $(M_{i,j} \mid i \in I)$ is called the j**th column of** M. The $J \times I$-matrix $M^{\top} := M \circ \top_{J \times I}$, satisfying $(M^{\top})_{j,i} = M_{i,j}$ for all $(j, i) \in J \times I$, is called the **transpose of** M, and is said to be obtained from M by **transposition**. If $m, n \in \mathbb{N}$, an $m^{\sqsubset} \times n^{\sqsubset}$-matrix or an $m^{\sqsupset} \times n^{\sqsupset}$-matrix is called an m-**by**-n-**matrix**. If m and n are small, the familiar bookkeeping scheme may be used: e.g.,

$$M =: \begin{pmatrix} M_{1,1} & M_{1,2} & M_{1,3} \\ M_{2,1} & M_{2,2} & M_{2,3} \end{pmatrix}.$$

The $I \times J$-matrix M is said to be a **square** matrix if $I = J$. In this case, M is said to be **symmetric** if $M^{\top} = M$, i.e., if $M_{j,i} = M_{i,j}$ for all $(i, j) \in I \times I$. If M is a (square) $I \times I$-matrix, the family $(M_{i,i} \mid \imath \in I)$ is called the **diagonal of** M.

Although families and sets must not be confused, we often wish to consider the family $(x \mid x \in S)$, where S is a given set; formally, this is just the identity mapping 1_S. It is a commonly accepted license to call this the **family** S or, more fully, the **set** S **self-indexed (as a family)**, and to use for it the same symbol S as for the set.

Let S be a set and U a subset of S. The family $\chi_{U \subset S}$ defined on S by

$$\chi_{U \subset S}(x) := \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{if } x \in S\backslash U \end{cases}$$

is called the **characteristic family of** U (**in** S), or the **characteristic function of** U (**in** S) when a codomain — usually $\mathbb{R}$, sometimes the doubleton $\{0, 1\}$ or $\mathbb{N}$ or the interval $[0, 1]$ — is specified. When the domain (index set) S is fixed throughout a discussion, the notation is frequently abbreviated to χ_U.

42A. Proposition. *Let the set S be given. The mapping*

$$U \mapsto \chi_{U \subset S} : \mathfrak{P}(S) \to \mathrm{Map}(S, \{0, 1\})$$

is invertible: its inverse is

$$f \mapsto f^{<}(\{1\}) : \mathrm{Map}(S, \{0, 1\}) \to \mathfrak{P}(S).$$

For every $x \in S$, the family $\delta_x^S := \chi_{\{x\} \subset S}$ is called the x**th Kronecker family** (or **function**). We note that δ_x^S is the xth row of the symmetric $S \times S$-matrix $\delta^S := \chi_{\triangle_S \subset S \times S}$, called the **Kronecker matrix of** S, and we have

$$(\delta_x^S)_y = \delta_{x,y}^S = \begin{cases} 1 & \text{if } x = y \\ \\ 0 & \text{if } x \neq y \end{cases} \qquad \text{for all } x, y \in S.$$

Let the sets D and C be given, and let F be a given subset of $\mathrm{Map}(D, C)$. We define the **evaluation family** ev^F on D, with terms in $\mathrm{Map}(F, C)$, by requiring ${\mathrm{ev}^F}_x$, the **evaluation at** x, to be the mapping $(f \mapsto f(x)) : F \to C$ for each $x \in D$. Note that ${\mathrm{ev}^F}_x = {\mathrm{ev}^{\mathrm{Map}(D,C)}}_x|_F$; when confusion is unlikely, the notation ev^F is abbreviated to ev. The mapping obtained from an evaluation family ev^F by specifying a suitable subset of $\mathrm{Map}(F, C)$ as codomain is called an **evaluation mapping**.

43. Families of sets

Let $(A_i \mid i \in I)$ be a family of sets, i.e., a family whose terms are sets. We use the notations

$$\bigcup_{i\in I} A_i := \bigcup\{A_i \mid i \in I\} \stackrel{*}{=} \{x \mid \exists i \in I, x \in A_i\}$$

$$\bigcap_{i\in I} A_i := \bigcap\{A_i \mid i \in I\} = \{x \in \bigcup_{i\in I} A_i \mid \forall i \in I, x \in A_i\} \quad \text{when } I \neq \emptyset$$

$$\bigcap_{i\in I}{}^{X} A_i := \bigcap{}^{X}\{A_i \mid i \in I\} = \{x \in X \mid \forall i \in I, x \in A_i\}$$

when $(A_i \mid i \in I)$ is in $\mathfrak{P}(X)$ for a given set X.

(Of course $\bigcap_{i\in I}{}^{X} A_i = \bigcap_{i\in I} A_i$ if $I \neq \emptyset$.)

We note that if $(A_i \mid i \in I)$ and $(B_i \mid i \in I)$ are families of sets such that $A_i \subset B_i$ for all $i \in I$, then: $\bigcup_{i\in I} A_i \subset \bigcup_{i\in I} B_i$; $\bigcap_{i\in I} A_i \subset \bigcap_{i\in I} B_i$ if $I \neq \emptyset$; and $\bigcap_{i\in I}{}^{X} A_i \subset \bigcap_{i\in I}{}^{X} B_i$ if $(B_i \mid i \in I)$ is in $\mathfrak{P}(X)$ for a given set X.

43A. Proposition. *Let the family of sets $(A_i \mid i \in I)$ and the sets B and X be given. The following rules hold:*

$$B \cap \bigcup_{i\in I} A_i = \bigcup_{i\in I}(B \cap A_i) \qquad \text{and} \qquad B \setminus \bigcup_{i\in I} A_i = \bigcap_{i\in I}{}^{B}(B \setminus A_i)$$

$$B \cup \bigcap_{i\in I} A_i = \bigcap_{i\in I}(B \cup A_i) \qquad \text{and} \qquad B \setminus \bigcap_{i\in I} A_i = \bigcup_{i\in I}(B \setminus A_i) \qquad \text{if } I \neq \emptyset$$

$$B \cup \bigcap_{i\in I}{}^{X} A_i = \bigcap_{i\in I}{}^{X}(B \cup A_i) \quad \text{and} \quad B \setminus \bigcap_{i\in I}{}^{X} A_i = \bigcup_{i\in I}(B \setminus A_i)$$

if $(A_i \mid i \in I)$ is in $\mathfrak{P}(X)$ and $B \subset X$.

43B. Proposition. *Let the mapping $f\colon D \to C$ and the families of sets $(U_i \mid i \in I)$ in $\mathfrak{P}(D)$ and $(V_j \mid j \in J)$ in $\mathfrak{P}(C)$ be given. The following rules hold:*

$$f_{>}\Big(\bigcup_{i\in I} U_i\Big) = \bigcup_{i\in I} f_{>}(U_i) \qquad f_{>}\Big(\bigcap_{i\in I}{}^{D} U_i\Big) \subset \bigcap_{i\in I}{}^{C} f_{>}(U_i)$$

$$f^{<}\Big(\bigcup_{j\in J} V_i\Big) = \bigcup_{j\in J} f^{<}(V_j) \qquad f^{<}\Big(\bigcap_{j\in J}{}^{C} V_j\Big) = \bigcap_{j\in J}{}^{D} f^{<}(V_j).$$

Proof. Proposition 23A. ■

For every family of sets $(A_i \mid i \in I)$ we define its **support** by $\operatorname{Supp}(A_i \mid i \in I) := \{i \in I \mid A_i \neq \emptyset\}$, a subset of I. The family $(A_i \mid i \in I)$ is said to be **disjoint** if

$A_j \cap A_k \neq \emptyset$ implies $j = k$ for all $j, k \in I$. A family $(A_i \mid i \in I)$ of *non-empty* sets clearly is disjoint if and only if it is injective and its range $\{A_i \mid i \in I\}$ is a disjoint collection of sets.

Let the set S be given. A family $(A_i \mid i \in I)$ of subsets of S is called a **classification of** S if it is disjoint and $\bigcup_{i\in I} A_i = S$. Classifications of S are related to partitions of S as follows.

43C. Proposition. *Let the set S be given.*

(a): *A family $(A_i \mid i \in I)$ of subsets of S is a classification of S if and only if its restriction $(A_i \mid i \in J)$ to $J := \mathrm{Supp}(A_i \mid i \in I)$ is injective and $\{A_i \mid i \in J\}$ is a partition of S.*

(b): *A collection $\mathcal{C}$ of subsets of S is a partition of S if and only if $\emptyset \notin \mathcal{C}$ and $\mathcal{C}$, self-indexed, is a classification of S.*

It is frequently necessary to determine whether a given family of mappings can be "patched together" to provide a single mapping from which each of the given mappings is obtained by adjustment. The following theorem gives a full answer to this question.

43D. Theorem. *Let the family $(f_i \mid i \in I)$ of mappings be given. Set $D := \bigcup_{i\in I} \mathrm{Dom} f_i$ and $C := \bigcup_{i\in I} \mathrm{Cod} f_i$. The following statements are equivalent:*

(i): *There is exactly one $g \in \mathrm{Map}(D, C)$ such that*

$$\forall i \in I, \qquad g|_{\mathrm{Dom} f_i} = f_i|^C. \tag{43.1}$$

Moreover, g satisfies

$$\mathrm{Gr} g = \bigcup_{i\in I} \mathrm{Gr} f_i. \tag{43.2}$$

(ii): *There is a $g \in \mathrm{Map}(D, C)$ such that* (43.1) *holds.*

(iii): $\forall j, k \in I, \quad f_j|^C_{\mathrm{Dom} f_j \cap \mathrm{Dom} f_k} = f_k|^C_{\mathrm{Dom} f_j \cap \mathrm{Dom} f_k}$ *(i.e., f_j and f_k agree on $\mathrm{Dom} f_j \cap \mathrm{Dom} f_k$).*

(iv): $\bigcup_{i\in I} \mathrm{Gr} f_i$ *is the graph of some mapping from D to C.*

We shall require the following auxiliary result.

43E. Lemma. *With the assumptions as in the statement of Theorem 43D, let $g \in \mathrm{Map}(D, C)$ be given. Then* (43.1) *holds if and only if* (43.2) *holds.*

Proof. (43.1) *implies* (43.2). Assume that (43.1) holds. Let $(x, y) \in \mathrm{Gr} g$ be given. Since $x \in D$, we may choose $j \in I$ such that $x \in \mathrm{Dom} f_j$. By (43.1) we have $y = g(x) = f_j(x)$, and therefore $(x, y) = (x, f_j(x)) \in \mathrm{Gr} f_j \subset \bigcup_{i\in I} \mathrm{Gr} f_i$. We have shown that

$$\mathrm{Gr}g \subset \bigcup_{i\in I} \mathrm{Gr}f_i. \tag{43.3}$$

Let $(x,y) \in \bigcup_{i\in I}\mathrm{Gr}f_i$ be given. We may choose $j \in I$ such that $(x,y) \in \mathrm{Gr}f_j$. Then $x \in \mathrm{Dom}f_j$, and by (43.1) we have $y = f_j(x) = g(x)$. Therefore $(x,y) = (x,g(x)) \in \mathrm{Gr}g$. We have shown that

$$\bigcup_{i\in I}\mathrm{Gr}f_i \subset \mathrm{Gr}g. \tag{43.4}$$

From (43.3) and (43.4) it follows that (43.2) holds.

(43.2) *implies* (43.1). Assume that (43.2) holds. Let $j \in I$ be given. For every $x \in \mathrm{Dom}f_j$ we have, by (43.2), $(x, f_j(x)) \in \mathrm{Gr}f_j \subset \mathrm{Gr}g$, and hence $f_j(x) = g(x)$. Thus $f_j|^C = g|_{\mathrm{Dom}f_j}$. Since $j \in I$ was arbitrary, we conclude that (43.1) holds. ■

Proof of Theorem 43D. (i) *implies* (ii). This is trivial.

(ii) *implies* (iii). By (ii) we may choose $g \in \mathrm{Map}(D,C)$ such that (43.1) holds. Let $j,k \in I$ be given. For every $x \in \mathrm{Dom}f_j \cap \mathrm{Dom}f_k$ we have, by (43.1), $f_j(x) = g(x) = f_k(x)$. Therefore f_j and f_k agree on $\mathrm{Dom}f_k \cap \mathrm{Dom}f_k$.

(iii) *implies* (iv). We define the mapping $F : D \to \mathfrak{P}(C)$ by

$$F(x) := \{y \in C \mid (x,y) \in \bigcup_{i\in I}\mathrm{Gr}f_i\} \qquad \text{for all } x \in D.$$

We assume that (iii) holds and claim that $F(x)$ is a singleton for every $x \in D$; this will establish (iv) (Proposition 22A).

Let $x \in D$ be given. Let $y,z \in F(x)$ be given. We may choose $j,k \in I$ such that $(x,y) \in \mathrm{Gr}f_j$ and $(x,z) \in \mathrm{Gr}f_k$. We find that $x \in \mathrm{Dom}f_j \cap \mathrm{Dom}f_k$ and, by (iii), $y = f_j(x) = f_k(x) = z$. Since $y,z \in F(x)$ were arbitrary, it follows (Proposition 15A) that $F(x)$ is either empty or a singleton.

Since $x \in D$, we may choose $j \in I$ such that $x \in \mathrm{Dom}f_j$. Then $(x, f_j(x)) \in \mathrm{Gr}f_j \subset \bigcup_{i\in I}\mathrm{Gr}f_i$, and therefore $f_j(x) \in F(x)$. We conclude that $F(x)$ is not empty, and hence is a singleton, as claimed.

(iv) *implies* (i). Assume (iv). Then there is exactly one $h \in \mathrm{Map}(D,C)$ such that

$$\mathrm{Gr}h = \bigcup_{i\in I}\mathrm{Gr}f_i \tag{43.5}$$

(Proposition 22A).

Let $g \in \mathrm{Map}(D,C)$ be given, and assume that (43.1) holds. By Lemma 43E, g also satisfies (43.2), and therefore, by (43.5), we must have $g = h$. On the other hand,

if we set $g := h$, then (43.2) holds, by (43.5); by virtue of Lemma 43E, g also satisfies (43.1). We conclude that $g := h$ is the only mapping from D to C that satisfies (43.1), and note that it also satisfies (43.2). ■

44. Products and direct unions

The (**Cartesian**) **product of** a family of sets $(A_i \mid i \in I)$ — called the family of **factors** — is defined to be the set of families

$$\bigtimes_{i\in I} A_i :\overset{*}{=} \{(x_i \mid i \in I) \mid \forall i \in I,\ x_i \in A_i\}.$$

For each $j \in I$ we have the j**th projection**, the mapping $\pi_j : \bigtimes_{i\in I} A_i \to A_j$ defined by the rule

$$\pi_j((x_i \mid i \in I)) := x_j \qquad \text{for each } (x_i \mid i \in I) \in \bigtimes_{i\in I} A_i.$$

Given sets I and A, the product of the family $(A \mid i \in I)$, with all terms equal to A, is denoted by $A^I := \bigtimes_{i\in I} A$. Thus A^I is the set of all families in A with index set I. By providing each such family with the codomain A, we may identify the product A^I with the set $\mathrm{Map}(I, A)$; we shall use this last remark only sparingly.

We note that if $(A_i \mid i \in I)$ and $(B_i \mid i \in I)$ are families of sets such that $A_i \subset B_i$ for all $i \in I$, then $\bigtimes_{i\in I} A_i \subset \bigtimes_{i\in I} B_i$.

Let us look at some special cases of the product of a family of sets $(A_i \mid i \in I)$. If $I = \emptyset$, careful consideration of the definitions shows that $\bigtimes_{i\in\emptyset} A_i = \{\emptyset\}$, where $\emptyset$ stands for the empty set self-indexed; the product of the empty family of sets is thus a singleton. Lists of length 2 (with index set $2^\sqsupset = \{1, 2\}$, say) may be identified with pairs, the 1st term becoming the former component of the pair, and the 2nd term becoming the latter component of the pair. Under this identification, the Cartesian product $\bigtimes_{i\in 2^\sqsupset} A_i$ of the list $(A_i \mid i \in 2^\sqsupset)$ is identified with the product set $A_1 \times A_2$ as defined in Section 17.

When $A_i = \emptyset$ for some $i \in I$, then obviously $\bigtimes_{i\in I} A_i = \emptyset$. The converse assertion, namely that $\bigtimes_{i\in I} A_i \neq \emptyset$ if $A_i \neq \emptyset$ for all $i \in I$, is easily verified when I is empty or a singleton or a doubleton, and can be proved in many other cases; its general validity, however, may be regarded as a matter for stipulation, embodied in the "•Axiom of Choice". We shall regard this converse assertion as valid (subject to later discussion in Chapter 17).

Let the family of sets $(A_i \mid i \in I)$ be given. In analogy with the mappings $(\cdot, b)$ and $(a, \cdot)$ defined in Section 25, we define for each $j \in I$ and $y \in \bigtimes_{i\in I\setminus\{j\}} A_i$ the mapping $(y, \cdot_j) : A_j \to \bigtimes_{i\in I} A_i$ defined by the rule

$$((y, \cdot_j)(z))_i := \begin{cases} y_i & \text{if } i \in I\setminus\{j\} \\ z & \text{if } i = j \end{cases} \qquad \text{for all } z \in A_j. \tag{44.1}$$

If f is a mapping with $\mathrm{Dom} f \subset \bigtimes_{i\in I} A_i$, we write $f(y,\cdot_j) := f\circ(y,\cdot_j)|^{\mathrm{Dom} f}$, so that

$$(f(x|_{I\setminus\{j\}},\cdot_j))(x_j) = f(x) \qquad \text{for all } x\in \mathrm{Dom} f. \tag{44.2}$$

44A. Proposition. *Let the family of sets $(A_i \mid i\in I)$ and the index $j\in I$ be given. The mapping*

$$(x\mapsto (x|_{I\setminus\{j\}},\ x_j)) : \bigtimes_{i\in I} A_i \to (\bigtimes_{i\in I\setminus\{j\}} A_i)\times A_j$$

is bijective.

Proof. The inverse is given by

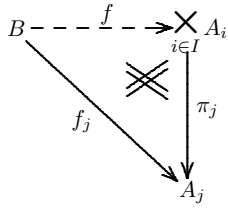
$$((y,z)\mapsto (y,\cdot_j)(z)) : (\bigtimes_{i\in I\setminus\{j\}} A_i)\times A_j \to \bigtimes_{i\in I} A_i. \ \blacksquare$$

44B. Proposition. *If the family of sets $(A_i \mid i\in I)$ satisfies $\bigtimes_{i\in I} A_i \neq \emptyset$, then the projection $\pi_j : \bigtimes_{i\in I} A_i \to A_j$ is surjective for each $j\in I$.*

Proof. It follows from Proposition 44A and the assumption that $\bigtimes_{i\in I\setminus\{j\}} A_i \neq \emptyset$. The mapping $((y,z)\mapsto z) : (\bigtimes_{i\in I\setminus\{j\}} A_i)\times A_j \to A_j$ is surjective (Proposition 32A). The projection π_j is the composite of the bijection of Proposition 44A and this surjection. $\blacksquare$

We now record an obvious but important fact about products.

44C. Proposition. *Let the family of sets $(A_i \mid i\in I)$ and the set B be given. For every family of mappings $(f_i \mid i\in I) \in \bigtimes_{i\in I} \mathrm{Map}(B,A_i)$. There is exactly one mapping $f\colon B\to \bigtimes_{i\in I} A_i$ such that $\pi_j\circ f = f_j$ for all $j\in I$.*



This mapping is given by the rule

$$f(y) := (f_i(y) \mid i\in I) \qquad \text{for all } y\in B.$$

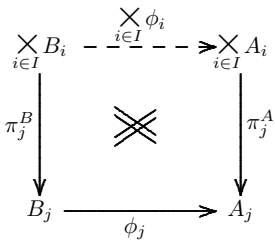
This formula describes a bijection from $\bigtimes_{i\in I} \mathrm{Map}(B,A_i)$ to $\mathrm{Map}(B, \bigtimes_{i\in I} A_i)$.

This bijection may serve to identify the two sets; then $(f_i \mid i\in I)$ is identified with f, and is evaluated termwise:

$$(f_i \mid i \in I)(y) := (f_i(y) \mid i \in I) \qquad \text{for all } y \in B.$$

Let us consider a special case of Proposition 44C. Let the families of sets $(A_i \mid i \in I)$ and $(B_i \mid i \in I)$ with the same index set be given. Let a family $(\phi_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(B_i, A_i)$ be given. We distinguish the projections associated with the families of sets by writing π_i^A, π_i^B. After setting $f_j := \phi_j \circ \pi_j^B : \bigtimes_{i\in I} B_i \to A_j$ for each $j \in I$, we may apply Proposition 44C and find that there is exactly one mapping $f\colon \bigtimes_{i\in I} B_i \to \bigtimes_{i\in I} A_i$ such that $\pi_j^A \circ f = \phi_j \circ \pi_j^B$ for all $j \in I$. This mapping f is denoted by $\bigtimes_{i\in I} \phi_i$, and is called the **product of** the family $(\phi_i \mid i \in I)$. Explicitly, we have

$$(\bigtimes_{i\in I} \phi_i)(y_i \mid i \in I) := (\phi_i(y_i) \mid i \in I) \qquad \text{for all } y \in \bigtimes_{i\in I} B_i,$$



In particular, if sets I, A, B, and a mapping $\phi : B \to A$ are given, we define $\phi^I := \bigtimes_{i\in I} \phi : B^I \to A^I$.

44D. Proposition. (a): *Let the families of sets $(A_i \mid i \in I)$, $(B_i \mid i \in I)$, $(C_i \mid i \in I)$ with the same index set, and the families $(\phi_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(B_i, A_i)$ and $(\psi_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(C_i, B_i)$ be given. Then*

$$\bigtimes_{i\in I} (\phi_i \circ \psi_i) = (\bigtimes_{i\in I} \phi_i) \circ (\bigtimes_{i\in I} \psi_i).$$

(b): *Let the family of sets $(A_i \mid i \in I)$ be given. Then* $\bigtimes_{i\in I} 1_{A_i} = 1_{\bigtimes_{i\in I} A_i}$.

We consider the preceding definitions in the special case in which the families are lists of length 2, identified with pairs (see p. 59). If (A_1, A_2) is a pair of sets and B is a set, the formula

$$f(y) := (f_1(y), f_2(y)) \qquad \text{for all } y \in B$$

describes a bijection $((f_1, f_2) \mapsto f) : \mathrm{Map}(B, A_1) \times \mathrm{Map}(B, A_2) \to \mathrm{Map}(B, A_1 \times A_2)$. This bijection may serve to identify the two sets; then (f_1, f_2) is identified with f, and is evaluated componentwise:

$$(f_1, f_2)(y) := (f_1(y), f_2(y)) \qquad \text{for all } y \in B.$$

If (A_1, A_2) and (B_1, B_2) are pairs of sets, and $(\phi_1, \phi_2) \in \mathrm{Map}(B_1, A_1) \times \mathrm{Map}(B_2, A_2)$, then the **product** $\phi_1 \times \phi_2 \in \mathrm{Map}(B_1 \times B_2, A_1 \times A_2)$ is defined by the rule

$$(\phi_1 \times \phi_2)((y_1, y_2)) := (\phi_1(y_1), \phi_2(y_2)) \qquad \text{for all } (y_1, y_2) \in B_1 \times B_2.$$

Let us look again at the family of sets $(A_i \mid i \in I)$. We should like to construct a set U consisting of all members of each A_i, but in such a way that each member of U "comes from" A_i for exactly one $i \in I$. If the family $(A_i \mid i \in I)$ is disjoint, then $U := \bigcup_{i \in I} A_i$ will fill the bill; if the family is not disjoint, the desired U cannot be constructed at all. As the next best thing, we construct U as the set consisting of "members of A_i with the label i attached, for all $i \in I$". More formally, the **direct union of** the family sets $(A_i \mid i \in I)$ is defined to be the set

$$\dot{\bigcup_{i \in I}} A_i := \{(j, x) \in I \times \bigcup_{i \in I} A_i \mid x \in A_j\}.$$

This set is sometimes called the *direct sum* or the *coproduct* of the family. For each $j \in I$ we have the j**th insertion**, the mapping $\sigma_j : A_j \to \dot{\bigcup_{i \in I}} A_i$ defined by the rule

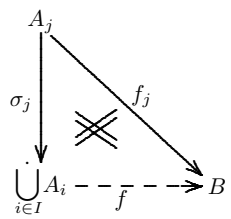
$$\sigma_j(x) := (j, x) \qquad \text{for each } x \in A_j.$$

This mapping is obviously injective. The mapping $((j, x) \mapsto x) : \dot{\bigcup_{i \in I}} A_i \to \bigcup_{i \in I} A_i$, on the other hand, is always surjective, and is bijective if and only if the family $(A_i \mid i \in I)$ is disjoint. (Prove!)

For given sets I and A, the direct union of the family of sets $(A \mid i \in I)$, with all terms equal to A, is $\dot{\bigcup_{i \in I}} A = I \times A$.

Corresponding to Proposition 44C, we have an obvious but important fact about direct unions.

44E. Proposition. *Let the family of sets $(A_i \mid i \in I)$ and the set B be given. For every family of mappings $(f_i \mid i \in I) \in \bigtimes_{i \in I} \mathrm{Map}(A_i, B)$, there is exactly one mapping $f : \dot{\bigcup_{i \in I}} A_i \to B$ such that $f \circ \sigma_j = f_j$ for all $j \in I$.*



This mapping is given by the rule

$$f((j,x)) := f_j(x) \qquad \textit{for all } (j,x) \in \dot{\bigcup_{i\in I}} A_i.$$

This formula describes a bijection from $\bigtimes_{i\in I} \mathrm{Map}(A_i, B)$ *to* $\mathrm{Map}(\dot{\bigcup_{i\in I}} A_i, B)$.

45. General associative and distributive laws

The most important general associative and distributive laws involving unions, intersections, and products of families of sets are given in the following theorem.

•**45A. Theorem**. *Let sets I and X, a family of sets $(J_i \mid i \in I)$, and a family of families of sets $((A_{i,j} \mid j \in J_i) \mid i \in I)$ be given. Set $U := \bigcup_{i\in I} J_i$ and $P := \bigtimes_{i\in I} J_i$. Assume that $\bigcup_{(i,j)\in U} A_{i,j} \subset X$. The following rules hold:*

$$\bigcup_{i\in I}\Big(\bigcup_{j\in J_i} A_{i,j}\Big) = \bigcup_{(i,j)\in U} A_{i,j} \tag{45.1}$$

$$\bigcap_{i\in I}{}^{X}\Big(\bigcap_{j\in J_i}{}^{X} A_{i,j}\Big) = \bigcap_{(i,j)\in U}{}^{X} A_{i,j} \tag{45.2}$$

$$\bullet\ \bigcap_{i\in I}{}^{X}\Big(\bigcup_{j\in J_i} A_{i,j}\Big) = \bigcup_{k\in P}\Big(\bigcap_{i\in I}{}^{X} A_{i,k_i}\Big) \tag{45.3}$$

$$\bullet\ \bigcup_{i\in I}\Big(\bigcap_{j\in J_i}{}^{X} A_{i,j}\Big) = \bigcap_{k\in P}{}^{X}\Big(\bigcup_{i\in I} A_{i,k_i}\Big) \tag{45.4}$$

$$\bullet\ \bigtimes_{i\in I}\Big(\bigcup_{j\in J_i} A_{i,j}\Big) = \bigcup_{k\in P}\Big(\bigtimes_{i\in I} A_{i,k_i}\Big) \tag{45.5}$$

$$\bigtimes_{i\in I}\Big(\bigcap_{j\in J_i} A_{i,j}\Big) = \bigcap_{k\in P}\Big(\bigtimes_{i\in I} A_{i,k_i}\Big) \quad \text{if } P \neq \emptyset. \tag{45.6}$$

Proof. Proof of (45.1). For every $(i', j') \in U$ we have $A_{i',j'} \subset \bigcup_{j\in J_{i'}} A_{i',j} \subset\subset \bigcup_{i\in I}\big(\bigcup_{j\in J_i} A_{i,j}\big)$. Therefore $\bigcup_{(i,j)\in U} A_{i,j} \subset \bigcup_{i\in I}\big(\bigcup_{j\in J_i} A_{i,j}\big)$. Conversely, for every $i' \in I$ and every $j' \in J_{i'}$ we have $(i', j') \in U$ and hence $A_{i',j'} \in \bigcup_{(i,j)\in U} A_{i,j}$; since $j' \in J_{i'}$, was arbitrary, we have $\bigcup_{j\in J_{i'}} A_{i',j} \subset \bigcup_{(i,j)\in U} A_{i,j}$; and since $i' \in I$ was arbitrary, we have $\bigcup_{i\in I}\big(\bigcup_{j\in J_i} A_{i,j}\big) \subset \bigcup_{(i,j)\in U} A_{i,j}$.

Proof of (45.2). This proof is similar to the preceding one and is left to the reader.

•*Proof of* (45.3). Let $k \in P$ be given. Since $A_{i,k_i} \subset \bigcup_{j\in J_i} A_{i,j}$ for each $i \in I$, we have $\bigcap_{i\in I}^X A_{i,k_i} \subset \bigcap_{i\in I}^X (\bigcup_{j\in J_i} A_{i,j})$; since $k \in P$ was arbitrary, we conclude that $\bigcup_{k\in P} (\bigcap_{i\in I}^X A_{i,k_i}) \subset \bigcap_{i\in I}^X (\bigcup_{j\in J_i} A_{i,j})$.

Conversely, let $x \in \bigcap_{i\in I}^X (\bigcup_{j\in J_i} A_{i,j})$ be given. For each $i \in I$, we have $x \in \bigcup_{j\in J_i} A_{i,j}$, and hence $L_i := \{j \in J_i \mid x \in A_{i,j}\} \neq \emptyset$. •Therefore $\bigtimes_{i\in I} L_i \neq \emptyset$. Choose $l \in \bigtimes_{i\in I} L_i \subset P$. Then $x \in A_{i,l_i}$ for each $i \in I$; hence $x \in \bigcap_{i\in I}^X A_{i,l_i} \subset \bigcup_{k\in P} (\bigcap_{i\in I}^X A_{i,k_i})$. We have thus shown that $\bigcap_{i\in I}^X (\bigcup_{j\in J_i} A_{i,j}) \subset \bigcup_{k\in P} (\bigcap_{i\in I}^X A_{i,k_i})$.

•*Proof of* (45.4). Set $B_{i,j} := X \backslash A_{i,j}$ for all $(i,j) \in U$, so that $A_{i,j} = X \backslash B_{i,j}$. Then Proposition 43A and •(45.3) yield

$$\begin{aligned} \bigcup_{i\in I} (\bigcap_{j\in J_i}^X A_{i,j})) &= \bigcup_{i\in I} (\bigcap_{j\in J_i}^X (X \backslash B_{i,j})) = X \backslash (\bigcap_{i\in I}^X (\bigcup_{j\in J_i} B_{i,j})) = \\ &= X \backslash (\bigcup_{k\in P} (\bigcap_{i\in I}^X B_{i,k_i})) = \bigcap_{k\in P}^X (\bigcup_{i\in I} (X \backslash B_{i,k_i}) = \bigcap_{k\in P}^X (\bigcup_{i\in I} A_{i,k_i}). \end{aligned}$$

•*Proof of* (45.5). This proof is similar to the proof of •(45.3), and is left to the reader.

Proof of (45.6). Let $k \in P$ be given. Since $\bigcap_{j\in J_i} A_{i,j} \subset A_{i,k_i}$ for each $i \in I$, we have $\bigtimes_{i\in I} (\bigcap_{j\in J_i} A_{i,j}) \subset \bigtimes_{i\in I} A_{i,k_i}$. Since $k \in P$ was arbitrary, we conclude that $\bigtimes_{i\in I} (\bigcap_{j\in J_i} A_{i,j}) \subset \bigcap_{k\in P} (\bigtimes_{i\in I} A_{i,k_i})$.

Conversely, let $(x_i \mid i \in I) \in (\bigcap_{k\in P} \bigtimes_{i\in I} A_{i,k_i})$ be given and let $i' \in I$ and $j' \in J'_i$ be given. Since $P \neq \emptyset$, the projection $\pi_{i'} : P \to J_{i'}$ is surjective (Proposition 44B). We may therefore choose $l \in P$ such that $l_{i'} = j'$. Then $(x_i \mid i \in I) \in \bigtimes_{i\in I} A_{i,l_i}$, whence $x_{i'} \in A_{i',l_{i'}} = A_{i',j'}$. Since $j' \in J_{i'}$ was arbitrary, we have $x_{i'} \in \bigcap_{j\in J_{i'}} A_{i',j}$. Since $i' \in I$ was arbitrary, $(x_i \mid i \in I) \in \bigtimes_{i\in I} (\bigcap_{j\in J_i} A_{i,j})$. We have shown that $\bigcap_{k\in P} (\bigtimes_{i\in I} A_{i,k_i}) \subset \bigtimes_{i\in I} (\bigcap_{j\in J_i} A_{i,j})$, and thus completed the proof. ■

45B. Proposition. *Let a set I, a family of sets $(J_i \mid i \in I)$, and a family of families of sets $((A_{i,j} \mid j \in J_i) \mid i \in I)$ be given. Set $U := \dot{\bigcup_{i\in I}} J_i$.*

(a): *The mapping*

$$(((a_{i,j} \mid j \in J_i) \mid i \in I) \mapsto (a_{i,j} \mid (i,j) \in U)) : \bigtimes_{i\in I} (\bigtimes_{j\in J_i} A_{i,j}) \to \bigtimes_{(i,j)\in U} A_{i,j}$$

is bijective.

(b): *The mapping*

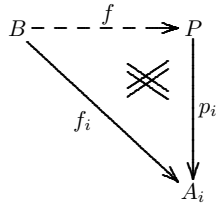
$$(i,(j,x)) \mapsto ((i,j),x) : \dot{\bigcup_{i\in I}} \dot{\bigcup_{j\in J_i}} A_{i,j} \to \dot{\bigcup_{(i,j)\in U}} A_{i,j}$$

is bijective.

46. Set-products and set-coproducts

For later use, we discuss in this section abstract generalizations of the concepts of Cartesian product and direct union of a family of sets. The properties that are the clue to the generalizations were described in Propositions 44C and 44E.

A **(set-)product of** a family of sets $(A_i \mid i \in I)$ is defined to be a set P together with a family of mappings $(p_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(P, A_i)$ such that for every set B and every family of mappings $(f_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(B, A_i)$ there is exactly one mapping $f\colon B \to P$ such that $f_i = p_i \circ f$ for all $i \in I$. The set P is called the **product-set** (sometimes the **product**, when confusion is unlikely), and for each $j \in I$ the mapping $p_j : P \to A_j$ is called the j**th projection**.



The first part of the next proposition asserts that a given family of sets has "essentially" at most one product: any two products "can be identified".

46A. Proposition. (a). *Let products of the family of sets $(A_i \mid i \in I)$ be given, with respective product-sets P and P' and respective families of projections $(p_i \mid i \in I)$ and $(p'_i \mid i \in I)$. Then the unique mappings $g\colon P \to P'$ and $g' : P' \to P$ that satisfy $p_i = p'_i \circ g$ and $p'_i = p_i \circ g'$ for all $i \in I$ are invertible, and each is the inverse of the other.*

(b): *Let a product of the family of sets $(A_i \mid i \in I)$ be given, with product-set P and family of projections $(p_i \mid i \in I)$. A given set Q and family of mappings $(q_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(Q, A_i)$ are the product-set and family of projections of a product of $(A_i \mid i \in I)$ if and only if the unique mapping $g : Q \to P$ that satisfies $q_i = p_i \circ g$ for all $i \in I$ is invertible.*

Proof. Proof of (a). We have $p_i \circ 1_P = p_i = p'_i \circ g = p_i \circ g' \circ g$ for all $i \in I$. By the uniqueness condition in the definition of product (with $B := P$ and $f_i := p_i$ for all $i \in I$), we conclude that $g' \circ g = 1_P$. Interchanging the roles of the products, we conclude that $g \circ g' = 1_{P'}$.

Proof of (b). The "only if" part is an immediate consequence of Part (a). To prove the "if" part, we assume that g is invertible. Let a set B and a family $(f_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(B, A_i)$ be given, and let $f\colon B \to P$ be the unique mapping that satisfies $f_i = p_i \circ f$ for all $i \in I$. A mapping $h\colon B \to Q$ satisfies $f_i = q_i \circ h$ for all $i \in I$ if and only if $f_i = p_i \circ g \circ h$ for all $i \in I$, and this is in turn the case if and only if $g \circ h = f$ or, equivalently, $h = g^{\leftarrow} \circ f$; there is thus exactly one mapping h with the required

property. ■

Has every family of sets a product? The affirmative answer is given by the following proposition.

46B. Proposition. *Let the family of sets* $(A_i \mid i \in I)$ *be given. Then the set* $\bigtimes_{i\in I} A_i$ *together with the family of projections* $(\pi_j \mid j \in I) \in \bigtimes_{j\in I} \mathrm{Map}(\bigtimes_{i\in I} A_i, A_j)$ *is a product of the family* $(A_i \mid i \in I)$.

Proof. Proposition 44C. ■

46C. Remarks. (a): Propositions 46A and 46B together show that the Cartesian product $\bigtimes_{i\in I} A_i$ and the family of projections $(\pi_i \mid i \in I)$ constitute "the essentially unique" product of the family of sets $(A_i \mid i \in I)$. The abstract concept of *set-product* is nevertheless useful. At the same time, the use of the term "product", chosen for the allusion to the Cartesian product, should not lead to confusion. The special product described in Proposition 44B may be called the **Cartesian product**, or the **standard product**, or simply *the* product, **of** $(A_i \mid i \in I)$.

(b): If I and A are given sets, the set $\mathrm{Map}(I, A)$, together with the family of evaluations $(a \mapsto a(i) \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(\mathrm{Map}(I, A), A)$, is a product of the family $(A \mid i \in I)$. If A and B are given sets, the set $A \times B$ together with the pair of mappings $(((x, y) \mapsto x) : A \times B \to A),\ ((x, y) \mapsto y) : A \times B \to B))$ is a product of the pair (A, B) (where pairs are identified with lists of length 2). ■

The following proposition could be proved by using Propositions 46A and 46B, together with our knowledge of the Cartesian product. We prefer to show how it can be obtained directly from the definition of product.

46D. Proposition. *Let a product of a family of sets* $(A_i \mid i \in I)$ *be given, with product-set* P *and family of projections* $(p_i \mid i \in I)$.

(a): $\forall x, y \in P,\ \ x = y \ \Leftrightarrow\ (\forall i \in I,\ p_i(x) = p_i(y))$.

(b): If $P \neq \emptyset$, *then* p_j *is surjective for every* $j \in I$.

Proof. Proof of (a). Let $x, y \in P$ be given, and assume that $p_i(x) = p_i(y)$ for all $i \in I$. Choose a non-empty set B (e.g., a singleton). Then $p_i \circ x_{B\to P} = p_i(x)_{B\to A_i} = p_i(y)_{B\to A_i} = p_i \circ y_{B\to P}$ for all $i \in I$. By the uniqueness condition in the definition of product, we must have $x_{B\to P} = y_{B\to P}$. Since $B \neq \emptyset$, this implies $x = y$. The reverse implication is trivial.

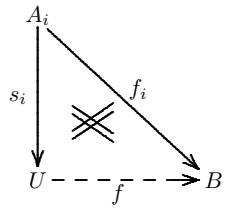
Proof of (b). Let $j \in I$ be given. Choose $x \in P$. Define the family of mappings $(f_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(A_j, A_i)$ by the rule

$$f_i := \begin{cases} 1_{A_j} & \text{if } i = j \\ \\ (x_i)_{A_j\to A_i} & \text{if } i \in I\backslash\{j\}. \end{cases}$$

By the definition of product, there is a mapping $f\colon A_j \to P$ such that $p_j \circ f = f_j = 1_{A_j}$. The p_j is right-invertible, and hence surjective. ■

We now introduce a concept similar to that of set-product, but "with the arrows going in the opposite direction". A **(set-)coproduct of** a family $(A_i \mid i \in I)$ is defined

to be a set U together with a family of mappings $(s_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(A_i, U)$ such that for every set B and every family of mappings $(f_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(A_i, B)$ there is exactly one mapping $f\colon U \to B$ such that $f_i = f \circ s_i$ for all $i \in I$. The set U is called the **coproduct-set** (sometimes the **coproduct**, if confusion is unlikely), and for each $j \in I$ the mapping $s_j : A_j \to U$ is called the j**th insertion**.



46E. Proposition. (a): *Let coproducts of the family of sets $(A_i \mid i \in I)$ be given, with respective coproduct-sets U and U' and respective families of insertions $(s_i \mid i \in I)$ and $(s'_i \mid i \in I)$. Then the unique mappings $g : U' \to U$ and $g' : U \to U'$ that satisfy $s_i = g \circ s'_i$ and $s'_i = g' \circ s_i$ for all $i \in I$ are invertible, and each is the inverse of the other.*

(b): *Let a coproduct of the family of sets $(A_i \mid i \in I)$ be given, with coproduct-set U and family of insertions $(s_i \mid i \in I)$. A given set V and family of mappings $(t_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(A_i, V)$ are the coproduct-set and family of insertions of a coproduct of $(A_i \mid i \in I)$ if and only if the unique mapping $g\colon U \to V$ that satisfies $t_i = g \circ s_i$ for all $i \in I$ is invertible.*

46F. Proposition. (a): *Let the family of sets $(A_i \mid i \in I)$ be given. Then the set $\dot{\bigcup_{i\in I}} A_i$ together with the family of insertions $(\sigma_j \mid j \in I) \in \bigtimes_{j\in I} \mathrm{Map}(A_j, \dot{\bigcup_{i\in I}} A_i)$ is a coproduct of the family $(A_i \mid i \in I)$.*

(b): *Let the disjoint family of sets $(A_i \mid i \in I)$ be given. Then the set $\bigcup_{i\in I} A_i$* together with the family of inclusion mappings

$$(1_{A_j \subset \bigcup_{i\in I} A_i} \mid j \in I) \in \bigtimes_{j\in I} \mathrm{Map}(A_j, \bigcup_{i\in I} A_i)$$

is a coproduct of the family $(A_i \mid i \in I)$.

Proof. Proof of (a). Proposition 44E.

Proof of (b). The mapping $((k, x) \mapsto x) : \dot{\bigcup_{i\in I}} A_i \to \bigcup_{i\in I} A_i$ is bijective when the family $(A_i \mid i \in I)$ is disjoint. Now the composite of σ_j with this mapping is precisely the inclusion mapping of A_j into $\bigcup_{i\in I} A_i$ for every $j \in I$. The conclusion follows from Part (a) and Proposition 46E,(b). ■

The coproduct described in Proposition 46F, (a) may be called the **standard coproduct of** the family of sets $(A_i \mid i \in I)$.

46G. Proposition. *Let a coproduct of a family* $(A_i \mid i \in I)$ *be given, with coproduct-set* U *and family of insertions* $(s_i \mid i \in I)$. *Then:*

(a): $U = \bigcup_{i\in I} \mathrm{Rng} s_i$.

(b): s_j *is injective for every* $j \in I$.

Proof. Proof of (a). Choose a set B that is neither empty nor a singleton (e.g., a doubleton), and choose $b, b' \in B$ such that $b \neq b'$. Define $f, f' : U \to B$ by $f := b_{U\to B}$ and by the rule

$$f'(x) := \begin{cases} b & \text{if } x \in \bigcup_{i\in I} \mathrm{Rng} s_i \\ \\ b' & \text{if } x \in U \backslash \bigcup_{i\in I} \mathrm{Rng} s_i. \end{cases}$$

Then $f \circ s_i = b_{A_i\to B} = f' \circ s_i$ for all $i \in I$. By the uniqueness condition in the definition of coproduct, we must have $f = f'$, and therefore $U = \bigcup_{i\in I} \mathrm{Rng} s_i$.

Proof of (b). Let $j \in I$ be given. If $A_j = \emptyset$, then $s_j = 1_{\emptyset\subset U}$ is injective. Assume now that $A_j \neq \emptyset$, and choose $y \in A_j$. Define the family of mappings $(f_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(A_i, A_j)$ by the rule

$$f_i := \begin{cases} 1_{A_j} & \text{if } i = j \\ \\ y_{A_i\to A_j} & \text{if } i \in I \backslash \{j\}. \end{cases}$$

By the definition of coproduct, there is a mapping $f\colon U \to A_j$ such that $f \circ s_j = f_j = 1_{A_j}$. Thus s_j is left-invertible, and hence injective. ■

Chapter 5

RELATIONS

51. Relations in a set

Consider the statement "6 is a multiple of 3". It is clear that if we replace "6" and "3" by any (names for) natural numbers we again obtain a meaningful statement. Thus "m is a multiple of n" is meaningful whenever m and n are members of the set $\mathbb{N}$ of natural numbers. Of course, whether "m is a multiple of n" is a true statement depends on what numbers m and n actually are. The sentence fragment "is a multiple of" determines a *relation* in the set $\mathbb{N}$ of all natural numbers. Similarly, the sentence fragments "does not exceed" and "is less than or equal to", both usually abbreviated "$\leq$", determine one and the same relation in the set $\mathbb{R}$ of all real numbers. The fragment becomes a complete assertion, or statement, when preceded and followed by names of real numbers. For instance, "Two does not exceed one", or "$2 \leq 1$", is a complete, meaningful statement; it happens to be false.

In general, a **relation** ρ **in** a set D is determined by a two-place predicate in which both places can be filled only with members of D. If $x, y \in D$ fill these places, respectively, the resulting assertion is usually written $x\ \rho\ y$, with the symbol for the relation between the symbols for x and y. The set D is called the **domain of** ρ, and is denoted by $\mathrm{Dom}\rho := D$.

Predicates that yield equivalent assertions when filled with the same pair of members of the same set are regarded as determining the *same* relation. Thus, if ρ and σ are relations, we have $\rho = \sigma$ if and only if

$$D := \mathrm{Dom}\rho = \mathrm{Dom}\sigma \qquad \text{and} \qquad \forall x, y \in D, \quad x\ \rho\ y\ \Leftrightarrow\ x\ \sigma\ y.$$

With each relation ρ in the set D we associate a subset $\mathrm{Gr}\rho$, called the **graph of** ρ, of the product set $D \times D$, as follows:

$$\mathrm{Gr}\rho := \{(x, y) \in D \times D \mid x\ \rho\ y\}.$$

Conversely, it is clear that every subset G of $D \times D$ is the graph of exactly one relation in D: namely, the relation ρ defined by the rule

$$\forall x, y \in D, \quad x \ \rho \ y \ :\Leftrightarrow \ (x, y) \in G.$$

This remark establishes a one-to-one correspondence between all relations in the set D and all subsets of $D \times D$. In view of this correspondence, many mathematicians say that a relation in D *is* a subset of $D \times D$, and make no distinction between a relation and its graph.

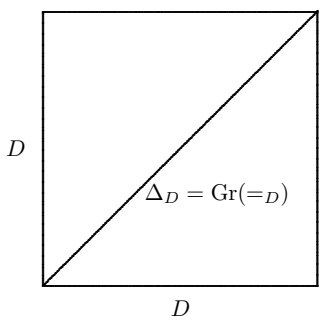
A trivial but important relation in D is **equality in** D, which may be written $=_D$; this is the relation defined by the rule

$$\forall x, y \in D, \quad x =_D y \ :\Leftrightarrow \ x = y,$$

so that $x =_D y$ for given $x, y \in D$ if and only if x and y are actually the same member of D. The graph of $=_D$ is the diagonal of $D \times D$:

$$\mathrm{Gr}(=_D) = \triangle_D = \{(x, x) \mid x \in D\}.$$

If D is an interval and we represent $D \times D$ as a square in a plane, this graph is then precisely a diagonal of this square.



52. Images and pre-images

Let the relation ρ in the set D be given. For every subset U of D, we define the **image of** U **under** ρ to be the set

$$\rho_>(U) := \{y \in D \mid x\ \rho\ y \text{ for some } x \in U\}. \tag{52.1}$$

The rule (52.1) defines a mapping $\rho_> : \mathfrak{P}(D) \to \mathfrak{P}(D)$, called the **image mapping induced by** ρ.

The mapping $(x \mapsto \rho_>(\{x\})) : D \to \mathfrak{P}(D)$ is of some independent interest: it assigns to each member of D the set of all members of D to which it is related by ρ. This mapping may serve to specify the relation, in the following sense. To every mapping $\phi : D \to \mathfrak{P}(D)$ there is exactly one relation ρ in D such that $\rho_>(\{x\}) = \phi(x)$ for all $x \in D$: namely, the relation ρ given by the rule

$$\forall x, y \in D, \quad x\ \rho\ y\ :\Leftrightarrow\ y \in \phi(x).$$

This remark establishes a one-to-one correspondence between all relations in the set D and all mappings from D to $\mathfrak{P}(D)$. In view of this correspondence, one might say that a relation in D *is* a mapping from D to $\mathfrak{P}(D)$.

Let us return to the given relation ρ in D. For every subset V of D, we define the **pre-image of** V **under** ρ to be the set

$$\rho^<(V) := \{x \in D \mid x\ \rho\ y \text{ for some } y \in V\}. \tag{52.2}$$

The rule (52.2) defines a mapping $\rho^< : \mathfrak{P}(D) \to \mathfrak{P}(D)$, called the **pre-image mapping induced by** ρ.

53. Reversal, composition, and restriction of relations

Let the relation ρ in the set D be given. We define a new relation $\rho^{\leftarrow}$ in D, called the **reverse of** ρ, by the rule

$$\forall x, y \in D, \quad x \ \rho^{\leftarrow} \ y \ :\Leftrightarrow \ y \ \rho \ x.$$

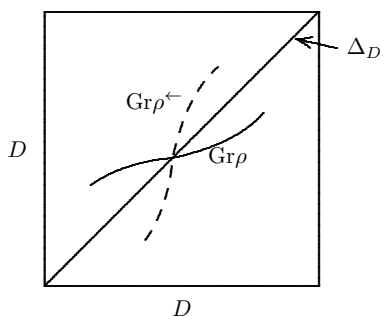
For example, the relation $\geq$ in $\mathbb{R}$ is the reverse of $\leq$; the relation "divides" in $\mathbb{N}$ is the reverse of "is a multiple of". This operation of reversal obviously satisfies the following rules:

$$\rho^{\leftarrow\leftarrow} = \rho, \tag{53.1}$$

$$Gr\rho^{\leftarrow} = (Gr\rho)^{\top} \tag{53.2}$$

$$(\rho^{\leftarrow})_{>} = \rho^{<}, \qquad (\rho^{\leftarrow})^{<} = \rho_{>}. \tag{53.3}$$

If D is an interval and $D \times D$ is represented as a square in a plane, the mapping $((x, y) \mapsto (y, x)) : D \times D \to D \times D$ corresponds to reflection with respect to the diagonal $\triangle_D$. This remark allows us to visualize the relationship (53.2) between the graph of a relation and the graph of its inverse.



Let the relations σ and ρ in D be given. We define a new relation $\sigma \circ \rho$ in D, called the **composite of** ρ **and** σ (notice the reversal of priority) by the rule

$$\forall x, y \in D, \quad x(\sigma \circ \rho)y \ :\Leftrightarrow \ (\exists z \in D, \quad x \ \rho \ z \quad \text{and} \quad z \ \sigma \ y).$$

This operation of **composition** obviously satisfies the following rules

$$(\sigma \circ \rho)_{>} = \sigma_{>} \circ \rho_{>} \qquad (\sigma \circ \rho)^{<} = \rho^{<} \circ \sigma^{<}, \tag{53.4}$$

$$(\sigma \circ \rho)^{\leftarrow} = \rho^{\leftarrow} \circ \sigma^{\leftarrow}, \tag{53.5}$$

and the composite relation $\sigma \circ \rho$ can be characterized by the rule

$$\forall x \in D, \quad (\sigma \circ \rho)_{>}(\{x\}) = \sigma_{>}(\rho_{>}(\{x\})). \tag{53.6}$$

(In (53.4), the right-hand sides are composites of *mappings*. The use of the same symbol for composition of relations and composition of mappings should be a help rather than a source of confusion.)

Composition of relations obeys the associative law in the following sense: if ρ, σ, τ are relations in D, then

$$(\tau \circ \sigma) \circ \rho = \tau \circ (\sigma \circ \rho).$$

This result follows at once from the definition of composition, or from (53.6). It permits us to write $\tau \circ \sigma \circ \rho$ without parentheses, and similarly for composites of more relations.

If ρ is a relation in the set D, and U is a subset of D, we define the relation $\rho|_U$ in U, called the **restriction of ρ to U**, by the rule

$$\forall x, y \in U, \quad x(\rho|_U)y \ :\Leftrightarrow \ x \ \rho \ y.$$

This restriction is characterized by either of the rules

$$\mathrm{Gr}(\rho|_U) = \mathrm{Gr}\rho \cap (U \times U)$$

$$\forall x \in U, \quad (\rho|_U)_{>}(\{x\}) = \rho_{>}(\{x\}) \cap U.$$

54. Relations from set to set; functional relations

This section constitutes a digression from the topic of this chapter, and may be skipped for the time being without loss of continuity.

We sometimes encounter a concept somewhat more general than that of a relation in a set, described in Section 51, in that the objects that fill the places in the two-place predicate are to be taken from possibly different sets. For instance, in the space of ordinary Euclidean geometry, we may consider assertions such as "the plane P is perpendicular to the line l". Here the places in the predicate "is perpendicular to" are to be filled with (the name of) a plane and (the name of) a line, respectively.

We therefore consider the concept of a **relation ρ from** the set D **to** the set C, in which the places in the predicate determining the relation are to be filled with a member of D and a member of C, respectively. D is called the **domain of** ρ, and C the **codomain of** ρ. A relation from D to D is then a relation in D as described in Section 51.

All the concepts introduced in Sections 51, 52, and 53 have their obvious generalizations to relations form set to set, and obey the same rules. Thus, if ρ is a relation from D to C, the **graph** $\mathrm{Gr}\rho$ is a subset of $D \times C$, the **induced mappings** are $\rho_> : \mathfrak{P}(D) \to \mathfrak{P}(C)$ and $\rho^< : \mathfrak{P}(C) \to \mathfrak{P}(D)$, and reversal of ρ produces the **reverse** $\rho^{\leftarrow}$, a relation from C to D.

One can compose relations ρ and σ to form the **composite** $\sigma \circ \rho$ only when the domain of σ is the codomain of ρ.

Let the mapping $f\colon D \to C$ be given. To this mapping we associate the relation $\overset{f}{\mapsto}$ from D to C defined by the rule

$$\forall (x, y) \in D \times C, \quad x \overset{f}{\mapsto} y \ :\Leftrightarrow \ y = f(x).$$

We note the following obvious rules satisfied by the relation $\overset{f}{\mapsto}$:

$$\mathrm{Gr}\overset{f}{\mapsto} = \mathrm{Gr}f, \tag{54.1}$$

$$(\overset{f}{\mapsto})_> = f_> \qquad (\overset{f}{\mapsto})^< = f^<. \tag{54.2}$$

If the mappings $f\colon D \to C$ and $g\colon C \to B$ are given, then

$$\overset{g}{\mapsto} \circ \overset{f}{\mapsto} = \overset{g \circ f}{\longmapsto}. \tag{54.3}$$

We observe that, for every set D, the equality relation $=_D$ in D defined in Section 51 is precisely $\overset{1_D}{\longmapsto}$.

A relation ρ from D to C is said to be **functional** if $\rho = \overset{f}{\mapsto}$ for some mapping $f\colon D \to C$. It is easy to see that ρ is functional if and only if $\rho_>(\{x\})$ is a singleton for every $x \in D$ (see also (54.1) in conjunction with Proposition 22A for a characterization in terms of the graph of ρ). If $f\colon D \to C$ is a mapping, the relation $(\overset{f}{\mapsto})^\leftarrow$ from C to D is functional if and only if f is invertible; in that case, $(\overset{f}{\mapsto})^\leftarrow = \overset{f^\leftarrow}{\longmapsto}$.

55. Properties of relations

We now return to the study of relations in a set. We introduce several important properties that a relation may have, and then list for each of them some obvious equivalent variants of the corresponding defining condition. Since the formulation of these conditions is rather stereotyped, we present them in a tabular form.

The relation ρ in the set D is said to be if

reflexive	$\forall x \in D,$	$x \ \rho \ x$
irreflexive	$\forall x \in D,$	not $x \ \rho \ x$
symmetric	$\forall x, y \in D,$	$x \ \rho \ y \Rightarrow y \ \rho \ x$
antisymmetric	$\forall x, y \in D,$	$(x \ \rho \ y$ and $y \ \rho \ x) \Rightarrow x = y$
strictly antisymmetric	$\forall x, y \in D,$	$x \ \rho \ y \Rightarrow$ (not $y \ \rho \ x$)
total	$\forall x, y \in D,$	$x \ \rho \ y$ or $y \ \rho \ x$ or $x = y$
transitive	$\forall x, y, z \in D,$	$(x \ \rho \ y$ and $y \ \rho \ z) \Rightarrow x \ \rho \ z.$

55A. Proposition. *Let the relation ρ in the set D be given. Then*

ρ is	*if and only if*	*and if and only if, for all $x \in D$,*
reflexive	$\triangle_D \subset \mathrm{Gr}\rho$	$x \in \rho_>(\{x\})$
irreflexive	$\mathrm{Gr}\rho \cap \triangle_D = \emptyset$	$x \in D \backslash \rho_>(\{x\})$
symmetric	$\rho^{\leftarrow} = \rho$	$\rho_>(\{x\}) = \rho^<(\{x\})$
antisymmetric	$\mathrm{Gr}\rho \cap \mathrm{Gr}\rho^{\leftarrow} \subset \triangle_D$	$\rho_>(\{x\}) \cap \rho^<(\{x\}) \subset \{x\}$
strictly antisymmetric	$\mathrm{Gr}\rho \cap \mathrm{Gr}\rho^{\leftarrow} = \emptyset$	$\rho_>(\{x\}) \cap \rho^<(\{x\}) = \emptyset$
total	$\mathrm{Gr}\rho \cup \mathrm{Gr}\rho^{\leftarrow} \cup \triangle_D = D \times D$	$\rho_>(\{x\}) \cup \rho^<(\{x\}) \cup \{x\} = D$
transitive	$\mathrm{Gr}(\rho \circ \rho) \subset \mathrm{Gr}\rho$	$\rho_>(\rho_>(\{x\})) \subset \rho_>(\{x\}).$

55B. Remarks. Let the relation ρ in the set D be given. Then:

(a): ρ is strictly antisymmetric if and only if ρ is antisymmetric and irreflexive.

(b): ρ is not both reflexive and irreflexive unless $D = \emptyset$.

(c): ρ is both symmetric and antisymmetric if and only if $\mathrm{Gr}\rho \subset \triangle_D$.

(d): ρ is not reflexive, symmetric, and antisymmetric unless ρ is the equality relation $=_D$.

(e): ρ is not both symmetric and strictly antisymmetric unless $\mathrm{Gr}\rho = \emptyset$.

(f): ρ is not reflexive, symmetric, and total unless $\mathrm{Gr}\rho = D \times D$.

(g): if ρ is transitive, then ρ is strictly antisymmetric if and only if ρ is irreflexive. ■

55C. Proposition. *Let the relation ρ in the set D and the subset U of D be given. If ρ is reflexive, then $\rho^{\leftarrow}$ and $\rho|_U$ are reflexive. The same implication holds with "reflexive" replaced by "irreflexive", "symmetric", "antisymmetric", "strictly antisymmetric", "total", or "transitive".*

55D. Examples. (a): The equality relation $=_D$ in a set D is reflexive, symmetric, antisymmetric, and transitive; it is neither irreflexive nor strictly antisymmetric unless $D = \emptyset$, nor is it total unless D is empty or a singleton.

(b)*: The relations $\leq$ and $<$ in $\mathbb{R}$ are antisymmetric, total, and transitive. The former is reflexive, hence neither irreflexive nor strictly antisymmetric; the latter is irreflexive and strictly antisymmetric, hence not reflexive. Neither relation is symmetric.

(c)*: The relation "divides" in $\mathbb{N}$ is reflexive, antisymmetric, and transitive; it is neither symmetric nor irreflexive nor strictly antisymmetric; moreover, it is not total (since "2 divides 5, or 5 divides 2, or $2 = 5$" is a false statement). The analogous relation "divides" in the set $\mathbb{Z}$ of all integers has the same properties, except that it is not antisymmetric (since -2 divides 2, and 2 divides -2, but $2 \neq -2$).

(d)*: In a model of a human population, the relation "is a child of" is irreflexive, indeed strictly antisymmetric; it is neither reflexive, symmetric, total, nor transitive (except for trivially restricted populations). The same assertions are valid for the relation "is a descendant of", except that this relation is transitive; it is a kind of "transitive closure" of the former relation. (The concept of the *transitive closure* of a relation will be examined in Section 73.) ■

56. Order

Among the properties of relations introduced in the preceding section there are some particularly important combinations. We discuss these in this section and the next.

A relation in a set is called an **order** if it is transitive, antisymmetric, and reflexive; it is called a **strict-order** if it is transitive and irreflexive, and hence also strictly antisymmetric (Remarks 55B,(g)). If the distinction from strict-orders is to be stressed, an order may be called a **lax order**. It follows from Proposition 55C that the reverse and each restriction of an order is again an order; an analogous assertion follows for strict-orders. It is customary to denote an order or a strict-order by $\prec$, $<$, $\lhd$, $\sqsubset$, or another symbol of a similar shape. If such a symbol is used, its left-to-right mirror image is used to denote the reverse relation.

There is an obvious correspondence between orders and strict-orders in a given set, exemplified by the relationship between the order $\leq$ and the strict-order $<$ in $\mathbb{R}$ (Examples 55D,(b)). These relations differ only by the inclusion or exclusion, respectively, of equality; to put it another way, by the inclusion or exclusion, respectively, of the diagonal in or from the graph. More formally, the correspondence is described by the following result.

56A. Proposition. *Let D be a given set. Then*:

(a): *For each order $\prec$ in D, define the relation $\precneqq$ in D by the rule*

$$\forall x, y \in D, \quad x \precneqq y \ :\Leftrightarrow \ (x \prec y \text{ and } x \neq y)$$

(i.e., $\mathrm{Gr}(\precneqq) := \mathrm{Gr}(\prec) \backslash \triangle_D$*). Then $\precneqq$ is a strict order.*

(b): *for each strict-order $\prec$ in D, define the relation $\preceq$ in D by the rule*

$$\forall x, y \in D, \quad x \preceq y \ :\Leftrightarrow \ (x \prec y \text{ or } x = y)$$

i.e., $\mathrm{Gr}(\preceq) := \mathrm{Gr}(\prec) \cup \triangle_D$*). Then $\preceq$ is an order.*

(c): *The mapping $\prec \mapsto \precneqq$ from the set of all orders in D to the set of all strict-orders in D, and the mapping $\prec \mapsto \preceq$ from the set of all strict-orders in D to the set of all orders in D, are inverses of each other.*

(d): *If $\prec$ is an order in D, then $\prec$ is total if and only if $\precneqq$ is total, and this is the case if and only if*

$$\forall x, y \in D, \quad x \precneqq y \ \Leftrightarrow \ (\textit{not } x \succ y).$$

Proposition 56A allows us to use the expressions "an order and the *corresponding* strict-order", "a strict-order and the *corresponding* (lax) order" unambiguously. Note that this correspondence is preserved under reversal and restriction. When dealing with this kind of relation, it is often useful to have both the order and the corresponding strict-order available, and it is a matter of expediency whether one considers the one or the other as primary. It is usually more convenient to work with the (lax) order.

56B. Examples. (a)*: The relations $\leq$ and $<$ in $\mathbb{R}$ (as well as their respective restrictions to, say $\mathbb{N}$) are an order and the corresponding strict-order, respectively. They are total.

(b)*: The relation "divides" in $\mathbb{N}$ (usually abbreviated to "|") is an order; it is not total (Example 55D,(c)).

(c)*: In the example of a model of a human population (Example 55D,(d)), the relation "is a descendant of" is a strict-order; it is usually not total.

(d): The most important example of an order is the **inclusion relation** $\subset_S$ in the power-set $\mathfrak{P}(S)$ of a given set S, defined by the rule

$$\forall U, V \in \mathfrak{P}(S), \quad U \subset_S V \ :\Leftrightarrow\ U \subset V.$$

The corresponding strict-order is the **proper-inclusion relation** $\subsetneq_S$. These relations are not total, unless S is empty or a singleton.

(e): Let the set D be given. If ρ and σ are relations in D, then ρ is said to be **narrower than** σ, and σ is said to be **broader than** ρ, if

$$\forall x, y \in D, \quad x\ \rho\ y \ \ \Rightarrow\ \ x\ \sigma\ y.$$

This condition is obviously equivalent to $\mathrm{Gr}\rho \subset \mathrm{Gr}\sigma$. Since the inclusion relation $\subset_{D\times D}$ is an order in $\mathfrak{P}(D \times D)$ (see (d)), it follows that the relation "is narrower than" is an order in the set of all relations in D, and "is broader than" is its reverse.■

As noted in Proposition 56A,(d), an order is total if and only if the corresponding strict-order is total. If $\prec$ is an order or a strict-order in a set D, then the members $x, y \in D$ are said to be $\prec$-**comparable** if either $x \prec y$ or $y \prec x$ or $x = y$. Thus the order or strict-order $\prec$ is total if and only if x and y are $\prec$-comparable for all $x, y \in D$. Some mathematicians prefer to use the term *partial order* instead of *order*, to indicate that there may be non-comparable pairs of members of the domain. In the older literature, the term *order* is often used to mean what we have called *total order*. Our usage reflects the increasingly prevalent view that (partial) order is a more fundamental concept than total order.

A set with a specified order in it is an important mathematical object in its own right; we shall devote three entire chapters to *ordered sets*.

57. Equivalence relations

Let the set D and a mapping f with $\mathrm{Dom} f = D$ be given. Define the relation ρ in D by the rule

$$\forall x, y \in D, \quad x \ \rho \ y \ :\Leftrightarrow \ f(x) = f(y).$$

It is an immediate consequence of the logical properties of equality that ρ is reflexive, symmetric, and transitive. We shall see later that these properties are characteristic, i.e., that every reflexive, symmetric, and transitive relation ρ in D is obtained in this manner from a suitable mapping f with $\mathrm{Dom} f = D$ (Corollary 57G). For this reason, we call such relations *equivalence relations* ("equivalent" means "having the same value").

Thus, a relation in a set is called an **equivalence relation** if it is reflexive, symmetric, and transitive. In addition to the examples that may be obtained by the process described above, we mention a few others.

57A. Examples. (a): In every set D, the equality relation $=_D$ and the trivial relation whose graph is $D \times D$ are equivalence relations.

(b)*: If D, C are given sets, the relation $\sim$ in $\mathrm{Map}(D, C)$ defined by the rule

$$\forall f, g \in \mathrm{Map}(D, C), \quad f \sim g \ :\Leftrightarrow \ \{x \in D \mid f(x) \neq g(x)\} \text{ is a finite set}$$

is an equivalence relation. (To prove this, we use the facts that $\emptyset$ is a finite set and that every subset of the union of two finite sets is a finite set.)

(c)*: Define the relation $\sim$ in $\mathbb{N} \times \mathbb{N}$ by the rule

$$\forall (m, n), (m', n') \in \mathbb{N} \times \mathbb{N}, \quad (m, n) \sim (m', n') \ :\Leftrightarrow \ m + n' = m' + n.$$

Then $\sim$ is an equivalence relation.

(d)*: Let $m \in \mathbb{Z}$ be given, and define the relation $\equiv_m$ in $\mathbb{Z}$ by the rule

$$\forall n, n' \in \mathbb{Z}, \quad n \equiv_m n' \ :\Leftrightarrow \ m \text{ divides } n' - n$$

(the definiendum is usually written "$n \equiv n' (\mathrm{mod}\ m)$"). Then $\equiv_m$, which is called *congruence modulo* m, is an equivalence relation. ■

In characterizing equivalence relations, it is sometimes convenient to combine the definitions of symmetry and transitivity into a single formula resembling the definition of transitivity alone.

57B. Proposition. *Let the reflexive relation ρ in the set D be given. The following statements are equivalent:*

(i): *ρ is an equivalence relation.*

(ii): $\forall x, y, z \in D, \ \ (x \ \rho \ y \text{ and } y \ \rho \ z) \ \Rightarrow \ z \ \rho \ x.$

(iii): $\forall x, y, z \in D, \ \ (x \ \rho \ y \text{ and } x \ \rho \ z) \ \Rightarrow \ y \ \rho \ z.$

(iv): $\forall x, y, z \in D, \ \ (x \ \rho \ z \text{ and } y \ \rho \ z) \ \Rightarrow \ x \ \rho \ y.$

The central facts about equivalence relations have to do with their relationship to partitions and mappings, as described by the following results.

57C. Theorem. *Let the equivalence relation ρ in the set D be given. Then the collection $\{\rho_>(\{x\}) \mid x \in D\}$ is a partition of D.*

Proof. Set $\mathcal{R} := \{\rho_>(\{x\}) \mid x \in D\}$. For each $x \in D$ we have $x \in \rho_>(\{x\})$, since ρ is reflexive. Therefore $\bigcup \mathcal{R} = D$ and $\emptyset \notin \mathcal{R}$. It remains to show that the collection $\mathcal{R}$ is disjoint.

Let $x, y \in D$ be such that $\rho_>(\{x\}) \cap \rho_>(\{y\}) \neq \emptyset$; choose $z \in \rho_>(\{x\}) \cap \rho_>(\{y\})$. Let $u \in \rho_>(\{x\})$ be given. We then have $x \ \rho \ u$, $x \ \rho \ z$, $y \ \rho \ z$. By Proposition 57B we have $y \ \rho \ u$, i.e., $u \in \rho_>(\{y\})$. Since $u \in \rho_>(\{x\})$ was arbitrary, it follows that $\rho_>(\{x\}) \subset \rho_>(\{y\})$. Interchanging x, y in the preceding argument, we conclude that the reverse inclusion holds, so that $\rho_>(\{x\}) = \rho_>(\{y\})$. ■

In the following theorem, $\mathcal{P}$ is a partition of D, and $\Omega_\mathcal{P} : D \to \mathcal{P}$ is the corresponding *partition mapping* (Section 24).

57D. Theorem. *Let the set D, the relation ρ in D, and the partition $\mathcal{P}$ of D be given. The following statements are equivalent:*

(i): *ρ is an equivalence relation, and $\mathcal{P} = \{\rho_>(\{x\}) \mid x \in D\}$.*

(ii) *ρ is reflexive, and $\mathcal{P} = \{\rho_>(\{x\}) \mid x \in D\}$.*

(iii): *$\Omega_\mathcal{P}(x) = \rho_>(\{x\})$ for all $x \in D$.*

(iv): $\forall x, y \in D, \quad x \ \rho \ y \ \Leftrightarrow \ \Omega_\mathcal{P}(x) = \Omega_\mathcal{P}(y)$.

(v): $\forall x, y \in D, \quad x \ \rho \ y \ \Leftrightarrow \ (\exists E \in \mathcal{P}, \quad x, y \in E)$.

(vi): $\mathrm{Gr}\rho = \bigcup_{E \in \mathcal{P}} (E \times E)$.

Proof. (i) ⇒ (ii). This implication is trivial.

(ii) ⇒ (iii). Assume that (ii) holds. For given $x \in D$ we have $x \in \Omega_\mathcal{P}(x) \in \mathcal{P}$ and $x \in \rho_>(\{x\}) \in \mathcal{P}$, the former by the definition of $\Omega_\mathcal{P}$, the latter by (ii). Thus $\Omega_\mathcal{P}(x), \rho_>(\{x\}) \in \mathcal{P}$ and $\Omega_\mathcal{P}(x) \cap \rho_>(\{x\}) \neq \emptyset$, whence $\Omega_\mathcal{P}(x) = \rho_>(\{x\})$.

(iii) ⇒ (iv). For all $x, y \in D$, we have $x \ \rho \ y$ if and only if $y \in \rho_>(\{x\})$, and hence, when (iii) holds, if and only if $y \in \Omega_\mathcal{P}(x)$, which is equivalent to $\Omega_\mathcal{P}(x) = \Omega_\mathcal{P}(y)$.

(iv) ⇒ (v). This implication is trivial, upon setting $E := \Omega_\mathcal{P}(x)$.

(v) ⇒ (vi). Assume that (v) holds. For given $x, y \in D$ we have the chain of equivalences

$$(x, y) \in \mathrm{Gr}\rho \ \Leftrightarrow \ x \ \rho \ y \ \Leftrightarrow \ (\exists E \in \mathcal{P}, \quad x, y \in E) \ \Leftrightarrow$$

$$\Leftrightarrow \ (\exists E \in \mathcal{P}, \quad (x, y) \in E \times E) \ \Leftrightarrow \ (x, y) \in \bigcup_{E \in \mathcal{P}} (E \times E).$$

(vi) ⇒ (iii). Assume that (vi) holds. For every $x, y \in D$ we have the chain of equivalences

$$y \in \rho_>(\{x\}) \ \Leftrightarrow \ x \ \rho \ y \ \Leftrightarrow \ (x, y) \in \mathrm{Gr}\rho \ \Leftrightarrow \ (x, y) \in \bigcup_{E \in \mathcal{P}} (E \times E) \ \Leftrightarrow$$

$$\Leftrightarrow \ (\exists E \in \mathcal{P}, \quad x, y \in E) \ \Leftrightarrow \ y \in \Omega_\mathcal{P}(x).$$

Since $y \in D$ was arbitrary, we conclude that $\rho_>(\{x\}) = \Omega_{\mathcal{P}}(x)$ for all $x \in D$.

(iii) $\Rightarrow$ (i). Assume that (iii) holds. Then $\{\rho_>(\{x\}) \mid x \in D\} = \{\Omega_{\mathcal{P}}(x) \mid x \in D\} = \mathcal{P}$. For all $x, y \in D$, we have the chain of equivalences

$$x \ \rho \ y \ \Leftrightarrow \ y \in \rho_>(\{x\}) \ \Leftrightarrow \ y \in \Omega_{\mathcal{P}}(x) \ \Leftrightarrow \ \Omega_{\mathcal{P}}(x) = \Omega_{\mathcal{P}}(y).$$

As we pointed out at the beginning of this section, ρ is then reflexive, symmetric, and transitive. ■

57E. Corollary. *Let the relation ρ in the set D be given. Then ρ is an equivalence relation if and only if it is reflexive and $\{\rho_>(\{x\}) \mid x \in D\}$ is a partition of D.*

Proof. The "only if" part follows form theorem 57C. The "if" part follows from Theorem 57D, ((ii) $\Rightarrow$ (i)). ■

57F. Theorem. *Let the set D, the relation ρ in D, and the mapping f with* $\mathrm{Dom} f = D$ *be given. The following statements are equivalent:*

(i): *ρ is an equivalence relation, and* $\mathrm{Part} f = \{\rho_>(\{x\}) \mid x \in D\}$.

(ii): $\rho_> = f^< \circ f_>$.

(iii): $\forall x, y \in D, \ \ x \ \rho \ y \ \Leftrightarrow \ f(x) = f(y)$.

Proof. (i) $\Rightarrow$ (ii). Assume that (i) holds. For given $x \in D$ we have $x \in f^<(\{f(x)\}) \in \mathrm{Part} f$ and $x \in \rho_>(\{x\}) \in \mathrm{Part} f$. Thus $\rho_>(\{x\}), f^<(\{f(x)\}) \in \mathrm{Part} f$ and $\rho_>(\{x\}) \cap f^<(\{f(x)\}) \neq \emptyset$, and therefore $\rho_>(\{x\}) = f^<(\{f(x)\}) = f^<(f_>(\{x\}))$. For every $U \in \mathfrak{P}(D)$ we have

$$\rho_>(U) = \rho_>(\bigcup_{x \in U} \{x\}) = \bigcup_{x \in U} \rho_>(\{x\}) = \bigcup_{x \in U} f^<(f_>(\{x\})) = f^<(f_>(\bigcup_{x \in U} \{x\})) = f^<(f_>(U)).$$

Therefore $\rho_> = f^< \circ f_>$.

(ii) $\Rightarrow$ (iii). Assume that (ii) holds. For all $x, y \in D$ we have the chain of equivalences

$$\begin{aligned} x \ \rho \ y \ \Leftrightarrow \ y \ \in \rho_>(\{x\}) \ \Leftrightarrow \ y \in f^<(f_>(\{x\})) \ &\Leftrightarrow \ f(y) \in f_>(\{x\}) \Leftrightarrow \\ &\Leftrightarrow \ f(y) = f(x). \end{aligned}$$

(iii) $\Rightarrow$ (i). Assume that (iii) holds. As was pointed out at the beginning of this section, ρ is then an equivalence relation. For every $x, y \in D$ we have the chain of equivalences

$$y \in \rho_>(\{x\}) \ \Leftrightarrow \ x \ \rho \ y \ \Leftrightarrow \ f(x) = f(y) \ \Leftrightarrow \ y \in f^<(\{f(x)\}).$$

Since $y \in D$ was arbitrary, we have $\rho_>(\{x\}) = f^<(\{f(x)\})$ for all $x \in D$, and therefore $\{\rho_>(\{x\}) \mid x \in D\} = \{f^<(\{f(x)\}) \mid x \in D\} = \mathrm{Part}\, f$. ■

57G. Corollary. *Let the relation ρ in the set D be given. Then ρ is an equivalence relation if and only if there exists a mapping f with* $\mathrm{Dom} f = D$ *and such that*

$$\forall x, y \in D, \quad x \ \rho \ y \ \Leftrightarrow \ f(x) = f(y).$$

Proof. The "if" part is the trivial remark at the beginning of this section (cf. Theorem 57F, ((iii) $\Rightarrow$ (i)). The "only if" part follows from Theorem 57C and Theorem 57D, ((i) $\Rightarrow$ (iv)), with f defined as the partition mapping of the partition $\{\rho_>(\{x\}) \mid x \in D\}$. ■

Our last result in this section concerns a reflexive, transitive relation and shows how the possible lack of antisymmetry, but for which it would be an order, can be remedied by moving from the domain to a suitable partition.

57H. Proposition. *Let the set D and the reflexive, transitive [and total] relation σ in D be given. Define the relation ρ in D by the rule*

$$\forall x, y \in D, \quad x \ \rho \ y \ :\Leftrightarrow \ (x \ \sigma \ y \text{ and } y \ \sigma \ x).$$

Then:

(a): *ρ is an equivalence relation.*

(b): *Let $\mathcal{P} := \{\rho_>(\{x\}) \mid x \in D\}$ be the partition associated with the equivalence relation ρ, and define the relation $\prec$ in $\mathcal{P}$ by the rule*

$$\forall A, B \in \mathcal{P}, \quad A \prec B \quad :\Leftrightarrow \quad (\exists (x, y) \in A \times B, \quad x \ \sigma \ y). \tag{57.1}$$

Then $\prec$ is a [total] order in $\mathcal{P}$ and satisfies

$$\forall A, B \in \mathcal{P}, \quad A \prec B \ \Leftrightarrow \ (\forall (x, y) \in A \times B, \quad x \ \sigma \ y), \tag{57.2}$$

$$\forall x, y \in D, \quad x \ \sigma \ y \ \Leftrightarrow \ \Omega_{\mathcal{P}}(x) \prec \Omega_{\mathcal{P}}(y). \tag{57.3}$$

Proof of (b). We first show that $\prec$ satisfies (57.2). Let $A, B \in \mathcal{P}$ be given, and assume that $A \prec B$. Let $x \in A$ and $y \in B$ be given. By (57.1) we may choose $x' \in A$ and $y' \in B$ such that $x' \ \sigma \ y'$. Now $x, x' \in A \in \mathcal{P}$, and hence $x \ \rho \ x'$, which implies $x \ \sigma \ x'$; similarly, we have $y \ \rho \ y'$, which implies $y' \ \sigma \ y$. Since σ is transitive, we conclude that $x \ \sigma \ y$. The reverse implication in (57.2) follows from (57.1), since neither A nor B is empty. (57.3) is a reformulation of (57.2).

Since no member of $\mathcal{P}$ is empty, and since σ is reflexive, $\prec$ is also reflexive. If $A, B \in \mathcal{P}$ satisfy $A \prec B$ and $B \prec A$, we choose $x \in A$ and $y \in B$; by (57.2) we have $x \ \sigma \ y$ and $y \ \sigma \ x$, hence $x \ \rho \ y$, hence $A = B$. This shows that $\prec$ is antisymmetric. The transitivity of $\prec$ follows at once from (57.2) and the transitivity of σ. [If σ is total, (57.1) at once implies that $\prec$ is total.] ■

This page intentionally left blank

Chapter 6

ORDERED SETS

61. Basic Concepts

A set D endowed with structure by the prescription of an order $\prec$ in D is called an **ordered set**; more specifically, the set D **ordered by** $\prec$. It is denoted simply by D, or, if for some reason the order must be made explicit, by $(D;\prec)$. Together with the order $\prec$ we also have the corresponding strict-order $\precneqq$ in D. An ordered set $(D;\prec)$ is unambiguously specified by prescribing a strict-order in the set D and requiring that this strict-order be $\precneqq$ (cf. Proposition 56A). An ordered set is said to be **totally ordered** if its order is total, and we speak of the set D **totally ordered by** $\prec$.

Let the set D ordered by $\prec$ be given. Every subset S of D is naturally endowed with the structure of an ordered set by the restriction $\prec|_S$ of $\prec$ to S. The set S ordered by $\prec|_S$ is called an **ordered subset of** D (the term should be "sub-ordered-set", but this is stylistically inadmissible). When confusion is unlikely, we refer to this ordered subset simply as S, or as S **ordered by** $\prec$. If D is totally ordered, every ordered subset is totally ordered. If the ordered subset S is totally ordered (even though D might not be), S is called a **chain of** D.

We introduce a collection of concepts concerning the structure of an ordered set. These concepts will be illustrated in Examples 61D. Throughout this section we consider a given set D ordered by $\prec$.

Let $x, y \in D$ be given. If $x \prec y$, we say that x **precedes** y, and that y **follows** x. If $x \precneqq y$, we say that x **strictly precedes** y, and that y **strictly follows** x. We shall use freely such contractions as $x \prec y \precneqq z$ to mean that $x \prec y$ and $y \precneqq z$, and similarly with other — and perhaps longer — combinations of $\prec$ and $\precneqq$. If $x \prec y$ we define the **order-intervals**

$$\begin{aligned}
[\![x,y]\!] &:= \{z \in D \mid x \prec z \prec y\} = \prec_>(\{x\}) \cap \prec^<(\{y\}) \\
]\!]x,y]\!] &:= [\![x,y]\!] \backslash \{x\} = \{z \in D \mid x \precneqq z \prec y\} \\
[\![x,y[\!] &:= [\![x,y]\!] \backslash \{y\} = \{z \in D \mid x \prec z \precneqq y\} \\
]\!]x,y[\!] &:= [\![x,y]\!] \backslash \{x,y\} = \{z \in D \mid x \precneqq z \precneqq y\} \ .
\end{aligned}$$

Order-intervals of the form $[\![x,y]\!]$ are said to be **closed**. We say that x **immediately precedes** y, and that y **immediately follows** x, if $[\![x,y[\!] = \{x\}$ or, equivalently, if $x \precneqq y$ and $]\!]x,y[\!] = \emptyset$. D is said to be **densely ordered**, and its order $\prec$ is said to be **dense**, if, for all $x,y \in D, x \precneqq y$ implies $]\!]x,y[\!] \neq \emptyset$. A subset A of D is said to be **order-convex** if

$$\forall x,y \in A, \quad x \prec y \ \Rightarrow \ [\![x,y]\!] \subset A.$$

It is obvious that every order-interval is order-convex.

If A is a subset of D and y a member of D, y is called an **upper bound of** A if y follows every member of A, i.e., if $A \subset \prec^<(\{y\})$; and y is called a **lower bound of** A if y precedes every member of A, i.e., if $A \subset \prec_>(\{y\})$. We introduce, for each subset A of D, the set $\mathrm{Ub}(A)$ of all its upper bounds and the set $\mathrm{Lb}(A)$ of all its lower bounds:

$$\begin{aligned}
\mathrm{Ub}(A) &:= \{y \in D \mid \forall x \in A,\ x \prec y\} = \bigcap_{x \in A}{}^{D} \prec_> (\{x\}) \\
\mathrm{Lb}(A) &:= \{y \in D \mid \forall x \in A,\ y \prec x\} = \bigcap_{x \in A}{}^{D} \prec^< (\{x\})
\end{aligned} \qquad \text{for all } A \in \mathfrak{P}(D).$$

The members of $\mathrm{Ub}(A)\backslash A$ are called **strict upper bounds of** A, and those of $\mathrm{Lb}(A)\backslash A$ are called **strict lower bounds of** A.

A subset of A of D is said to be **bounded from above** [**bounded from below**] (sometimes, in full, **order-bounded from above** [**order-bounded from below**]) if $\mathrm{Ub}(A) \neq \emptyset$ [$\mathrm{Lb}(A) \neq \emptyset$]. A is said to be **order-bounded** if $\mathrm{Ub}(A) \neq \emptyset$ and $\mathrm{Lb}(A) \neq \emptyset$. A subset A of D is said to be **cofinal** [**coinitial**] **in** D if $A \cap \mathrm{Ub}(\{x\}) \neq \emptyset$ [$A \cap \mathrm{Lb}(\{x\}) \neq \emptyset$] for every $x \in D$. (The word should properly be "confinal", but it is too late to change.)

Let a subset S of D and a subset A of S be given. It is clear that a member of S is an upper bound of A in the ordered subset S if and only if it is an upper bound of A in D itself; and similarly for lower bounds. We denote the sets of all upper bounds and of all lower bounds of A in the ordered subset S by $\mathrm{Ub}_S(A)$ and $\mathrm{Lb}_S(A)$, respectively, and find that

(61.1) $\mathrm{Ub}_S(A) = S \cap \mathrm{Ub}(A)$, $\mathrm{Lb}_S(A) = S \cap \mathrm{Lb}(A)$ for all $S \in \mathfrak{P}(D)$ and $A \in \mathfrak{P}(S)$.

In particular, of course, $\mathrm{Ub}_D(A) = \mathrm{Ub}(A)$ and $\mathrm{Lb}_D(A) = \mathrm{Lb}(A)$.

Let the subset A of D be given. Since the order is antisymmetric, each of the sets $\mathrm{Ub}_A(A) = A \cap \mathrm{Ub}(A)$ and $\mathrm{Lb}_A(A) = A \cap \mathrm{Lb}(A)$ is either empty or a singleton. If $\mathrm{Ub}_A(A) \neq \emptyset$ [$\mathrm{Lb}_A(A) \neq \emptyset$], we define the **maximum** [**minimum**] **of** A to be $\max A :\in \mathrm{Ub}_A(A)$ [$\min A :\in \mathrm{Lb}_A(A)$], i.e., the unique member of A that follows [precedes] every member of A.

Let the subset A of D be given. If $\mathrm{Ub}(A)$ has a minimum [$\mathrm{Lb}(A)$ has a maximum], we define the **supremum** [**infimum**] **of** A to be $\sup A := \min \mathrm{Ub}(A)$ [$\inf A := \max \mathrm{Lb}(A)$]. We note that, in this case, $\mathrm{Ub}(A) = {\prec_>}(\{\sup A\}) = \mathrm{Ub}(\{\sup A\})$ [$\mathrm{Lb}(A) = {\prec^<}(\{\inf A\}) = \mathrm{Lb}(\{\inf A\})$].

Instead of *maximum, minimum, supremum, infimum*, one sometimes prefers to use *greatest* (or *last*) *member, least* (or *first*) *member, least upper bound, greatest lower bound*, respectively; or similarly suggestive terms. (Instead of the symbols sup and inf one sometimes encounters lub and glb.)

It is important to distinguish carefully between maximum and supremum, and between minimum and infimum. For instance, a subset of D may have a supremum but fail to have a maximum (Examples 61D,(b),(c),(e)). The reverse situation, however, is impossible.

61A. Proposition. *Let the ordered set D and the subset A of D be given.*

Then:

(a): *If A has a maximum [minimum], then A has a supremum [infimum] and* $\sup A = \max A$ *[*$\inf A = \min A$*].*

(b): *If A has a supremum [infimum], then A has a maximum [minimum] if and only if* $\sup A \in A$ *[*$\inf A \in A$*].*

Let the subset S of D and the subset A of S be given. If A has a supremum [infimum] when regarded as a subset of the ordered set S, i.e., if $\mathrm{Ub}_S(A)$ has a minimum [$\mathrm{Lb}_S(A)$ has a maximum], we call that member of S the **supremum** [**infimum**] **of** A **with respect to** S, and denote it by $\sup_S A$ [$\inf_S A$]. In view of (61.1) we have

(61.2)
$$\begin{aligned} &\sup_S A := \min\, \mathrm{Ub}_S(A) = \min(S \cap \mathrm{Ub}(A)) \\ &\inf_S A := \max\, \mathrm{Lb}_S(A) = \max(S \cap \mathrm{Lb}(A)) \end{aligned} \qquad \text{for all } S \in \mathfrak{P}(D) \text{ and } A \in \mathfrak{P}(S).$$

Of course $\sup_D A = \sup A$ and $\inf_D A = \inf A$, while $\sup_A A = \max A$ and $\inf_A A = \min A$. The existence of a supremum of A with respect to S neither implies, nor is implied by, the existence of a supremum of A (with respect to D); and similarly for infima. There are, however, some useful consequences of (61.2).

61B. Proposition. *Let the set D ordered by $\prec$, the subsets S and T of D with $T \subset S$, and the subset A of T be given. Then:*

(a): *If A has a supremum [infimum] with respect to S, and also with respect to T, then*

$$\sup_S A \prec \sup_T A \qquad [\inf_T A \prec \inf_S A].$$

(b): *If A has a supremum [infimum] with respect to S, and $\sup_S A \in T$ [$\inf_S A \in T$], then A also has a supremum [infimum] with respect to T, and $\sup_T A = \sup_S A$ [$\inf_T A = \inf_S A$].*

We require yet another pair of concepts concerning a given subset A of D. A member m of A is of course followed and preceded by m itself; if m is not followed [preceded] by any other member of A, then m is called a **maximal** [**minimal**] **member of** A; this condition can be restated as $\mathrm{Ub}_A(\{m\}) = \{m\}$ [$\mathrm{Lb}_A(\{m\}) = \{m\}$]. It is again important to distinguish carefully between maximum and maximal member, and between minimum and minimal member. For instance, a subset of D may have maximal members, but fail to have a maximum (Examples 61D,(a),(c),(d),(e)): it is quite easy to exhibit one such example with a single maximal member. All of this cannot occur, however, if the order is total, and the reverse situation is altogether impossible.

61C. Proposition. *Let the ordered set D and the subset A of D be given. Then:*

(a): *If A has a maximum [minimum] then $\max A$ [$\min A$] is the unique maximal [minimal] member of A.*

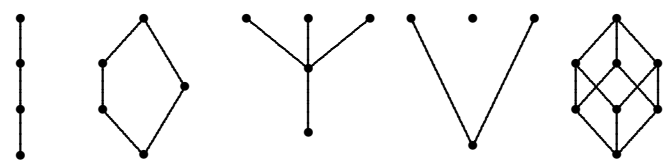
(b): *If the order is total, then a member of A is a maximal [minimal] member of A (if and) only if it is the maximum [minimum] of A.*

(c): *If A is cofinal [coinitial] in D, then a member of A is a maximal [minimal] member of A if and only if it is a maximal [minimal] member of D.*

(d): *If $\mathrm{Ub}(A)$ has an infimum [$\mathrm{Lb}(A)$ has a supremum], then that infimum [supremum] is the supremum [infimum] of A.*

(e): *A member of $\mathrm{Ub}(A)$ [of $\mathrm{Lb}(A)$] is a maximal member of $\mathrm{Ub}(A)$ [a minimal member of $\mathrm{Lb}(A)$] if and only if it is a maximal [minimal] member of D.*

61D. Examples. (a)*: An ordered finite set with reasonably few members can be represented graphically by dots and connecting line segments in the following way: x strictly precedes y if and only if the dot representing x can be connected to the dot representing y by an "ascending chain" of segments. Examples of such graphic representations are shown below.



(A) represents a totally ordered set. (B) represents an ordered set that has a maximum and a minimum. (C) represents an ordered set that has a minimum but no maximum; it has three maximal members. (D) represents an ordered set that has neither maximum nor minimum; it has three maximal and two minimal members. (E) represents the set $\mathfrak{P}(S)$ ordered by $\subset_S$, for a set S that has three members.

(b)*: Consider the set $\mathbb{P}$ of all positive real numbers (including 0), ordered by

the usual order $\leq$. This is a totally ordered set. The order-intervals $[\![s,t]\!]$, etc., are the usual intervals $[s,t]$, etc., in $\mathbb{P}$. If $s,t \in \mathbb{P}$ satisfy $s < t$, the sets $[s,t]$ and $]s,t]$ have a maximum, namely t, but $[s,t[$ and $]s,t[$ have none. $\mathbb{P}$ itself has no maximum, while $\min\mathbb{P} = 0$; but $\mathbb{P}^\times := \mathbb{P}\backslash\{0\}$ has no minimum. It is a consequence of the fundamental properties of the Real-Number System that every non-empty subset of $\mathbb{P}$ has an infimum (e.g., $\inf]s,t] = s$ if $s < t$, $\inf\{\frac{1}{n} \mid n \in \mathbb{N}^\times\} = 0$), and every subset of $\mathbb{P}$ that has *some* upper bound actually has a supremum (e.g., $\sup\{r \in \mathbb{P} \mid r \text{ rational}, r^2 < 2\} = \sqrt{2}$, $\sup_{\mathbb{P}}\emptyset = 0$).

(c)*: Consider the set $\mathbb{N}$ ordered by the relation "divides". Let the subset A of $\mathbb{N}$ be given. The upper bounds of A are the *common multiples of* A, and the lower bounds of A are the *common divisors of* A. The subset A always has a supremum, called the *least common multiple of* A and denoted by $\mathrm{lcm}A$: namely, $\mathrm{lcm}A = 0$ if $0 \in A$ or A is infinite; $\mathrm{lcm}A = 1$ if $A = \emptyset$; and $\mathrm{lcm}A$ is the usual least common multiple otherwise. A also always has an infimum, called the *greatest common divisor of* A, and denoted by $\gcd A$: namely, $\gcd A = 0$ if $A = \emptyset$ or $A = \{0\}$, and $\gcd A$ is the usual greatest common divisor otherwise. However, if $A \subset \mathbb{N}^\times$, A has a supremum with respect to $\mathbb{N}^\times$ if and only if A is finite, and an infimum with respect to $\mathbb{N}^\times$ if and only if A is not empty. Every non-empty subset of $\mathbb{N}$ has minimal members; in particular, the minimal members of $\mathbb{N}^\times\backslash\{1\}$ are the prime numbers. The special terminology mentioned in this example serves to prevent confusion with the concepts associated with the set $\mathbb{N}$ (totally) ordered by the usual order $\leq$.

(d): Every set D can trivially be endowed with the structure of an ordered set by prescribing the equality relation $=_D$ as the order. In this ordered set, every member of a subset is both a maximal and a minimal member. A subset has a maximum, a minimum, a supremum, or an infimum only when it is a singleton; the sole exception is that, when D itself is a singleton, the empty set also has a supremum and an infimum.

(e): Let the set S be given, and consider the collection $\mathfrak{P}(S)$ ordered by $\subset_S$. This ordered set is not totally ordered unless S is empty or a singleton. Every subcollection $\mathcal{U}$ of $\mathfrak{P}(S)$ has a supremum and an infimum, namely $\bigcup\mathcal{U}$ and $\bigcap^S\mathcal{U}$, respectively, as is easily verified. For $U,V \in \mathfrak{P}(S)$, the subcollection $\{U,V\}$ has neither maximum nor minimum unless $U \subset V$ or $V \subset U$. The minimal members of $\mathfrak{P}^\times(S)$ are the singletons of S.

(f): More generally, let a collection of sets $\mathcal{D}$ be given. In $\mathcal{D}$, the relation "is included in", defined by set-inclusion, is an order. The ordered set so obtained is called the collection $\mathcal{D}$ **ordered by inclusion**; it may be regarded as an ordered subset of $\mathfrak{P}(\bigcup\mathcal{D})$ ordered by $\subset_{\bigcup\mathcal{D}}$. In $\mathcal{D}$ ordered by inclusion, we may use the terms **largest member** and **smallest member** instead of *maximum* and *minimum*, respectively. If $\mathcal{D}$ is totally ordered by inclusion, $\mathcal{D}$ is called a **nest**.

(g): Let the set D be given. In Example 56B,(e) we defined the order "narrower than" in the set of all relations in D, by specifying that, for given relations ρ and σ in D, ρ is said to be **narrower than** σ if $\mathrm{Gr}\rho \subset \mathrm{Gr}\sigma$; this is thus essentially a special case of (e), with $S := D \times D$. An interesting ordered subset of the set of all

relations in D ordered by "narrower than" is $\mathrm{Ord}(D)$, the set of all *orders* in D. The next result characterizes the maximal members of $\mathrm{Ord}(D)$. ■

61E. Proposition. *Let the set D be given. An order ρ in D is a maximal member of $\mathrm{Ord}(D)$, ordered by the order "narrower than", if and only if ρ is a total order.*

Proof. The "if" part follows immediately from the fact that every member of $\mathrm{Ord}(D)$ is an antisymmetric relation.

To prove the "only if" part by contraposition, let the order $\rho \in \mathrm{Ord}(D)$ be given, and assume that ρ is not total. Choose $a, b \in D$ such that a and b are not ρ-comparable, and define the relation σ in D by

$$\forall x, y \in D, \quad x \ \sigma \ y :\Leftrightarrow \ (x \ \rho \ y \ \text{ or } \ (x \ \rho \ a \ \text{ and } \ b \ \rho \ y)).$$

This relation is obviously reflexive, and it is easily verified that σ is also antisymmetric and transitive, hence $\sigma \in \mathrm{Ord}(D)$. Moreover, ρ is clearly narrower than σ; since a and b are not ρ-comparable, we have $(a, b) \in \mathrm{Gr}\sigma \backslash \mathrm{Gr}\rho$, and we conclude that ρ is strictly narrower than σ. Hence ρ is not a maximal member of $\mathrm{Ord}(D)$. ■

61F. Remark. Together with the set D ordered by $\prec$ we may also consider the same set D ordered by the reverse order $\succ$. If A is a subset of D, the upper bounds of A in D ordered by $\prec$ are the lower bounds of A in D ordered by $\succ$, and vice versa. In the same manner, when the order is reversed, bounded from above becomes bounded from below, cofinal becomes coinitial, maximum becomes minimum, supremum becomes infimum, and maximal members become minimal members, of A; and vice versa. This observation allows us to cut in half many proofs and indeed statements, such as those of Propositions 61A, 61B, and 61C. This will be carried out as follows. A proposition may have an arrow $\uparrow$ added to its designation. This indicates that another proposition may be obtained from it, on account of this remark, by replacing "upper bound" by "lower bound" and vice versa, and making the other interchanges mentioned before. This other proposition is not *stated*, but may be *quoted* by the same designation with an arrow $\downarrow$ added instead.

Since the reverse of a total order is a total order, this remark extends to propositions concerning totally ordered sets. ■

An example of the convention in Remark 61F is the following technical result.

61G↑. Proposition. *In every totally ordered set every non-empty order-interval has a supremum.*

Proof. Let the non-empty order interval J in the totally ordered set D be given. If J has a maximum, then J has a supremum (Proposition 61A,(a)). Assume that J has no maximum; then $J = [\![a, b[\![$ or $J =]\!]a, b[\![$ for suitable $a, b \in D$ with $a \precneqq b$. Choose $c \in J$. Now $b \in \mathrm{Ub}(J)$. Let $x \in \mathrm{Ub}(J)$ be given; then $c \prec x$. If $x \precneqq b$ we should have $x \in J$, which is excluded, since J has no maximum. Since D is totally ordered, we must have $b \prec x$. Thus b is the minimum of $\mathrm{Ub}(J)$, hence the supremum of J. ■

62. Isotone mappings

Let the sets D and D' ordered by $\prec$ and $\prec'$, respectively, be given. A mapping $f\colon D \to D'$ is said to be **isotone** (in full, **$\prec$-$\prec'$-isotone**), and is called an **order-morphism from** the ordered set D **to** the ordered set D', if it satisfies the condition

$$\forall x, y \in D, \quad x \prec y \;\Rightarrow\; f(x) \prec' f(y).$$

A mapping $f\colon D \to D'$ is said to be **strictly isotone** (in full, **strictly $\prec$-$\prec'$-isotone**), and is called a **strict-order-morphism from D to D'**, if it satisfies the condition

$$\forall x, y \in D, \quad x \precneqq y \;\Rightarrow\; f(x) \precneqq' f(y).$$

The set of all isotone mappings from D to D' is denoted by $\mathrm{Isot}(D, D')$ or, when it is indispensable to record the orders, by $\mathrm{Isot}((D;\prec),(D';\prec'))$; thus

$$\mathrm{Isot}(D, D') := \mathrm{Isot}((D;\prec),(D';\prec')) := \{f \in \mathrm{Map}(D, D') \mid f \text{ is } \prec\text{-}\prec'\text{-isotone}\}.$$

62A. Remarks. (a): Every strictly isotone mapping is isotone. Every injective isotone mapping is strictly isotone. A strictly isotone mapping, however, need not be injective: for instance, if the sets D and D' are given, and ordered by the respective equality relations $=_D$ and $=_{D'}$, then *every* mapping from D to D' is strictly isotone. On the other hand, an isotone mapping with a *totally* ordered domain is strictly isotone (if and) only if it is injective, and in that case, moreover, it satisfies

$$\forall x, y \in D, \quad x \precneqq y \;\Leftrightarrow\; f(x) \precneqq' f(y).$$

(b): If S is an ordered subset of an ordered set D, the inclusion mapping $1_{S\subset D}$ is strictly isotone; in particular, for every ordered set D the identity mapping 1_D is strictly isotone from D to itself.

(c): Let f be a [strictly] isotone mapping from D to D', and let S be a subset of D and S' a subset of D' such that $f_>(S) \subset S'$. Then $f|_S^{S'}$ is [strictly] isotone from the ordered subset S to the ordered subset S'. ■

62B. Proposition. *Let the [strictly] isotone mappings f from D to D' and g from D' to D'' be given. Then $g \circ f$ is a [strictly] isotone mapping from D to D''.*

In the following proposition, the terms *maximum, supremum*, etc., and the symbols Ub, max, etc., refer to the set D ordered by $\prec$ or to the set D' ordered by $\prec'$, as the case may require; when the restrictions of $\prec$ and $\prec'$ to $D \cap D'$ do not agree, special precautions must of course be taken to prevent confusion.

62C↑. Proposition. *Let the isotone mapping f from D to D' be given. Let the subsets S of D and S' of D' satisfy $f_>(S) \subset S'$. For every subset A of S we have:*

(a): $$f_>(\mathrm{Ub}_S(A)) \subset \mathrm{Ub}_{S'}(f_>(A)).$$

(b): *If A has a maximum then $f_>(A)$ has a maximum, and* $\max f_>(A) = f(\max A)$.

(c): *If A has a supremum with respect to S and $f_>(A)$ has a supremum with respect to S', then* $\sup_{S'} f_>(A) \prec' f(\sup_S A)$.

(d): *If f is strictly isotone and $m \in A$ is such that $f(m)$ is a maximal member of $f_>(A)$, then m is a maximal member of A.*

Let the sets D and D' ordered by $\prec$ and $\prec'$, respectively, be given. For each $f \in \mathrm{Isot}(D, D')$ we have the question: Is there $g \in \mathrm{Isot}(D', D)$ such that $g \circ f = 1_D$ and $f \circ g = 1_{D'}$? The existence of such a g plainly requires f to be an invertible, or equivalently, a bijective, mapping; and if this is the case, any such g must be precisely $f^{\leftarrow}$. The answer to the question is therefore affirmative if and only if f is an invertible mapping and $f^{\leftarrow} \in \mathrm{Isot}(D', D)$; equivalently, if and only if f is bijective and

$$\forall x, y \in D, \quad x \prec y \;\Leftrightarrow\; f(x) \prec' f(y). \tag{62.1}$$

If these conditions are satisfied, f is called an **order-isomorphism from D to D'**. Obviously, if f is an order-isomorphism, so is $f^{\leftarrow}$. We say that the ordered set D is **order-isomorphic to** the ordered set D', or that D **and** D' **are order-isomorphic**, if there exists an order-isomorphism from D to D'.

It should be noted that an isotone mapping may be bijective and yet fail to be an order-isomorphism. For instance, choose an set D ordered by some order $\prec$ distinct from the equality relation $=_D$ (such an order can be found unless D is empty or a singleton). Then $1_D \in \mathrm{Isot}((D; =_D), (D; \prec))$, but $(1_D)^{\leftarrow} = 1_D \notin \mathrm{Isot}((D; \prec), (D; =_D))$. This situation is ruled out, however, if the domain of the bijection is totally ordered.

62D. Proposition. *An isotone mapping from a totally ordered set to an ordered set is an order-isomorphism if (and only if) it is bijective.*

62E. Proposition. *Let the ordered set D be given. The mapping $x \mapsto \mathrm{Lb}(\{x\})$ is an order-isomorphism from D to a subcollection of $\mathfrak{P}(D)$, ordered by inclusion.*

62F. Remarks. (a): Every order-isomorphism is injective; hence both it and its inverse are strictly isotone (Remark 62A,(a)). It follows that there is no separate notion of "strict-order-isomorphism".

(b): Every composite of order-isomorphisms is an order-isomorphism.

(c): Every order-isomorphism "preserves the structure of the ordered sets". More precisely, if f is an order-isomorphism from D to D', then x precedes y in D if and only if $f(x)$ precedes $f(y)$ in D' (cf. (62.1)); D is totally ordered if and only if D' is totally ordered; and with the notations used in Proposition 62C↑, $f_>(\mathrm{Ub}_S(A)) = \mathrm{Ub}_{f_>(S)}(f_>(A))$; A has a maximum if and only if $f_>(A)$ has a maximum; A has a supremum with respect to S if and only if $f_>(A)$ has a supremum with respect to $f_>(S)$, and $\sup_{f_>(S)} f_>(A) = f(\sup_S A)$; for every $m \in A$, m is a maximal member of A if and only if $f(m)$ is a maximal member of $f_>(A)$; etc. ■

When dealing with sets D and D' ordered by $\prec$ and $\prec'$, respectively, it is sometimes convenient to consider also the same sets ordered by the reverse orders $\succ$ and $'\succ$. It is clear that a mapping $f\colon D \to D'$ is [strictly] $\prec$-$\prec'$-isotone if and only if it is [strictly] $\succ$-$'\succ$-isotone. A mapping $f\colon D \to D'$ is said to be [**strictly**] **antitone** (in full, [**strictly**] $\prec$-$\prec'$-**antitone**) if it is [strictly] $\prec$-$'\succ$-isotone, i.e., if it satisfies the

condition

$$\forall x, y \in D, \quad x \prec y \;\Rightarrow\; f(y) \prec' f(x)$$

$$[\forall x, y \in D, \quad x \preceq y \;\Rightarrow\; f(y) \preceq' f(x)].$$

An order-isomorphism from D ordered by $\prec$ to D' ordered by $'\succ$ is called an **order-antimorphism from** D ordered by $\prec$ **to** D' ordered by $\prec'$. The properties of antitone mappings and of their interactions with isotone mappings are simple translations of properties of isotone mappings, and need not be recorded here.

62G. Examples. (a)*: The mapping $(t \mapsto t^3) : \mathbb{R} \to \mathbb{R}$ is an order-isomorphism; $(t \mapsto t^2) : \mathbb{P} \to \mathbb{P}$ is an order-isomorphism; $(t \mapsto t^2) : \mathbb{P} \to \mathbb{R}$ is strictly isotone, but $(t \mapsto t^2) : \mathbb{R} \to \mathbb{P}$ is not isotone. Here $\mathbb{R}$ and $\mathbb{P}$ are ordered by the usual order $\leq$.

(b)*: Consider the ordered sets $(\mathbb{N}^\times; \leq)$ and $(\mathbb{N}^\times;$"divides"$)$. The mapping $1_{\mathbb{N}^\times}$ is strictly isotone from $(N^\times;$"divides"$)$ to $(\mathbb{N}^\times; \leq)$. (Here we use the fact that $m \leq mn$ for all $m, n \in \mathbb{N}^\times$.) However, $(1_{\mathbb{N}^\times})^{\leftarrow} = 1_{\mathbb{N}^\times}$ is not isotone from $(\mathbb{N}^\times; \leq)$ to $(\mathbb{N}^\times;$ "divides"$)$: for instance, $2 \leq 5$, but 2 does not divide 5.

(c): Let the set S be given. The **complementation mapping** $C_S : \mathfrak{P}(S) \to \mathfrak{P}(S)$ defined by $C_S(U) := S \backslash U$ for all $U \in \mathfrak{P}(S)$ is an order-antimorphism from $\mathfrak{P}(S)$ ordered by inclusion to itself. Not every ordered set has an order-antimorphism from it to itself; for instance, if the ordered set has a maximum but no minimum there cannot be such an order-antimorphism. ■

Let the ordered sets D and D' be given. A mapping $f\colon D \to D'$ is said to be **monotone** if either f is isotone or f is antitone. (We mention that the now current terms "isotone", "antitone", "strictly isotone", "strictly antitone" provide crisp, unambiguous replacements for the more old-fashioned "monotone increasing", "monotone decreasing", "monotone strictly increasing", "monotone strictly decreasing", respectively; or, *but only when D and D' are totally ordered*, for the even less transparent terms "monotone non-decreasing", "monotone non-increasing", "monotone increasing", "monotone decreasing", respectively.)

We present a useful characterization of monotone mappings for certain ordered sets.

62H. Proposition. *Let the ordered sets D and D' (order in both denoted by $\prec$) and the mapping $f\colon D \to D'$ be given. Assume that in D every doubleton is order-bounded. (This assumption is satisfied in particular if D is totally ordered.) The following statements are equivalent:*

(i): *f is monotone.*

(ii): $\forall u, v, w \in D, \quad u \prec v \prec w \;\Rightarrow\; (f(u) \prec f(v) \prec f(w)$ *or* $f(u) \succ f(v) \succ f(w))$.

Proof. (i) obviously implies (ii). To prove that (ii) implies (i), assume that f is not monotone. We may then choose $v, w, v', w' \in D$ such that $v \prec w$ and $v' \prec w'$, but such that

$$\text{neither } f(v) \prec f(w) \text{ nor } f(v') \succ f(w'). \tag{62.2}$$

Choose a lower bound u of $\{v, v'\}$ and an upper bound z of $\{w, w'\}$. Then $u \prec v \prec w \prec z$ and $u \prec v' \prec w' \prec z$. If (ii) held, it would follow from (62.2) that

$$f(u) \succ f(v) \succneqq f(w) \quad \text{and} \quad f(v) \succneqq f(w) \succ f(z) \text{ and}$$

$$f(u) \prec f(v') \precneqq f(w') \quad \text{and} \quad f(v') \precneqq f(w') \prec f(z),$$

whence $f(u) \succneqq f(z)$ and $f(u) \precneqq f(z)$, a contradiction. We conclude that (ii) cannot hold. By contraposition, (ii) implies (i). ■

62I. Remark. The assumption on the ordered set D may not be omitted in the statement of Proposition 62H. Indeed, consider the ordered sets D and D' and the mapping f indicated in the following graphic representation.

D $\qquad\qquad$ D'

Here D and D' are finite, D' is totally ordered, and in D every doubleton has an upper bound — indeed, a supremum. the mapping f satisfies (ii), but is not monotone. ■

63. Products

Let a family of ordered sets $((A_i;\prec_i) \mid i \in I)$ be given. We define the **product of** $((A_i;\prec_i) \mid i \in I)$ — the family of **factors** — to be the set $\bigtimes_{i\in I} A_i$ ordered by the **product order** $\prec$, defined by the rule

$$\forall x, y \in \bigtimes_{i\in I} A_i, \quad x \prec y \;:\Leftrightarrow\; (\forall i \in I, \quad x_i \prec_i y_i). \tag{63.1}$$

It is plain that $\prec$ is indeed an order; it is said to be **defined termwise**.
We note that $\succ$ is the product order of the family $((A_i;_i \succ) \mid i \in I)$.

The product order can be characterized in various ways involving the projections $\pi_j : \bigtimes_{i\in I} A_i \to A_j$ for $j \in I$.

63A. Proposition. *Let the family of ordered sets $((A_i;\prec_i) \mid i \in I)$ be given, and let $\prec$ be its product order. Then:*

(a): *π_j is $\prec$-$\prec_j$-isotone for all $j \in I$.*

(b): *The product order $\prec$ is broader than every order $\prec'$ in $\bigtimes_{i\in I} A_i$ such that π_j is $\prec'$-$\prec_j$-isotone for all $j \in I$.*

Proof. (a) is an immediate consequence of the definition. To prove (b), let $\prec'$ be an order as specified. For all $x, y \in \bigtimes_{i\in I} A_i$ we have the chain of implications and equivalences

$$x \prec' y \;\Rightarrow\; (\forall i \in I,\ \pi_i(x) \prec_i \pi_i(y)) \;\Leftrightarrow\; (\forall i \in I,\ x_i \prec_i y_i) \;\Leftrightarrow\; x \prec y. \blacksquare$$

▼ **63B. Proposition**. *Let the family of ordered sets $((A_i;\prec_i) \mid i \in I)$ be given. There is exactly one order $\prec$ in the product set $\bigtimes_{i\in I} A_i$ with the following properties:*

(a): $\pi_j \in \mathrm{Isot}((\bigtimes_{i\in I} A_i;\prec), (A_j;\prec_j))$ *for all $j \in I$;*

(b): *for every ordered set $(B;\prec')$ and every family of isotone mappings $(f_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Isot}((B;\prec'), (A_i;\prec_i))$ there is exactly one $f \in \mathrm{Isot}((B;\prec'), (\bigtimes_{i\in I} A_i;\prec))$* such that *$f_j = \pi_j \circ f$ for all $j \in I$.*

This unique order is the product order, and the unique f in (b) *is defined by the rule*

$$f(y) := (f_i(y) \mid i \in I) \qquad \text{for all } y \in B. \tag{63.2}$$

Proof. 1. Let $\prec$ be an order in $\bigtimes_{i\in I} A_i$ that satisfies (a) and (b). From (a) and Proposition 63A,(b) we infer that $\prec$ is narrower than the product order.

We now apply (b) to the special case in which $B := \bigtimes_{i\in I} A_i, \prec'$ is the product order, and $f_i := \pi_i$ for each $i \in I$. By Proposition 63A,(a), each f_i is indeed isotone, so that (b) is applicable. Now we know a mapping $f\colon \bigtimes_{i\in I} A_i \to \bigtimes_{i\in I} A_i$ such that

$\pi_j = f_j = \pi_j \circ f$ for all $j \in I$, namely the identity mapping of $\bigtimes_{i\in I} A_i$; by Proposition 44C, it is the only one. This identity mapping must therefore be precisely the $\prec$-$\prec$-isotone mapping from $\bigtimes_{i\in I} A_i$ to $\bigtimes_{i\in I} A_i$ guaranteed by (b). It follows that $\prec$ is broader than the product order $\prec'$.

Since $\prec$ is both broader and narrower than the product order, it must itself be the product order.

2. Conversely, let $\prec$ be the product order. It follows from Proposition 63A,(a) that $\prec$ satisfies (a); it remains to show that it satisfies (b). Let the set B ordered by $\prec'$ and the family of isotone mappings $(f_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Isot}((B;\prec'),(A_i;\prec_i))$ b be given. By Proposition 44C, there is exactly one $f \in \mathrm{Map}(B, \bigtimes_{i\in I} A_i)$, and hence at most one $f \in \mathrm{Isot}((B;\prec'),(\bigtimes_{i\in I} A_i;\prec))$, such that $f_j = \pi_j \circ f$ for all $j \in I$; moreover, that unique f is defined by the rule (63.2). To complete the proof, we must show that this f is in fact $\prec'$-$\prec$-isotone. This follows from the chain of implications and equivalences obtained from (63.2) and (63.1), and valid for all $y, z \in B$:

$$y \prec' z \;\Rightarrow\; (\forall i \in I, f_i(y) \prec_i f_i(z)) \;\Leftrightarrow\; (\forall i \in I, (f(y))_i \prec_i (f(z))_i) \Leftrightarrow$$

$$\Leftrightarrow\; f(y) \prec f(z). \blacksquare$$

▲

63C. Remark. The product of a family of totally ordered sets is in general not totally ordered: indeed, if it is not empty, it is totally ordered only when all factors are singletons with at most one exception, and the exceptional factor is totally ordered.■

In the next proposition, Ub, *maximum, supremum*, etc., refer to each set A_i ordered by $\prec_i$ or to the product $\bigtimes_{i\in I} A_i$ ordered by the product order $\prec$, as the case may require.

63D↑. Proposition. *Let the family of ordered sets $((A_i;\prec_i) \mid i \in I)$ be given, and let $\prec$ be its product order. Let the subset B of $\bigtimes_{i\in I} A_i$ be given. Then:*

(a): $$\mathrm{Ub}(B) = \bigtimes_{i\in I} \mathrm{Ub}((\pi_i)_>(B)).$$

(b): *If B has a maximum, then $(\pi_i)_>(B)$ has a maximum for every $i \in I$, and* $\max B = (\max(\pi_i)_>(B) \mid i \in I)$.

(c): *B has a supremum if and only if $(\pi_i)_>(B)$ has a supremum for every $i \in I$; and in that case* $\sup B = (\sup(\pi_i)_>(B) \mid i \in I)$.

We conclude this section with some variants of the product order. Let a pair of ordered sets $((A';\prec'),(A'';\prec''))$ be given. We then define the **product of** the pair to be the set $A' \times A''$ ordered by the **product order** $\prec$, defined by the rule

$$\forall (x',x''),(y',y'') \in A' \times A'', \quad (x',x'') \prec (y',y'') \quad :\Leftrightarrow$$

$$\Leftrightarrow\; (x' \prec' y' \text{ and } x'' \prec'' y'').$$

This product corresponds to the product defined earlier in this section under the identification of the product set $A' \times A''$ with the Cartesian product of a list of length

2 (Section 44, p. 59), and therefore all the preceding results of this section can be applied to it with the appropriate translation.

Let the set S and the set D ordered by $\prec$ be given. We then define in $\mathrm{Map}(S, D)$ the **valuewise order** $\prec$ by the rule

$$\forall f, g \in \mathrm{Map}(S, D), \quad f \prec g \ :\Leftrightarrow \ (\forall s \in S, \quad f(s) \prec g(s)).$$

Thus the valuewise order $\prec$ is merely the product order on the Cartesian product D^S under the identification of $\mathrm{Map}(S, D)$ with that product (Section 44, p. 59). The use of the same symbol for the order on D and for the valuewise order on $\mathrm{Map}(S, D)$ should not cause any confusion.

64. Properties of ordered sets

The order of an ordered set may have some useful special properties. One instance is totality; totally ordered sets have already been considered in the preceding sections. We mention a few other properties in this section.

Let the set D ordered by $\prec$ be given. If for all $x, y \in D$ there is a member of D that follows both x and y, i.e., if $\mathrm{Ub}(\{x, y\}) \neq \emptyset$ for all $x, y \in D$, then the ordered set D is called a **directed set**; specifically, the set D **directed by** $\prec$. If $\{x, y\}$ has both a supremum and infimum for all $x, y \in D$, the ordered set D is called a **lattice**, and the order $\prec$ is called a **lattice-order**. Every totally ordered set is a lattice, and every lattice is a directed set. If $\prec$ is a lattice-order, so is the reverse order $\succ$. These concepts will not be discussed any further at this time. We shall, however, devote the entire next chapter to ordered sets with a property stronger than that of being a lattice: a set D ordered by $\prec$ is said to be **completely ordered** (**by** $\prec$) if *every* subset of D has both a supremum and an infimum. A completely ordered set is also called a **complete lattice**.

A set D ordered by $\prec$ is said to be **well-founded** (**by** $\prec$) if every non-empty subset of D has at least one minimal member. This kind of ordered set is related to the notions of "proof by induction" and "recursive definition" that will be discussed at length in Chapter 8. Finally, a set D ordered by $\prec$ is said to be **well-ordered** (**by** $\prec$) if every non-empty subset of D has a minimum. We note that every ordered subset of a well-founded [well-ordered] set is also well-founded [well-ordered].

64A. Proposition. *An ordered set is well-ordered if and only if it is both totally ordered and well-founded.*

Proof. An ordered set is totally ordered if and only if every doubleton included in it has a minimum. The assertion then follows from Proposition 61C. ■

64B. Proposition. *Let the ordered set D and D' be given. If D' is well-founded and there exists at least one strictly isotone mapping from D to D', then D is also well-founded.*

Proof. This follows at once from Proposition 62C↓,(d). ■

64C. Proposition. *The product of a pair of well-founded sets is well-founded.*

Proof. Let $((A'; \prec'), (A''; \prec''))$ be a pair of well-founded sets, and let $\prec$ be the product order in $A' \times A''$, defined by (63.3). Let the non-empty subset B of $A' \times A''$ be given. Choose a minimal member m' of the non-empty subset $\{x' \in A' \mid \exists x'' \in A'',\ (x', x'') \in B\}$ of A', and choose a minimal member m'' of the non-empty subset $\{x'' \in A'' \mid (m', x'') \in B\}$ of A''. Then (m', m'') is a minimal member of B in the product $A' \times A''$ ordered by $\prec$. ■

64D. Proposition. *Let the ordered set D and the collection $\mathcal{C}$ of ordered subsets of D be given. Assume that*

$$\bigcup \mathcal{C} = D \tag{64.1}$$

every member of $\mathcal{C}$ is well-founded (64.2)

(64.3) $$\forall A \in \mathcal{C},\ \forall x \in A, \quad \mathrm{Lb}(\{x\}) \subset A.$$

Then D is well-founded.

Proof. Let the non-empty subset S of D be given. By (64.1) we may choose $A \in \mathcal{C}$ such that $A \cap S \neq \emptyset$. By (64.2) we may choose a minimal member m of $A \cap S$. By (64.3) we have $S \cap \mathrm{Lb}(\{m\}) = S \cap A \cap \mathrm{Lb}(\{m\}) = \{m\}$, so that m is a minimal member of S. ■

64E. Examples. (a): For every set S, the collection $\mathfrak{P}(S)$ is completely ordered by inclusion (Example 61D,(e)). *We mention that $\mathfrak{P}(S)$ is well-founded by inclusion if and only if S is a finite set (Section 102).

(b)*: The set $\mathbb{R}$ ordered by $\leq$ is totally ordered. It is neither completely ordered nor well-ordered, since it has no minimum. Its ordered subset $\mathbb{P}$, though it has a minimum, is neither completely ordered (it has no maximum) nor well-ordered ($\mathbb{P}^\times$ has no minimum). If $s, t \in \mathbb{R}$ satisfy $s < t$, then the ordered subset $[s, t]$ is completely ordered, but not well-ordered. The ordered subset $\mathbb{N}$ is well-ordered, but not completely ordered (it has no maximum).

(c)*: The set $\mathbb{N}$ ordered by the relation "divides" is completely ordered (Example 61D,(c)), but not totally ordered. The ordered subset $\mathbb{N}^\times$ is a lattice, but it is not completely ordered (since it has no maximum). It is, however, well-founded, by Proposition 64B, since, as pointed out in Example 62G,(b), the mapping $1_{\mathbb{N}^\times}$ is strictly isotone from $\mathbb{N}^\times$ ordered by "divides" to $\mathbb{N}^\times$ well-ordered by $\leq$. ■

▾ 65. Lexicographic products and ordered direct unions

In this section we sketch some additional ways of constructing new ordered sets from old ones. Proofs (some tedious, but all straightforward) are left to the reader.

Let a family of ordered sets $((A_i;\prec_i) \mid i\in I)$ be given, where I is a *well-ordered* set (we avoid naming its order here). For all $x,y\in \bigtimes_{i\in I} A_i$ with $x\neq y$ we may define $k(x,y):=\min\{i\in I \mid x_i\neq y_i\}$ in the well-ordered set I. We then define the relation lex (in full lex_A) in $\bigtimes_{i\in i} A_i$ by

$$\forall x,y\in \bigtimes_{i\in I} A_i;\quad x \text{ lex } y \;:\Leftrightarrow\; (x\neq y \;\Rightarrow\; x_{k(x,y)} \precneqq_{k(x,y)} y_{k(x,y)}).$$

65A. Proposition. *Let the family of ordered sets $((A_i;\prec_i) \mid i\in I)$ be given, where I is a well-ordered set.*

(a): *The relation* lex *in $\bigtimes_{i\in I} A_i$ is an order, and it is broader than the product order.*

(b): *Assume that $\bigtimes_{i\in I} A_i\neq\emptyset$. Then* lex *is total if and only if $\prec_i$ is total for every $i\in I$.*

The order lex in $\bigtimes_{i\in I} A_i$ is called **lexicographic order**, and the ordered set $(\bigtimes_{i\in I} A_i;\text{lex})$ is called the **lexicographic product of** the family $((A_i;\prec_i) \mid i\in I)$.

For a given pair of ordered spaces $((A';\prec'),(A'';\prec''))$, the appropriate variant definition of the order lex in the product $A'\times A''$ (yielding the **lexicographic product** $(A'\times A'';\text{lex})$ **of** the pair) is

$$\begin{aligned}\forall (x',x''),(y',y'') \;&\in A'\times A'',\quad (x',x'') \text{ lex } (y',y'') \;:\Leftrightarrow\\ &\Leftrightarrow (x' \precneqq' y' \text{ or } (x'=y' \text{ and } x''\prec'' y'')).\end{aligned}$$

The following result is proved exactly like Proposition 64C.

65B. Proposition *The lexicographic product of a pair of well-founded sets is well-founded.*

Let now a family of ordered sets $((A_i;\prec_i) \mid i\in I)$ be given, where I is an *ordered* set (we use the word *precedes* to denote the order). We define the relation $\prec$ in $\dot{\bigcup_{i\in I}} A_i$ as follows.

$$\forall (j,a),(k,b)\in \dot{\bigcup_{i\in I}} A_i, (j,a)\prec (k,b) \;:\Leftrightarrow\; (j \text{ strictly precedes } k \text{ or } (j=k \text{ and } a\prec_j b)).$$

65C. Proposition. *Let the family of ordered sets* $((A_i;\prec_i) \mid i \in I)$ *be given, where* I *is an ordered set.*

(a): *The relation* $\prec$ *in* $\dot{\bigcup_{i\in I}} A_i$ *is an order.*

(b): *The ordered set* $(\dot{\bigcup_{i\in I}} A_i;\prec)$ *has any one of the following properties: totally ordered, completely ordered, well-founded, well-ordered, if and only if the ordered subset* $\mathrm{Supp}(A_i \mid i \in I)$ *of* I *as well as* $(A_i;\prec_i)$ *for every* $i \in I$ *have the same property.*

The ordered set $(\dot{\bigcup_{i\in I}} A_i;\prec)$ thus defined is called the **ordered direct union of** the family $((A_i;\prec_i) \mid i \in I)$. ▲

This page intentionally left blank

Chapter 7

COMPLETELY ORDERED SETS

71. Completely ordered sets

We recall a definition from Section 64. A set D ordered by $\prec$ is said to be **completely ordered** (**by** $\prec$) if every subset of D has both a supremum and an infimum.

71A. Remarks. (a): If D is a completely ordered set, then D has both a maximum and a minimum, since $\sup_D D = \max D$ and $\inf_D D = \min D$. In particular, D is not empty.

(b): If D is completely ordered by $\prec$, then D is also completely ordered by the reverse order $\succ$. The convention introduced in Remark 61F is therefore in force for propositions concerning completely ordered sets.

(c): For every set S, the collection $\mathfrak{P}(S)$ is completely ordered by inclusion; the supremum and infimum of a subcollection $\mathcal{U}$ of $\mathfrak{P}(S)$ are, respectively, $\bigcup \mathcal{U}$ and $\bigcap^S \mathcal{U}$ (Examples 61D,(e) and 64E,(a)).

(d): Let the set D be given. By (c), the collection $\mathfrak{P}(D \times D)$ is completely ordered by inclusion. Consequently, the set of all relations in D is completely ordered by the relation "narrower than" (Example 56B,(e)). ■

In defining a completely ordered set we required that every subset have both a supremum and an infimum; it turns out that requiring the existence of either would have been enough.

71B↑. Proposition. *An ordered set D is completely ordered if (and only if) every subset of D has a supremum.*

Proof. Assume that every subset of D has a supremum. Let the subset A of D be given, and set $s := \sup \mathrm{Lb}(A)$. By Proposition 61C,(d), s is the infimum of A. ■

71C. Example. Let the set S be given, and consider the collection $\mathrm{Part}(S)$ of all partitions of S (a subcollection of $\mathfrak{P}(\mathfrak{P}(S))$). The relation $\sqsubset_S$ in $\mathrm{Part}(S)$, defined by the rule

$$\forall \mathcal{P}, \mathcal{Q} \in \mathrm{Part}(S), \quad \mathcal{P} \sqsubset_S \mathcal{Q} \ :\Leftrightarrow \ \mathcal{P} \sqsubset \mathcal{Q}$$

is an order (see Proposition 18D). We shall now show that, in the collection $\mathrm{Part}(S)$ ordered by $\sqsubset_S$, every subcollection of $\mathrm{Part}(S)$ has a supremum. It will then follow

by Proposition 71B↑ that $\mathrm{Part}(S)$ is completely ordered by $\sqsubset_S$.

Let the subcollection Γ of $\mathrm{Part}(S)$ be given. We claim that the collection

$$\mathcal{C} := \{\bigcap_{\mathcal{P}\in\Gamma}{}^S\Omega_{\mathcal{P}}(x) \mid x \in S\} \subset \mathfrak{P}(S)$$

is a partition of S and that, in fact, $\mathcal{C}$ is the supremum of Γ. For every $x \in S$, we have $x \in \Omega_{\mathcal{P}}(x)$ for all $\mathcal{P} \in \Gamma$, and hence $x \in \bigcap_{\mathcal{P}\in\Gamma}{}^S\Omega_{\mathcal{P}}(x)$. Therefore $\bigcup\mathcal{C} = S$ and $\emptyset \notin \mathcal{C}$. Let $x, y \in S$ be given, and suppose that $(\bigcap_{\mathcal{P}\in\Gamma}{}^S\Omega_{\mathcal{P}}(x)) \cap (\bigcap_{\mathcal{P}\in\Gamma}{}^S\Omega_{\mathcal{P}}(y)) \neq \emptyset$. Then $\Omega_{\mathcal{P}}(x) \cap \Omega_{\mathcal{P}}(y) \neq \emptyset$, whence $\Omega_{\mathcal{P}}(x) = \Omega_{\mathcal{P}}(y)$, for each $\mathcal{P} \in \Gamma$. Therefore $\bigcap_{\mathcal{P}\in\Gamma}{}^S\Omega_{\mathcal{P}}(x) = \bigcap_{\mathcal{P}\in\Gamma}{}^S\Omega_{\mathcal{P}}(y)$. Since $x, y \in S$ were arbitrary, we conclude that $\mathcal{C}$ is a partition of S.

For every $\mathcal{Q} \in \Gamma$ we have $\bigcap_{\mathcal{P}\in\Gamma}{}^S\Omega_{\mathcal{P}}(x) \subset \Omega_{\mathcal{Q}}(x)$ for each $x \in S$; hence $\mathcal{Q} \sqsubset \mathcal{C}$. Since $\mathcal{Q} \in \Gamma$ was arbitrary, we conclude that $\mathcal{C}$ is an upper bound of Γ. On the other hand, suppose that the partition $\mathcal{R}$ of S is an upper bound of Γ. For each $\mathcal{P} \in \Gamma$ we then have $\mathcal{P} \sqsubset \mathcal{R}$, and therefore $\Omega_{\mathcal{R}}(x) \subset \Omega_{\mathcal{P}}(x)$ for every $x \in S$. It follows that $\Omega_{\mathcal{R}}(x) \subset \bigcap_{\mathcal{P}\in\Gamma}{}^S\Omega_{\mathcal{P}}(x)$ for every $x \in S$, and therefore $\mathcal{C} \sqsubset \mathcal{R}$. We conclude that $\mathcal{C}$ is the minimum of the upper bounds of Γ, i.e., the supremum of Γ. ■

71D↑. Proposition. (a): *Let the family of completely ordered sets $((A_i; \prec_i) \mid i \in I)$ be given. Then its product is completely ordered. If B is a subset of $\bigtimes_{i\in I} A_i$, then*

$$\sup B = (\sup(\pi_i)_>(B) \mid i \in I).$$

(b): *Let the set S and the completely ordered set D be given. Then the set $\mathrm{Map}(S, D)$ is completely ordered by the valuewise order. If F is a subset of $\mathrm{Map}(S, D)$, then*

$$(\sup F)(s) = \sup\{f(s) \mid f \in F\} \qquad \text{for all } s \in S.$$

Proof. This follows at once from Propositions 63D↑,(c) and 71B↑. ■

Let D be a completely ordered set, and E a subset of D. There are some useful criteria that ensure that the ordered subset E is itself completely ordered. We give one here (Proposition 71F); another will be found in Proposition 72E,(b).

A subset E of D is said to be **infimum-stable** if $\inf_D A \in E$ for every subset A of E. (This implies, in particular, that $\max D = \inf_D \emptyset \in E$, and $\inf_D E \in E$.) When the ordered set D happens to be $\mathfrak{P}(S)$ ordered by inclusion, for some set S (Remark 71A,(c)), the term *infimum-stable* may be replaced by the more specific term **intersection-stable**. Since infimum-stable subsets of completely ordered sets occur with great frequency in mathematics, it is convenient to learn how to make new ones out of given ones; one way is to construct intersections, as the next proposition shows.

71E. Proposition. *Let the completely ordered set D be given. Then the collection of all infimum-stable subsets of D is an intersection-stable subcollection of $\mathfrak{P}(D)$.*

71F. Proposition. *Let the completely ordered set D and the infimum-stable subset E of D be given. Then the ordered subset E is completely ordered, and $\inf_E A = \inf_D A$ for all subsets A of E.*

Proof. For every subset A of E, we have $\inf_D A \in E$. By Proposition 61B,(b), A has an infimum with respect to E, and $\inf_E A = \inf_D A$. By Proposition 71B↓, the ordered subset E is completely ordered. ■

71G. Remarks. (a): Under the assumptions of Proposition 71F, it is by no means the case in general that $\sup_E A = \sup_D A$.

(b): In every ordered set, the collection of order-convex sets is intersection-stable.

(c): Let the set D be given, and consider the set of all relations in D, completely ordered by the relation "narrower than", and the corresponding collection $\mathfrak{P}(D \times D)$, ordered by inclusion, of their graphs (cf. Remark 71A,(d)). For each of the following properties of relations in D, the collection of graphs of all relations having this property is intersection-stable: reflexive, symmetric, transitive. (But neither antisymmetric nor total, unless D is empty or a singleton, since $D \times D$ is not the graph of an antisymmetric relation, and the intersection of the graphs of all total ones is empty.) Consequently, the set of all reflexive relations in D is infimum-stable, as are the sets of all symmetric and of all transitive relations in D. By Proposition 71E, the set of all relations in D having any given combination of these properties is also infimum-stable. In particular, the set of all equivalence relations in D is infimum-stable. ■

We recall that every order-interval in an ordered set D is an order-convex subset of D. We now examine the converse implication when D is totally ordered.

71H. Proposition. *Let the totally ordered set D be given. Then D is completely ordered if and only if*

(71.1) *every order-convex subset of D is an order-interval.*

Proof. Proof of the "only if" part. Assume that D is completely ordered. Let the order-convex subset S of D be given. If $S = \emptyset$, then $S = [\![\min D, \min D[\![$, an order-interval. Assume now that $S \neq \emptyset$, and set $a := \inf S$, $b := \sup S$. Then $a \prec b$ and

$$S \subset [\![a, b]\!]. \tag{71.2}$$

Let $z \in]\!]a, b[\![$ be given. Then z is neither a lower nor an upper bound of S; since D is totally ordered, we may choose $x, y \in S$ such that $x \prec z \prec y$. Since S is order-convex, it follows that $z \in S$. We have shown that

$$S \supset]\!]a, b[\![. \tag{71.3}$$

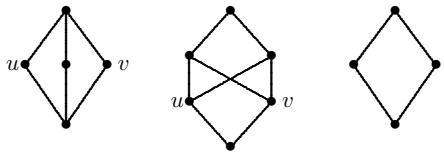
It follows from (71.2) and (71.3) that S is one of the order-intervals $[\![a, b]\!]$, $[\![a, b[\![$, $]\!]a, b]\!]$, $]\!]a, b[\![$.

Proof of the "if" part. Assume that D satisfies (71.1). There are no order-intervals in an empty ordered set, but D itself is an order-convex subset of D. By (71.1), D is not empty and is an order-interval. Therefore D has a minimum and a maximum.

Let the subset A of D be given. Then $\mathrm{Ub}(A)$ is order-convex and contains $\max D$. By (71.1), $\mathrm{Ub}(A)$ is a non-empty order-interval. By Proposition 61G↓, $\mathrm{Ub}(A)$ has an infimum, and by Proposition 61C,(d) this infimum is the supremum of A.

We have shown that every subset of D has a supremum. By Proposition 71B↑ it follows that D is completely ordered. ■

71I. Remark. The ordered sets described by the graphic representations shown below (cf. Examples 61D,(a)) illustrate the relationships among the conditions "D is totally ordered", "D is completely ordered", "D satisfies (71.1)". None of the three ordered sets is totally ordered.



(A) (B) (C)

(A) represents an ordered set D that is completely ordered, but fails to satisfy (71.1): the subset $\{u, v\}$ is order-convex, but is not an order-interval. (B) represents an ordered set D that satisfies (71.1), as may be verified by a tedious but straightforward accounting; but D is not completely ordered: the subset $\{u, v\}$ has no supremum. (C) represents an ordered set D that is completely ordered and satisfies (71.1); these two conditions together do not imply that D is totally ordered; see, however, the next result. ■

71J. Proposition. *Let the densely ordered set D be given, and assume that D satisfies* (71.1). *Then D is totally and completely ordered.*

Proof. Suppose that D is not totally ordered. We may then choose $x, y \in D$ such that x and y are not $\prec$-comparable. Then $\{x, y\}$ is an order-convex subset of D; by (71.1) we have $\{x, y\} =]\!]a, b[\![$ for suitable $a, b \in D$ with $a \precneqq x \precneqq b$. Since D is densely ordered, we may choose $z \in]\!]a, x[\![$. But then $z \in]\!]a, b[\![\setminus \{x, y\} = \emptyset$, which is impossible. We conclude that D is totally ordered. It follows from Proposition 71H that D is completely ordered. ■

72. Pre-completely ordered sets

Some very important ordered sets are not completely ordered, but have nearly the same properties. An example is $\mathbb{R}$ ordered by $\leq$, which has neither a maximum nor a minimum; all its closed order-intervals, however, considered as ordered subsets, are completely ordered. We therefore introduce a new concept. A set D ordered by $\prec$ is said to be **pre-completely ordered** (**by** $\prec$) if every non-empty subset of D that has an upper bound has a supremum and every non-empty subset of D that has a lower bound has an infimum. (Instead of the term *pre-completely ordered* one encounters *conditionally completely ordered* or *relatively completely ordered.* Some mathematicians prefer to use the term *completely ordered* for this concept, and make appropriate modifications when referring to completely ordered sets as defined in these notes.)

72A. Remarks. (a): The empty set is pre-completely ordered by its only relation.

(b): Every completely ordered set is pre-completely ordered (see also Proposition 72E,(a)).

(c): If the set D is pre-completely ordered by $\prec$, then it is also pre-completely ordered by the reverse order $\succ$. The convention introduced in Remark 61F is therefore in force for propositions concerning pre-completely ordered sets. ■

In the definition of a pre-completely ordered set, either half of the defining condition is redundant, as we now show.

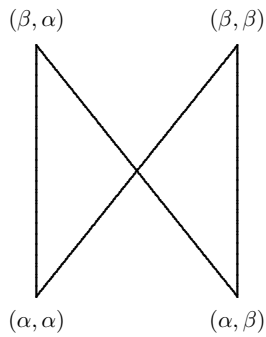
72B↑. Proposition. *An ordered set D is pre-completely ordered if (and only if) every non-empty subset of D that has an upper bound has a supremum.*

Proof. Assume that every non-empty subset of D that has an upper bound has a supremum. Let the non-empty subset A of D be given and assume that A has a lower bound. Then $\mathrm{Lb}(A)$ is a non-empty subset of D that has an upper bound (any member chosen in the non-empty subset A), and therefore a supremum, say $s := \sup \mathrm{Lb}(A)$. It now follows by Proposition 61C,(d) that A has an infimum (namely s). ■

If D is a pre-completely ordered set, then every non-empty order-bounded subset of D has both a supremum and an infimum. One might be tempted to conjecture that the converse is true; but we now produce a counterexample. We shall show later that the conjecture is right when D is totally ordered (Proposition 72J).

72C. Example. Choose distinct objects α and β, and define the relation $\prec$ in $\{\alpha,\beta\}\times\{\alpha,\beta\}$ by the rule

$$\forall (u,v),(u',v')\in\{\alpha,\beta\}\times\{\alpha,\beta\},\quad (u,v)\prec(u',v')\ :\Leftrightarrow\ (u=\alpha \text{ and } u'=\beta).$$



It is trivial that $\prec$ is a strict-order. Consider set $\{\alpha,\beta\}\times\{\alpha,\beta\}$ ordered by $\preceq$ (represented in the figure in the manner described in Example 61D,(a)). It is easily verified that each non-empty order-bounded subset of $\{\alpha,\beta\}\times\{\alpha,\beta\}$ has both a supremum and an infimum (indeed, a maximum and a minimum). The set $\{(\alpha,\alpha),(\alpha,\beta)\}$, however, has two upper bounds but no supremum. Hence $\{\alpha,\beta\}\times\{\alpha,\beta\}$ is not pre-completely ordered by $\preceq$. ■

72D↑. Proposition. *The assertion of Proposition* 71D↑ *remains valid if "completely ordered" is replaced everywhere in it by "pre-completely ordered".*

Proof. This follows at once from Propositions 63D↑,(c) and 72B↑. ■

We next examine several ways in which pre-completely ordered sets are related to completely ordered sets.

72E. Proposition. *Let the set D pre-completely ordered by $\prec$ be given. Then:*

(a): *D is completely ordered by $\prec$ if (and only if) D has a maximum and a minimum.*

(b): If $a,b\in D$ and $a\prec b$, then the ordered subset $[\![a,b]\!]$ is completely ordered, and the ordered subsets $[\![a,b[\![$, $]\!]a,b]\!]$, $]\!]a,b[\![$ are pre-completely ordered.

Proof. Proof of (a). Assume that D has a maximum and a minimum. Let the subset A of D be given. If $A\neq\emptyset$, then $\max D\in \mathrm{Ub}(A)$, and therefore A has a supremum. If, on the other hand, $A=\emptyset$, then $\sup_D A=\min D$. In either case, A has a supremum. The assertion follows from Proposition 71B↑.

Proof of (b): Let E be any one of the order-intervals specified. Let A be a non-empty subset of E, and choose $y\in A$. If $\mathrm{Ub}_E(A)=E\cap\mathrm{Ub}_D(A)$ is not empty, choose $z\in\mathrm{Ub}_E(A)$. By the assumption, A has a supremum (with respect to D) and $y\prec\sup_D A\prec z$; since $y,z\in E$, we have $\sup_D A\in E$. by Proposition 61B,(b) it follows that A has a supremum in E (and in fact $\sup_E A=\sup_D A$).

Since A was an arbitrary non-empty subset of E, it follows from Proposition 72B↑ that the ordered subset E is pre-completely ordered. If $E=[\![a,b]\!]$, it follows from (a) that the ordered subset E is completely ordered. ■

72F. Remarks. (a): It follows from the proof of Proposition 72E,(b) and from Remark 72A,(c) that if E is an order-interval in the pre-completely ordered set D, then every non-empty subset A of E that has a supremum [infimum] with respect to E satisfies $\sup_E A = \sup_D A$ [$\inf_E A = \inf_D A$].

(b): Example 72C is an instance of an ordered set in which every closed interval is completely ordered; yet the ordered set itself is not pre-completely ordered. There is therefore no converse to Proposition 72E,(b); but see Proposition 72J. ■

72G. Theorem. *Let the ordered set $(D;\prec)$ be given. Then $(D;\prec)$ is pre-completely ordered if and only if there exists a completely ordered set $(D';\prec')$ such that $(D;\prec)$ is order-isomorphic to the ordered subset $]\!]\min D', \max D'[\!]$ of $(D';\prec')$.*

Proof. Proof of the "if" part. A completely ordered set $(D';\prec')$ is pre-completely ordered (Remark 72A,(b)), and the ordered subset $]\!]\min D', \max D'[\!]$ is also pre-completely ordered (Proposition 72E,(b)). If $(D;\prec)$ is order-isomorphic to this ordered subset, it is itself pre-completely ordered.

Proof of the "only if" part. Assume that $(D;\prec)$ is pre-completely ordered. Choose distinct objects α and ω, neither of which is a member of D, and set $D' := D \cup \{\alpha, \omega\}$. Define the relation $\prec'$ in D' by the rule

$$\forall x, y \in D', \quad x \prec' y \ :\Leftrightarrow\ (x = \alpha \text{ or } y = \omega \text{ or } (x, y \in D \text{ and } x \prec y)).$$

It is easy to verify that $\prec'$ is an order in D'. In the ordered set $(D';\prec')$ we have $\alpha = \min D'$, $\omega = \max D'$, and $D =]\!]a, \omega[\!] =]\!]\min D', \max D'[\!]$. It is also clear that 1_D is an order-isomorphism from $(D;\prec)$ to the ordered subset $]\!]\min D', \max D'[\!]$ of $(D';\prec')$. It remains to prove that $(D';\prec')$ is completely ordered.

Let the subset A of D' be given. If $\omega \in A$, then $\sup_{D'} A = \max A = \omega$. If $A \subset \{\alpha\}$, then $\sup_{D'} A = \alpha$. Suppose now that A does not contain ω and is not a subset of $\{\alpha\}$. Then $A \cap D \neq \emptyset$. If $A \cap D$ has no upper bound in $(D;\prec)$, then $\omega \in \mathrm{Ub}_{D'}(A) \subset \mathrm{Ub}_{D'}(A \cap D) = \{\omega\}$; hence $\mathrm{Ub}_{D'}(A) = \{\omega\}$, and $\sup_{D'} A = \omega$. If, on the other hand, $A \cap D$ does have an upper bound in $(D;\prec)$, which is the ordered subset $]\!]\alpha, \omega[\!]$ of $(D';\prec')$, it has a supremum there. Since $\omega \notin A$, we have $\mathrm{Ub}_{D'}(A) = \mathrm{Ub}_{D'}(A \cap D) = [\![\sup_D(A \cap D), \omega]\!]$. Hence $\sup_D(A \cap D)$ is the minimum of $\mathrm{Ub}_{D'}(A)$, so that $\sup_D(A \cap D) = \sup_{D'} A$. We have shown that every subset A of D' has a supremum in $(D';\prec')$. By Proposition 71B↑, we conclude that $(D';\prec')$ is completely ordered. ■

72H↑. Remark. Let the ordered set $(D;\prec)$ be given. If D has a maximum, the assertion of Theorem 72G remains valid if $]\!]\min D', \max D'[\!]$ is replaced by $]\!]\min D', \max D'[\!]$ (just set $D' := D \cup \{\alpha\}$ in the proof). If D has both a maximum and a minimum, the assertion remains valid with $]\!]\min D', \max D'[\!]$ replaced by $[\![\min D', \max D']\!]$, i.e., D' itself; but this is the same as Proposition 72E,(a). ■

72I. Examples. (a)*: The set $\mathbb{N}$ is pre-completely ordered by $\leq$. $\mathbb{N}$ has a minimum namely 0, but no maximum. By choosing an object ω that is not a member of $\mathbb{N}$ and requiring $n \leq \omega$ for all $n \in \mathbb{N}$, as well as $\omega \leq \omega$, we obtain a set $\mathbb{N} \cup \{\omega\}$ completely ordered by $\leq$ (Remark 72H↓).

(b)*: The set $\mathbb{R}$ is pre-completely ordered by $\leq$; $\mathbb{R}$ has neither maximum nor

minimum. By choosing distinct objects $-\infty,\ \infty$, neither of which is a member of $\mathbb{R}$, and requiring $-\infty \leq r \leq \infty$ for all $r \in \mathbb{R}$, as well as $-\infty \leq -\infty \leq \infty \leq \infty$, we obtain a set $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ completely ordered by $\leq$. ■

For totally ordered sets, the condition for pre-completeness is easier to express.

72J. Proposition. *Let the totally ordered set D be given. The following statements are equivalent*:

(i): *D is pre-completely ordered.*

(ii): *Every non-empty order-bounded subset of D has both a supremum and an infimum.*

(iii): *Every closed order-interval of D is completely ordered.*

(iv): *Every order-convex subset of D is either an order-interval; or D itself; or* $\mathrm{Ub}(\{c\})$ *or* $\mathrm{Ub}(\{c\})\backslash\{c\}$ *or* $\mathrm{Lb}(\{c\})$ *or* $\mathrm{Lb}(\{c\})\backslash\{c\}$ *for some $c \in D$.*

Proof. (i) ⇒ (ii). This implication is trivial.

(ii) ⇒ (iii). Assume that (ii) holds. Let $a, b \in D$ with $a \prec b$ be given, and let the subset A of $[\![a, b]\!]$ be given. If $A = \emptyset$, then $\sup_{[\![a,b]\!]} A = a$. We now assume that $A \neq \emptyset$. Since A has the upper bound b and the lower bound a, A has a supremum $\sup_D A$, and $\sup_D A \preceq b$. Since $\emptyset \neq A \subset [\![a, b]\!]$, we also have $a \preceq \sup_D A$. Thus $\sup_D A \in [\![a, b]\!]$. By Proposition 61B,(b), we find that A has a supremum with respect to $[\![a, b]\!]$. By Proposition 71B↑, the ordered subset $[\![a, b]\!]$ is completely ordered.

(iii) ⇒ (i). Assume that (iii) holds. Let the non-empty subset A of D be given, and assume that A has an upper bound, say b. Choose $a \in A$. Then the ordered subset $[\![a, b]\!]$ is completely ordered. Therefore $A \cap [\![a, b]\!]$ has a supremum with respect to $[\![a, b]\!]$, say $s := \sup_{[\![a,b]\!]}(A \cap [\![a, b]\!])$. We claim that s is the supremum of A with respect to D. Since A was an arbitrary non-empty subset of D having an upper bound, it will then follow from Proposition 72B↑ that D is pre-completely ordered, as asserted.

To establish our claim concerning s, we use the assumption that D is totally ordered. Let $x \in A$ be given; then $x \prec b$, and $a \prec x$ or $x \prec a$. If $a \prec x \prec b$, then $x \in A \cap [\![a, b]\!]$, so that $x \prec s$; if, on the other hand $x \prec a$, then $x \prec a \prec s$. In either case we have found $x \prec s$. Since $x \in A$ was arbitrary, we conclude that $s \in \mathrm{Ub}_D(A)$.

Let now $y \in \mathrm{Ub}_D(A)$ be given. Then $a \prec y$, and $b \prec y$ or $y \prec b$. If $b \prec y$, then $s \prec b \prec y$; if, on the other hand, $a \prec y \prec b$, then $y \in [\![a, b]\!] \cap \mathrm{Ub}_D(A) = \mathrm{Ub}_{[\![a,b]\!]}(A)$, so that $s \prec y$. In either case we have found $s \prec y$. Since $y \in \mathrm{Ub}_D(A)$ was arbitrary, we conclude that $s = \min \mathrm{Ub}_D(A) = \sup_D A$, and our claim is established.

(i) ⇒ (iv). Assume that D is pre-completely ordered. By Theorem 72G we may assume that D is the ordered subset $]\!]\min D', \max D'[\![$ of a completely ordered set D'; since D is totally ordered, so is D'. By Proposition 71H, every order-convex subset of D is an order-interval of D' that is included in D. It is an easy, though somewhat tedious, task to verify that the order-intervals of D' that are included in D are precisely the subsets of D listed in (iv).

(iv) ⇒ (iii). Let $a, b \in D$ be given, and assume that $a \prec b$. We claim that every order-convex subset of the totally ordered closed order-interval $[\![a, b]\!]$ is an order-interval of $[\![a, b]\!]$; it will follow by Proposition 71H that $[\![a, b]\!]$ is completely ordered.

This will establish (iii).

Let the order-convex subset S of $[\![a, b]\!]$ be given. Then S is an order-convex subset of D and hence, by (iv), one of the sets listed in that statement. Careful analysis, with attention to the fact that D is totally ordered, shows that every one of these sets, when included in $[\![a, b]\!]$, is actually an order-interval. For instance, if $S = \mathrm{Ub}(\{c\})\backslash\{c\}$ for a suitable $c \in D$, then either $S = \emptyset = [\![a, a[\![$, or else $S \neq \emptyset$ and $c \prec b$ and $b = \max D$ and $S = [\![a, b]\!]$ or $S =]\!]c, b]\!]$, this last according as $c \precneqq a$ or $a \prec c$. ■

73. Closure mappings

Throughout this section we consider a set D completely ordered by $\prec$.

An isotone mapping $\omega \in \mathrm{Isot}(D, D)$ is called a **closure mapping** (**in** D) if it satisfies the following conditions:

(Aug) : $x \prec \omega(x)$ for all $x \in D$,

(Idp) : $\omega \circ \omega = \omega$.

A mapping $\omega : D \to D$ that satisfies (Aug) may be said to be **augmenting**. (The term most frequently used for this is *increasing*, but that use may produce a clash in certain contexts in which the same term has traditionally been used to mean "[strictly] isotone".) This condition may be rephrased as $1_D \prec \omega$ in the valuewise order of $\mathrm{Map}(D, D)$. We recall that a mapping satisfying (Idp) is said to be *idempotent*.

73A. Proposition. *Let the set D completely ordered by $\prec$ be given. A mapping $\omega : D \to D$ is a closure mapping if and only if it satisfies* (Aug) *and*

$$x \prec \omega(y) \;\Rightarrow\; \omega(x) \prec \omega(y) \quad \textit{for all } x, y \in D. \tag{73.1}$$

73B. Corollary. *The set of all closure mappings in a completely ordered set D is an infimum-stable subset of the set* $\mathrm{Map}(D, D)$ *completely ordered by the valuewise order.*

73C. Examples. (a): Let $a \in D$ be given, and define $\omega : D \to D$ by the rule $\omega(x) := \sup\{a, x\}$. Then ω is a closure mapping.

(b): Let the set S and the relation ρ in S be given. Then the mapping $\rho_>$ is always isotone from $\mathfrak{P}(S)$ completely ordered by inclusion to itself, but it is a closure mapping in $\mathfrak{P}(S)$ if and only if ρ is reflexive and transitive. The same assertion holds for the mapping $\rho^<$. ■

The fundamental result about closure mappings is a characterization of their ranges and the sets of their fixed points.

73D. Theorem. *Let a completely ordered set D be given.*

(a): *If $\omega : D \to D$ is a closure mapping, then* $\mathrm{Rng}\omega$ *is infimum-stable and*

$$\mathrm{Rng}\omega = \mathrm{Fix}\omega. \tag{73.2}$$

(b): *If E is an infimum-stable subset of D, then the mapping* $\mathrm{clos}_E : D \to D$ *defined by the rule*

$$\mathrm{clos}_E(x) := \inf(E \cap \mathrm{Ub}(\{x\})) \qquad \textit{for all } x \in D$$

is a closure mapping.

(c): *If $\omega : D \to D$ is a closure mapping and E is an infimum-stable subset of D, then* $\mathrm{Rng}\omega = E$ *if and only if* $\mathrm{clos}_E = \omega$.

Proof. Proof of (a). Since ω is idempotent, (73.2) holds, by Proposition 26C.

Let a subset A of $\mathrm{Rng}\omega$ be given. By (73.2), $\omega_>(A) = A$. Since $\inf A \in \mathrm{Lb}(A)$ and ω is isotone, we have $\omega(\inf A) \in \mathrm{Lb}(\omega_>(A)) = \mathrm{Lb}(A)$, whence $\omega(\inf A) \prec \inf A$. Since

ω is augmenting, we have $\inf A \prec \omega(\inf A)$; consequently, $\inf A = \omega(\inf A) \in \mathrm{Rng}\omega$. Since A was an arbitrary subset of $\mathrm{Rng}\omega$, we conclude that $\mathrm{Rng}\omega$ is infimum-stable.

Proof of (b). For each $x \in D$, we have $x = \inf \mathrm{Ub}(\{x\}) \prec \inf(E \cap \mathrm{Ub}(\{x\})) = \mathrm{clos}_E(x)$, so that $\omega := \mathrm{clos}_E$ satisfies (Aug).

Let $x, y \in D$ be given. If $x \prec \mathrm{clos}_E(y) = \inf(E \cap \mathrm{Ub}(\{y\}))$, it follows that $E \cap \mathrm{Ub}(\{y\}) \subset \mathrm{Ub}(\{x\})$, and hence that $E \cap \mathrm{Ub}(\{y\}) \subset E \cap \mathrm{Ub}(\{x\})$; from this we obtain $\mathrm{clos}_E(x) \prec \mathrm{clos}_E(y)$. Since $x, y \in D$ were arbitrary, we conclude that $\omega := \mathrm{clos}_E$ satisfies (73.1). By Proposition 73A, clos_E is a closure mapping.

Proof of (c). Assume first that $\mathrm{Rng}\omega = E$, and let $x \in D$ be given. We have $\omega(x) \in E \cap \mathrm{Ub}(\{x\})$, since ω is augmenting, so that $\mathrm{clos}_E(x) \prec \omega(x)$. On the other hand, for every $y \in E \cap \mathrm{Ub}(\{x\})$ we have $\omega(y) = y$ (from (a)) and $x \prec y$, and therefore $\omega(x) \prec \omega(y) = y$. If follows that $\omega(x) \in \mathrm{Lb}(E \cap \mathrm{Ub}(\{x\}))$, and therefore $\omega(x) \prec \mathrm{clos}_E(x)$. We have shown that $\mathrm{clos}_E(x) = \omega(x)$ for all $x \in D$, i.e., that $\mathrm{clos}_E = \omega$.

Assume, conversely, that $\mathrm{clos}_E = \omega$, and let $x \in D$ be given. If $x \in E$, then $\omega(x) = \inf(E \cap \mathrm{Ub}(\{x\})) = x$. On the other hand, if $\omega(x) = x$, then $x = \mathrm{clos}_E(x) = \inf(E \cap \mathrm{Ub}(\{x\})) \in E$, since E is infimum-stable. Since $x \in D$ was arbitrary, we conclude that $E = \mathrm{Fix}\omega$. It now follows from (a) that $\mathrm{Rng}\omega = E$. ■

73E. Remark. Theorem 73D,(c) asserts that each of the mappings $\omega \mapsto \mathrm{Rng}\omega$ and $E \mapsto \mathrm{clos}_E$ between the set of all closure mappings in D and the collection of all infimum-stable subsets of D is the inverse of the other. ■

73F. Examples. (a)*: Consider the set $\bar{\mathbb{R}}$, completely ordered by $\leq$, defined in Example 72I,(b) (where $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$), and the infimum-stable subset $\bar{\mathbb{Z}} := \mathbb{Z} \cup \{-\infty, +\infty\}$. The mapping $\mathrm{clos}_{\bar{\mathbb{Z}}}|^{\mathbb{R}}_{\mathbb{R}}$ is the *ceiling-function* $\lceil\colon \mathbb{R} \to \mathbb{R}$, defined by the rule $\lceil t := \min\{n \in \mathbb{Z} \mid t \leq n\}$ for all $t \in \mathbb{R}$. The more familiar *floor-function* $\lfloor\colon \mathbb{R} \to \mathbb{R}$ defined by the rule $\lfloor t := \max\{n \in \mathbb{Z} \mid n \leq t\}$ for all $t \in \mathbb{R}$ is obtained by a similar adjustment from the closure mapping associated with $\bar{\mathbb{Z}}$ in the set $\bar{\mathbb{R}}$ completely ordered by the reverse order $\geq$.

(b)*: Consider the collection $\mathfrak{P}(\mathbb{R})$ completely ordered by inclusion. The collection $\mathcal{I}$ of all *intervals* in $\mathbb{R}$, i.e., order-convex subsets of $\mathbb{R}$, is intersection-stable (Remark 71G,(b)). The mapping $\mathrm{clos}_{\mathcal{I}}$ assigns to each subset of $\mathbb{R}$ the smallest interval including it. For instance, if A is a subset of $\mathbb{R}$ that has a maximum but no minimum, then $\mathrm{clos}_{\mathcal{I}}(A)$ is $]\inf A, \max A]$ or $]-\infty, \max A]$, according as A has or fails to have a lower bound. ■

73G. Corollary. *Let the completely ordered set D and the closure mappings ω and ω' in D be given. Then $\omega' \circ \omega$ is a closure mapping if and only if $\omega'_>(\mathrm{Rng}\omega) \subset \mathrm{Rng}\omega$. This is the case, in particular, if ω and ω' commute.*

Proof. Assume first that $\omega' \circ \omega$ is a closure mapping. For every $x \in \omega'_>(\mathrm{Rng}\omega) = \mathrm{Rng}(\omega' \circ \omega)$ we have, by Theorem 73D,(a), $x \prec \omega(x) \prec \omega'(\omega(x)) = x$, so that $x = \omega(x) \in \mathrm{Rng}\omega$. This proves that $\omega'_>(\mathrm{Rng}\omega) \subset \mathrm{Rng}\omega$.

Assume, conversely, that $\omega'_>(\mathrm{Rng}\omega) \subset \mathrm{Rng}\omega$. Now $\omega' \circ \omega$ is obviously isotone and augmenting. For every $x \in D$ we have $\omega'(\omega(x)) \in \mathrm{Rng}\omega$, and Theorem 73D,(a) implies $\omega(\omega'(\omega(x))) = \omega'(\omega(x))$. Therefore $\omega'(\omega(\omega'(\omega(x)))) = \omega'(\omega'(\omega(x))) = \omega'(\omega(x))$. We

have shown that $\omega' \circ \omega$ is also idempotent; hence it is a closure mapping. ■

Let a set S be given. It was pointed out in Remark 71G,(c) that the set of all transitive relations in S is an infimum-stable subset of the set of all relations in S, completely ordered by the order "narrower than". By Theorem 73D,(b), there is a closure mapping associated with this infimum-stable subset. If ρ is a relation in S, the value of this closure mapping at ρ is called the **transitive closure of** ρ; it is the narrowest of all transitive relations that are broader than ρ.

73H. Proposition. *Let ρ be a relation in the set S, and let τ be the transitive closure of ρ. Then*

$$\forall x, y \in S, \quad x \;\tau\; y \;\Leftrightarrow\; (x \;\rho\; y \text{ or } (\exists z \in S,\; x \;\tau\; z \text{ and } z \;\rho\; y)) \tag{73.3}$$

$$\forall x, y \in S, \quad x \;\tau\; y \;\Leftrightarrow\; (x \;\rho\; y \text{ or } (\exists z \in S,\; x \;\rho\; z \text{ and } z \;\tau\; y)). \tag{73.4}$$

Proof. Define the relation σ in S by the rule

$$\forall x, y \in S, \quad x \;\sigma\; y \;:\Leftrightarrow\; (x \;\rho\; y \text{ or } (\exists z \in S,\; x \;\tau\; z \text{ and } z \;\rho\; y)).$$

Since ρ is narrower than τ, and τ is transitive, we have for all $x, y \in S$ the chain of implications

$$x \;\sigma\; y \;\Rightarrow\; (x \;\tau\; y \text{ or } (\exists z \in S,\; x \;\tau\; z \text{ and } z \;\tau\; y)) \Rightarrow\; x \;\tau\; y,$$

so that σ is narrower than τ.

For all $x, y, z \in S$ we have the chain of implications

$$(x \;\sigma\; y \text{ and } y \;\sigma\; z) \;\Rightarrow\; \left\{ \begin{array}{l} x \;\tau\; y \text{ and } y \;\rho\; z, \text{or} \\ \\ x \;\tau\; y \text{ and } (\exists w \in S,\;\; y \;\tau\; w \text{ and } w \;\rho\; z) \end{array} \right\} \Rightarrow$$

$$\Rightarrow \left\{ \begin{array}{l} x \;\tau\; y \text{ and } y \;\rho\; z, \text{ or} \\ \\ \exists w \in S, \quad x \;\tau\; w \text{ and } w \;\rho\; z \end{array} \right\} \Rightarrow x \;\sigma\; z.$$

Therefore σ is transitive. Since σ is obviously broader than ρ, it follows that σ is also broader than the transitive closure τ of ρ. Hence $\sigma = \tau$, and (73.3) is proved.

The proof of (73.4) is completely similar. It may also be obtained by applying (73.3) to the reverse relation $\rho^{\leftarrow}$ instead of to ρ; indeed, relation-reversal is an involutory order-isomorphism of the set of all relations on D, ordered by "narrower than", to itself, and it preserves transitivity (cf. Proposition 55C). ■

We conclude this section with a technical result describing the relationship between suprema and infima with respect to D and with respect to some infimum-stable subset E. We recall that, according to Proposition 71F, the ordered subset E is completely ordered.

73I. Proposition. *Let the completely ordered set D and the infimum-stable subset E of D be given. Then:*

(a): *For every subset A of D we have* $\mathrm{clos}_E(\sup_D A) = \sup_E(\mathrm{clos}_E)_>(A)$ *and* $\mathrm{clos}_E(\inf_D A) \prec \inf_E(\mathrm{clos}_E)_>(A)$.

(b): *For every subset A of E,* $\sup_E A = \mathrm{clos}_E(\sup_D A)$ *and* $\inf_E A = \inf_D A$.

Proof. Proof of (a). Since clos_E is augmenting, Proposition 61B,(a) yields $\sup_D A \prec \sup_D(\mathrm{clos}_E)_>(A) \prec \sup_E(\mathrm{clos}_E)_>(A)$. By Theorem 73D,$E = \{x \in D \mid \mathrm{clos}_E(x) = x\}$; since clos_E is isotone, we find

$$\mathrm{clos}_E(\sup_D A) \prec \mathrm{clos}_E(\sup_E(\mathrm{clos}_E)_>(A)) = \sup_E(\mathrm{clos}_E)_>(A).$$

On the other hand, for each $x \in A$ we have $x \prec \sup_D A$, whence $\mathrm{clos}_E(x) \prec \mathrm{clos}_E(\sup_D A) \in E$; therefore $\mathrm{clos}_E(\sup_D A) \in \mathrm{Ub}_E((\mathrm{clos}_E)_>(A))$, whence $\sup_E(\mathrm{clos}_E)_>(A) \prec \mathrm{clos}_E(\sup_D A)$. We conclude that equality must hold.

For each $x \in A$ we have $\inf_D A \prec x$, whence $\mathrm{clos}_E(\inf_D A) \prec \mathrm{clos}_E(x)$; since $\mathrm{Rng}\,\mathrm{clos}_E = E$, we then have $\mathrm{clos}_E(\inf_D A) \in \mathrm{Lb}_E((\mathrm{clos}_E)_>(A))$, whence $\mathrm{clos}_E(\inf_D A) \prec \inf_E(\mathrm{clos}_E)_>(A)$.

Proof of (b). Since $A \subset E$ we have $(\mathrm{clos}_E)_>(A) = A$. Part (a) then yields $\mathrm{clos}_E(\sup_D A) = \sup_E A$. The equality $\inf_E A = \inf_D A$ follows from Proposition 71F. ■

74. Galois correspondences

Let the sets D and D', completely ordered by $\prec$ and $\prec'$, respectively, be given. A pair of mappings $(\phi, \phi') \in \mathrm{Map}(D, D') \times \mathrm{Map}(D', D)$ is called a **Galois correspondence (from D to D')** if both ϕ and ϕ' are antitone and both $\phi' \circ \phi : D \to D$ and $\phi \circ \phi' : D' \to D'$ are augmenting.

74A. Remarks. (a): (ϕ, ϕ') is a Galois correspondence from D to D' if and only if (ϕ', ϕ) is a Galois correspondence from D' to D.

(b): A pair of mappings $(\phi, \phi') \in \mathrm{Map}(D, D') \times \mathrm{Map}(D', D)$ is a Galois correspondence if and only if both ϕ and ϕ' are antitone and

$$x \prec \phi'(x') \;\Leftrightarrow\; x' \prec' \phi(x) \quad \text{for all } (x, x') \in D \times D'. \tag{74.1}$$

■

74B. Example. For every mapping $f\colon S \to T$, the pair $(f_>, f^<)$ is a Galois correspondence from $\mathfrak{P}(S)$ ordered by inclusion to $\mathfrak{P}(T)$ ordered by the reverse of inclusion, on account of Proposition 23A. ■

The next proposition gives a complete account of all Galois correspondences from $\mathfrak{P}(S)$ to $\mathfrak{P}(T)$, both ordered by inclusion, for given sets S and T. Here we require the concept of a *relation from S to T* (Section 54).

74C. Proposition. *Let the sets S and T and the relation ρ from S to T be given. Then there exists exactly one Galois correspondence (Φ, Ψ) from $\mathfrak{P}(S)$ to $\mathfrak{P}(T)$, both ordered by inclusion, such that*

$$\Phi(\{x\}) = \rho_>(\{x\}) \qquad \textit{for all } x \in S; \tag{74.2}$$

it is defined by the rules

$$\Phi(U) := \bigcap_{x \in U}{}^{T} \rho_>(\{x\}) \qquad \textit{for all } U \in \mathfrak{P}(S) \tag{74.3}$$

$$\Psi(V) := \bigcap_{y \in V}{}^{S} \rho^<(\{y\}) \qquad \textit{for all } V \in \mathfrak{P}(T). \tag{74.4}$$

Proof. Let (Φ, Ψ) be a Galois correspondence satisfying (74.2). For every $x \in S$ and $V \in \mathfrak{P}(T)$ we have, by Remark 74A,(b), the chain of equivalences

$$\begin{aligned} x \in \Psi(V) \;\Leftrightarrow\; V \subset \Phi(\{x\}) \;\Leftrightarrow\; V \subset \rho_>(\{x\}) \;&\Leftrightarrow\; (\forall y \in V, \quad x \;\rho\; y) \Leftrightarrow \\ &\Leftrightarrow\; x \in \bigcap_{y \in V}{}^{S} \rho^<(\{y\}). \end{aligned}$$

Therefore (74.4) holds. In particular, $\Psi(\{y\}) = \rho^<(\{y\})$ for every $y \in T$. Using this instead of (74.2), we verify (74.3) in the same manner.

Conversely, the pair of mappings (Φ, Ψ) defined by (74.3) and (74.4) obviously satisfies (74.2), and it is a matter of direct verification that it is a Galois correspondence. ■

Remark. If the mapping $\Phi : \mathfrak{P}(S) \to \mathfrak{P}(T)$ is given, there is exactly one relation ρ from S to T satisfying (74.2). Proposition 74C therefore establishes a one-to-one correspondence between the set of all relations from S to T and the set of all Galois correspondences from $\mathfrak{P}(S)$ to $\mathfrak{P}(T)$, both ordered by inclusion.

74D. Examples. (a): Let the ordered set D be given. Then $(\mathrm{Ub}_D, \mathrm{Lb}_D)$ is a Galois correspondence from $\mathfrak{P}(D)$ ordered by inclusion, to itself.

(b): Let the set D be given. For each subset F of $\mathrm{Map}(D, D)$ we define the **commutator of** F to be

$$\mathrm{Comm}_D(F) := \{g \in \mathrm{Map}(D, D) \mid \forall f \in F, \quad g \circ f = f \circ g\}.$$

Then $(\mathrm{Comm}_D, \mathrm{Comm}_D)$ is the Galois correspondence associated with relation "commutes with" in $\mathrm{Map}(D, D)$ according to Proposition 74C.

(c): Let the set D and a subset Γ of $\mathrm{Map}(D, D)$ be given. Define the mappings $\Phi : \mathfrak{P}(\Gamma) \to \mathfrak{P}(D)$ and $\Psi : \mathfrak{P}(D) \to \mathfrak{P}(\Gamma)$ by the rules

$$\Phi(\Delta) := \bigcap_{f \in \Delta}^{D} \mathrm{Fix} f \quad \text{for all } \Delta \in \mathfrak{P}(\Gamma)$$

$$\Psi(A) := \{f \in \Gamma \mid A \subset \mathrm{Fix} f\} \quad \text{for all } A \in \mathfrak{P}(D).$$

Then (Φ, Ψ) is the Galois correspondence from $\mathfrak{P}(\Gamma)$ to $\mathfrak{P}(D)$ associated with the relation ρ from Γ to D give by

$$\forall (f, x) \in \Gamma \times D, \quad f \ \rho \ x \ \Leftrightarrow \ f(x) = x.$$

An instance of this construction plays a central part in the work of Évariste Galois (1811-1832) on polynomial equations; it is this instance that originated the term "Galois correspondence". ■

The fundamental result about Galois correspondences relates them to closure mappings.

74E. Theorem. *Let the completely ordered sets D and D' and the Galois correspondence (ϕ, ϕ') from D to D' be given. Then:*

(a): *$\phi \circ \phi' \circ \phi = \phi$ and $\phi' \circ \phi \circ \phi' = \phi'$.*

(b): *$\phi' \circ \phi$ is a closure mapping in D and $\phi \circ \phi'$ is a closure mapping in D'. Moreover,* $\mathrm{Rng}(\phi' \circ \phi) = \mathrm{Rng}\phi'$ *and* $\mathrm{Rng}(\phi \circ \phi') = \mathrm{Rng}\phi$.

(c): *Each of the antitone mappings $\phi|^{\mathrm{Rng}\phi}_{\mathrm{Rng}\phi'}$ and $\phi'|^{Rng\phi'}_{\mathrm{Rng}\phi}$ is the inverse of the other; they are order-antimorphisms.*

Proof. Proof of (a). Let the respective orders of D and D' be $\prec$ and $\prec'$. For every $x \in D$ we have $x \prec (\phi' \circ \phi)(x)$, and therefore $(\phi \circ \phi' \circ \phi)(x) = \phi((\phi' \circ \phi)(x)) \prec' \phi(x)$; on the other hand, $\phi(x) \prec' (\phi \circ \phi')(\phi(x)) = (\phi \circ \phi' \circ \phi)(x)$; hence equality must hold. Since $x \in D$ was arbitrary, we conclude that $\phi \circ \phi' \circ \phi = \phi$. The proof of $\phi' \circ \phi \circ \phi' = \phi'$ is similar.

Proof of (b). $\phi' \circ \phi$ is isotone, since it is the composite of the antitone mappings ϕ and ϕ'; it is augmenting by assumption. It is also idempotent, since (a) implies

$(\phi' \circ \phi) \circ (\phi' \circ \phi) = \phi' \circ (\phi \circ \phi' \circ \phi) = \phi' \circ \phi$. Thus $\phi' \circ \phi$ is a closure mapping in D. Moreover, $\mathrm{Rng}\phi' = \mathrm{Rng}(\phi' \circ \phi \circ \phi') \subset \mathrm{Rng}(\phi' \circ \phi) \subset \mathrm{Rng}\phi'$, so that equality holds. The proof for $\phi \circ \phi'$ is similar.

Proof of (c). By (b) and Theorem 73D,(a) we have

$$(\phi'|^{\mathrm{Rng}\phi'}_{\mathrm{Rng}\phi}) \circ (\phi|^{\mathrm{Rng}\phi}_{\mathrm{Rng}\phi'}) = (\phi' \circ \phi)|^{\mathrm{Rng}\phi'}_{\mathrm{Rng}\phi'} = (\phi' \circ \phi)|^{\mathrm{Rng}(\phi' \circ \phi)}_{\mathrm{Rng}(\phi' \circ \phi)} = 1_{\mathrm{Rng}(\phi' \circ \phi)} = 1_{\mathrm{Rng}\phi'},$$

and similarly for the composition in the reverse order. ■

▼ The following is a kind of converse of Theorem 74E.

74F. Proposition. *Let the completely ordered sets D and D' and the mapping ψ from the subset H of D to the subset H' of D' be given. Then there exists at most one Galois correspondence (ϕ, ϕ') from D to D' such that* $\mathrm{Rng}\phi = H'$, $\mathrm{Rng}\phi' = H$, *and* $\phi|^{H'}_{H} = \psi$. *This Galois correspondence exists if and only if H and H' are infimum-stable and ψ is an order-antimorphism; it is then given by* $\phi := (\psi|^{D'}) \circ (\mathrm{clos}_H|^H)$ *and* $\phi' := (\psi^{\leftarrow}|^D) \circ (\mathrm{clos}_{H'}|^{H'})$.

Theorem 74E,(b) shows that if (ϕ, ϕ') is a Galois correspondence, then $\phi' \circ \phi$ is a closure mapping. We now show that every closure mapping can be obtained in this way.

74G. Proposition. *Let the closure mapping ω in the completely ordered set D be given. Then there is a completely ordered set D' and a Galois correspondence (ϕ, ϕ') from D to D' such that $\omega = \phi' \circ \phi$.*

Proof. We choose D' to be the collection $\mathfrak{P}(\mathrm{Rng}\omega)$ completely ordered by inclusion and define $\phi : D \to \mathfrak{P}(\mathrm{Rng}\omega)$ and $\phi' : \mathfrak{P}(\mathrm{Rng}\omega) \to D$ by the rules

$$\begin{aligned} \phi(x) &:= \mathrm{Rng}\omega \cap \mathrm{Ub}(\{x\}) && \text{for all } x \in D \\ \phi'(A) &:= \inf A && \text{for all } A \in \mathfrak{P}(\mathrm{Rng}\omega). \end{aligned}$$

Then ϕ and ϕ' are obviously antitone. By Theorem 73D,

$$(\phi' \circ \phi)(x) = \inf(\mathrm{Rng}\omega \cap \mathrm{Ub}(\{x\})) = \mathrm{clos}_{\mathrm{Rng}\omega}(x) = \omega(x) \qquad \text{for all } x \in D.$$

Thus $\phi' \circ \phi = \omega$ and, in particular, $\phi' \circ \phi$ is augmenting. On the other hand, $A \subset \mathrm{Ub}(\{\inf A\})$ for all $A \in \mathfrak{P}(D)$. Therefore

$$A \subset \mathrm{Rng}\omega \cap \mathrm{Ub}(\{\inf A\}) = (\phi \circ \phi')(A) \qquad \text{for all } A \in \mathfrak{P}(\mathrm{Rng}\omega),$$

▲ so that $\phi \circ \phi'$ is also augmenting. We conclude that (ϕ, ϕ') is a Galois correspondence. ■

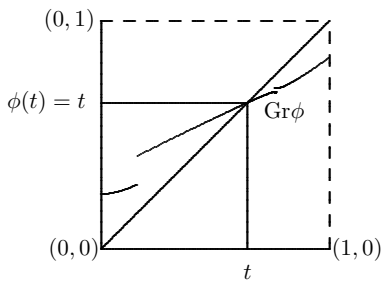
75. The fixed-point theorem for isotone mappings

Completely ordered sets play a large part in many branches of mathematics. Perhaps the best-known results for which the completeness of an ordered set is decisive are those that depend on the fact that closed order-intervals in $\mathbb{R}$ are completely ordered. Here we consider another important consequence of completeness of an ordered set: the existence of fixed points of isotone mappings.

75A. Theorem. (Knaster Fixed-Point Theorem). *Let the set D completely ordered by $\prec$ and the isotone mapping $\phi : D \to D$ be given. Then $u := \inf\{x \in D \mid \phi(x) \prec x\}$ is a fixed point of ϕ, i.e., $\phi(u) = u$.*

Proof. Set $E := \{x \in D \mid \phi(x) \prec x\}$. For all $x \in E$ we have $u = \inf E \prec x$, whence $\phi(u) \prec \phi(x) \prec x$; thus $\phi(u) \in \mathrm{Lb}(E)$, and therefore $\phi(u) \prec u$. This implies $\phi(\phi(u)) \prec \phi(u)$. Therefore $\phi(u) \in E$, and consequently $u = \inf E \prec \phi(u)$. The asserted equality $\phi(u) = u$ follows. ■

75B. Example*. Let the isotone mapping $\phi\colon [0,1] \to [0,1]$ be given. Then there is $t \in [0,1]$ such that $\phi(t) = t$. ■



We shall be using the Knaster Fixed-Point Theorem in the next chapter (Section 82). To give an immediate illustration of its power, however, we apply it to the proof of a fundamental theorem of set theory. (This proof was proposed independently by Irving Kaplansky (b. 1917) and by our late colleague Ignace Izaak Kolodner (1920-1996).)

75C. Theorem. (Schröder-Bernstein Theorem). *Let the sets X and Y be given. If there exists an injection from X to Y and also an injection from Y to X, then there exists a bijection from X to Y.*

Proof. Let the injections $g\colon X \to Y$ and $h\colon Y \to X$ be chosen. We plan to find a subset S of X and a subset T of Y such that

$$g_>(S) = Y \backslash T \qquad\qquad h_>(T) = X \backslash S. \tag{75.1}$$

Suppose we have subsets S and T satisfying (75.1). Then $g' := g|_S^{Y\setminus T}$ and $h' := h|_T^{X\setminus S}$ are bijections. We may therefore define the mappings $f\colon X \to Y$ and $f'\colon Y \to X$ by the rules

$$f(x) := \begin{cases} g'(x) & \text{if } x \in S \\ \\ h'^{\leftarrow}(x) & \text{if } x \in X\setminus S \end{cases} \qquad f'(y) := \begin{cases} h'(y) & \text{if } y \in T \\ \\ g'^{\leftarrow}(y) & \text{if } y \in Y\setminus T \end{cases}$$

and verify that $f' \circ f = 1_X$ and $f \circ f' = 1_Y$. Therefore f is invertible, hence a bijection.

It remains to produce S and T satisfying (75.1). We use the complementation mappings

$$C_X := (U \mapsto X\setminus U) : \mathfrak{P}(X) \to \mathfrak{P}(X) \qquad C_Y := (V \mapsto Y\setminus V) : \mathfrak{P}(Y) \to \mathfrak{P}(Y)$$

(Example 62G,(c)). With respect to the collections $\mathfrak{P}(X)$ and $\mathfrak{P}(Y)$, both completely ordered by inclusion, the mappings $g_> : \mathfrak{P}(X) \to \mathfrak{P}(Y)$ and $h_> : \mathfrak{P}(Y) \to \mathfrak{P}(X)$ are isotone, while C_X and C_Y are antitone. Therefore $\Phi := C_X \circ h_> \circ C_Y \circ g_> : \mathfrak{P}(X) \to \mathfrak{P}(X)$ is isotone. By Theorem 75A, we may choose $S \in \mathfrak{P}(X)$ such that $\Phi(S) = S$. If we set $T := Y\setminus g_>(S) = C_Y(g_>(S)) \in \mathfrak{P}(Y)$, we find

$$S = C_X(h_>(C_Y(g_>(S)))) = C_X(h_>(T)) = X\setminus h_>(T),$$

so that S and T satisfy (75.1). ■

Chapter 8

INDUCTION AND RECURSION

81. Proof by induction

In this chapter we shall explain precisely what is meant by the terms *proof by induction* and *recursive definition*, two procedures that are fundamental in all branches of mathematics. The structure underlying each instance of these procedures is a well-founded set.

We recall from Section 64 that an ordered set is said to be **well-founded** if every non-empty subset has at least one minimal member. As a special case of this, an ordered set is said to be **well-ordered** if every non-empty subset has a minimum. An ordered set is well-ordered if and only if it is both totally ordered and well-founded (Proposition 64A).

Let the set I ordered by $\prec$ be given. In this section and the next, an important part will be played by the set of all members of I that strictly precede a given member i of I; this is also the set of strict lower bounds of $\{i\}$. We denote this set by $\mathrm{Spr}(i)$, so that

$$\mathrm{Spr}(i) := \{j \in I \mid j \precneqq i\} = \precneqq^<(\{i\}) = \mathrm{Lb}(\{i\})\backslash\{i\} \quad \text{for all } i \in I.$$

The availability of proofs by induction rests on the following result.

81A. Proposition. *Let the ordered set I be given. The following statements are equivalent:*

(i): *I is well-founded;*

(ii): *the only subset J of I that satisfies*

$$\mathrm{Spr}(i) \subset J \;\Rightarrow\; i \in J \qquad \textit{for all } i \in I \tag{81.1}$$

is $J = I$.

Proof. Let the subset J of I be given. Then (81.1) fails to hold if and only if there exists $m \in I\backslash J$ such that $\mathrm{Spr}(m) \subset J$, i.e., such that $\mathrm{Spr}(m) \cap (I\backslash J) = \emptyset$; but this precisely describes a minimal member of $I\backslash J$. Thus (81.1) holds if and only if $I\backslash J$ has no minimal members. The equivalence of (i) and (ii) follows at once. ∎

81B. Inductive-Proof Scheme. Let I be a well-founded set, and $P(\ )$ a predicate describing a property that the members of I may have. We define another such predicate, $\mathrm{Ind}_P(\ )$, by requiring that the assertion $\mathrm{Ind}_P(i)$ hold whenever the validity of $P(j)$ for every $j \in I$ that strictly precedes i entails the validity of $P(i)$; the rule defining $\mathrm{Ind}_P(\ )$ is, then,

$$\mathrm{Ind}_P(i) :\Leftrightarrow ((\forall j \in \mathrm{Spr}(i),\ P(j)) \ \Rightarrow\ P(i)) \qquad \text{for all } i \in I.$$

Suppose that $\mathrm{Ind}_P(i)$ holds for *all* $i \in I$. Then the set $J := \{i \in I \mid P(i)\}$ satisfies (81.1) and, by Proposition 81A, we must have $J = I$. We have proved the implication

$$(\forall i \in I, \quad \mathrm{Ind}_P(i)) \ \Rightarrow\ (\forall i \in I, \quad P(i)). \tag{81.2}$$

This provides a scheme for a proof of the assertion that $P(i)$ holds for all $i \in I$: one *proves* that $\mathrm{Ind}_P(i)$ holds for all $i \in I$ — this part is called the *induction step* — and then applies (81.2) to obtain the desired conclusion. The statement "$\forall j \in \mathrm{Spr}(i),\ P(j)$" that occurs in the induction step is called the *induction hypothesis*. A proof according to this scheme is called a *proof by induction* or an *inductive proof*. ■

81C. Example*. The set $\mathbb{N}$ is well-ordered by $\leq$, and $\mathrm{Spr}(n) = n^{\sqsubset} = [\![0, n[\![$ for all $n \in \mathbb{N}$. For each $n \in \mathbb{N}$ we have $n \in (n+1)^{\sqsubset}$. Let the predicate $P(\ )$, describing a property that natural numbers may have, be given. Then obviously

$$P(0) \ \Rightarrow\ ((\forall m \in 0^{\sqsubset},\ P(m)) \ \Rightarrow\ P(0)) \ \Leftrightarrow\ \mathrm{Ind}_P(0)$$
$$(P(n) \ \Rightarrow\ P(n+1)) \ \Rightarrow\ ((\forall m \in (n+1)^{\sqsubset},\ P(m)) \ \Rightarrow\ P(n+1)) \ \Leftrightarrow\ \mathrm{Ind}_P(n+1)$$
$$\text{for all } n \in \mathbb{N}.$$

Since every natural number is either 0 or of the form $n+1$ for some $n \in \mathbb{N}$, the preceding implications together with (81.2) yield the implication

$$(P(0) \ \text{ and } \ (\forall n \in \mathbb{N},\ P(n) \ \Rightarrow\ P(n+1))) \ \Rightarrow\ (\forall n \in \mathbb{N},\ P(n)),$$

which provides the most usual inductive-proof scheme for $\mathbb{N}$. ■

82. Recursive definitions

Let the ordered set I be given. Suppose that one wishes to define a family a with index set I in the following manner: for each $i \in I$, the term a_i is prescribed as a member of a specified set A_i by a rule that involves the terms a_j for the j that strictly precede i. Is a family a well defined by such a prescription? More precisely, is there exactly one family satisfying it? In general, the answer is "no". We proceed to give some counterexamples; a more generic counterexample is discussed in Example 82D.

82A. Examples. (a)*: Consider the set $\mathbb{Z}$ ordered by $\leq$, and attempt to define a family $(a_n \mid n \in \mathbb{Z})$ in $\mathbb{N}$ by the rule $a_n := a_{n-1} + 1$ for all $n \in \mathbb{Z}$. There is no family satisfying this rule, for it would also satisfy $a_{-n} + n = a_0$ for all $n \in \mathbb{N}$, and therefore $a_{-a_0-1} + 1 = 0$ in $\mathbb{N}$, which is impossible.

(b)*: Now attempt to define the family $(a_n \mid n \in \mathbb{Z})$ in $\mathbb{N}$ by the rule $a_n := a_{n-1}$ for all $n \in \mathbb{Z}$. Every constant family satisfies this prescription.

(c)*: Consider the set $[0,1]$ completely ordered by $\leq$, and attempt to define a mapping $f\colon [0,1] \to [0,1]$ by the rule $f(t) := \sup f_>([0,t[)$ for all $t \in [0,1]$. Every isotone mapping $f\colon [0,1] \to [0,1]$ that satisfies $f(0) = 0$ and is left-continuous satisfies this prescription; e.g., $f := 0_{[0,1]\to[0,1]}$, $f := (t \mapsto \frac{1}{2}t)$, $f := (t \mapsto t^{27})$, $f := \chi_{]0,1]}$.

If, however, I is well-founded, then the answer to our questions is "yes", as we now show.

82B. Theorem. *Let the well-founded set I and the family of sets $(A_i \mid i \in I)$ be given. Let a family of mappings $(\phi_i \mid i \in I)$ also be given, with $\phi_i \in \mathrm{Map}(\bigtimes_{j\in \mathrm{Spr}(i)} A_j, A_i)$ for each $i \in I$. Then there is exactly one family $a \in \bigtimes_{i\in I} A_i$ such that*

$$a_i = \phi_i(a|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in I. \tag{82.1}$$

More generally, if J is a subset of I such that $\mathrm{Spr}(i) \subset J$ for every $i \in J$, then $b \in \bigtimes_{i\in I} A_i$ satisfies

$$b_i = \phi_i(b|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in J \tag{82.2}$$

(if and) only if $b = a|_J$.

Proof. 1. Define the mapping $\Gamma : \mathfrak{P}(\dot{\bigcup_{i\in I}} A_i) \to \mathfrak{P}(\dot{\bigcup_{i\in I}} A_i)$ by the rule

$$\Gamma(T) := \{(i, \phi_i(u)) \mid i \in I, u \in \bigtimes_{j\in \mathrm{Spr}(i)} A_j,\ \mathrm{Gr}(u) \subset T\} \qquad \text{for all } T \in \mathfrak{P}(\dot{\bigcup_{i\in I}} A_i).$$

This mapping is clearly isotone from $\mathfrak{P}(\dot{\bigcup_{i\in I}} A_i)$, ordered by inclusion, to itself. We set $C := \bigcap\{T \in \mathfrak{P}(\dot{\bigcup_{i\in I}} A_i) \mid \Gamma(T) \subset T\}$. By the Knaster Fixed-Point Theorem (Theorem 75A), we have

$$C = \Gamma(C) = \{(i, \phi_i(u)) \mid i \in I, u \in \bigtimes_{j \in \mathrm{Spr}(i)} A_j,\ \mathrm{Gr}(u) \subset C\}. \tag{82.3}$$

For each $i \in I$ we set $C_i := \{x \in A_i \mid (i,x) \in C\} \subset A_i$. By (82.3) we have

$$\begin{aligned} C_i = \{\phi_i(u) \mid u \in \bigtimes_{j \in \mathrm{Spr}(i)} A_j, \mathrm{Gr}(u) \subset C\} &= \{\phi_i(u) \mid u \in \bigtimes_{j \in \mathrm{Spr}(i)} C_j\} = \\ &= (\phi_i)_>(\bigtimes_{j \in \mathrm{Spr}(i)} C_j). \end{aligned} \tag{82.4}$$

2. Let $i \in I$ be given and assume that C_j is a singleton for every $j \in \mathrm{Spr}(i)$. Then $\bigtimes_{j \in \mathrm{Spr}(i)} C_j$ is a singleton and, by (82.4), C_i is also a singleton. Since $i \in I$ was arbitrary, we have proved by induction (Inductive-Proof Scheme 81B) that C_i is a singleton for every $i \in I$.

Since $\bigtimes_{i \in I} C_i$ is now a singleton, we may define $c \in \bigtimes_{i \in I} A_i$ by $c :\in \bigtimes_{i \in I} C_i$. For every $i \in I$ we have $\bigtimes_{j \in \mathrm{Spr}(i)} C_j = \{c|_{\mathrm{Spr}(i)}\}$; by (82.4) we therefore have $c_i \in C_i = \{\phi_i(c|_{\mathrm{Spr}(i)})\}$, i.e.,

$$c_i = \phi_i(c|_{\mathrm{Spr(i)}}) \qquad \text{for all } i \in I. \tag{82.5}$$

3. Let the subset J of I be given and assume that $\mathrm{Spr}(i) \subset J$ for every $i \in J$. Assume that $b \in \bigtimes_{i \in J} A_i$ satisfies (82.2). We prove by induction that for every $i \in I$, if $i \in J$, then $b_i = c_i$. Let $i \in I$ be given, and suppose that for every $j \in \mathrm{Spr}(i)$, if $j \in J$, then $b_j = c_j$ (induction hypothesis). Now if $i \in J$, then $\mathrm{Spr}(i) \subset J$, and therefore $b|_{\mathrm{Spr}(i)} = c|_{\mathrm{Spr}(i)}$. By (82.2) and (82.5) we have $b_i = \phi_i(b|_{\mathrm{Spr}(i)}) = \phi_i(c|_{\mathrm{Spr}(i)}) = c_i$. This completes the induction step, and we have shown that $b = c|_J$. Conversely, it is obvious that $b := c|_J$ satisfies (82.2).

Applying the preceding argument to $J := I$, we conclude that there is exactly one $a \in \bigtimes_{i \in I} A_i$ satisfying (82.1), namely $a := c$. ■

The following usual terminology is justified by Theorem 82B. Given the well-founded set I and the families $(A_i \mid i \in I)$ and $(\phi_i \mid i \in I)$, the unique family $a \in \bigtimes_{i \in I} A_i$ that satisfies (82.1) is said to be *defined recursively by the rule*

$$a_i := \phi_i(a|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in I. \tag{82.6}$$

82C. Example*. Let a sequence of sets $(A_n \mid n \in \mathbb{N})$, a member $z \in A_0$, and a sequence of mappings $(h_n \mid n \in \mathbb{N}) \in \bigtimes_{n \in \mathbb{N}} \mathrm{Map}(A_n, A_{n+1})$ be given. Then there exists exactly one sequence $a \in \bigtimes_{n \in \mathbb{N}} A_n$ such that $a_0 = z$ and $a_{n+1} = h_n(a_n)$ for all $n \in \mathbb{N}$.

This is an application of Theorem 82B with $\phi_0 := z_{\{\emptyset\}\to A_0}$ and $\phi_n(u) := h_{n-1}(u_{n-1})$ for all $n \in \mathbb{N}$ and $u \in \bigtimes_{j\in n^\sqsubset} A_j$ (cf. Example 81C).

In particular, if we choose $A_n := \mathbb{N}$ for all $n \in \mathbb{N}$, $z := 1$, and $h_n := (m \mapsto (n+1)m) : \mathbb{N} \to \mathbb{N}$ for all $n \in \mathbb{N}$, we obtain the sequence $(n! \mid n \in \mathbb{N})$ defined recursively by the rules

$$\begin{aligned} 0! &:= 1 \\ (n+1)! &:= (n+1)\cdot n! \qquad \text{for all } n \in \mathbb{N}. \blacksquare \end{aligned}$$

▼ **82D. Example.** The assumption that the ordered set I is well-founded is not only sufficient in Theorem 82B, but essentially necessary as well. More precisely, suppose that the ordered set I and the family of sets $(A_i \mid i \in I)$ are given. Suppose that there are a non-empty subset J of I and families $s, t \in \bigtimes_{i\in I} A_i$ such that J has no minimal members and $s_i \neq t_i$ for all $i \in J$. Then there exists one family of mappings $(\phi_i \mid i \in I)$ with $\phi_i \in \mathrm{Map}(\bigtimes_{j\in \mathrm{Spr}(i)} A_j, A_i)$ for all $i \in I$, such that there is *no* $a \in \bigtimes_{i\in I} A_i$ satisfying (82.1), and another such that there is *more than one* family satisfying (82.1). (The assumption about s and t cannot be omitted.)

For our first example, we define $(\phi_i \mid i \in I)$ by the rule

$$\phi_i(u) := \begin{cases} t_i & \text{if } i \in J \text{ and } u = s|_{\mathrm{Spr}(i)} \\ s_i & \text{otherwise} \end{cases} \qquad \text{for all } i \in I \text{ and } u \in \bigtimes_{j\in\mathrm{Spr}(i)} A_j. \tag{82.7}$$

Suppose that $a \in \bigtimes_{i\in I} A_i$ satisfied (82.1). Choose $i \in J$ (as we may, since $J \neq \emptyset$), and note that $\phi_i(s|_{\mathrm{Spr}(i)}) = t_i \neq s_i$, so that $a \neq s$. We might therefore choose a new $i \in I$ such that $a_i \neq s_i$. By (82.7), this would imply $i \in J$ and $a|_{\mathrm{Spr}(i)} = s|_{\mathrm{Spr}(i)}$. Since i would not be a minimal member of J, we might choose $j \in J \cap \mathrm{Spr}(i)$, and find that $a_j = s_j$. Now $\mathrm{Spr}(j) \subset \mathrm{Spr}(i)$, and therefore $a|_{\mathrm{Spr}(j)} = s|_{\mathrm{Spr}(j)}$. But then (82.1) and (82.7) would yield the contradiction

$$s_j = a_j = \phi_j(a|_{\mathrm{Spr}(j)}) = \phi_j(s|_{\mathrm{Spr}(j)}) = t_j \neq s_j.$$

Therefore there is no $a \in \bigtimes_{i\in I} A_i$ satisfying (82.1).

For our second example, we define $(\phi_i \mid i \in I)$ by the rule

$$\phi_i(u) := \begin{cases} s_i \text{ if } u = s|_{\mathrm{Spr}(i)} \\ t_i \text{ otherwise} \end{cases} \qquad \text{for all } i \in I \text{ and } u \in \bigtimes_{j\in\mathrm{Spr}(i)} A_j. \tag{82.8}$$

It is obvious that $a := s$ now satisfies (82.1). However, we claim that $a := b$ also satisfies (82.1), where $b \in \bigtimes_{i\in I} A_i$ is defined by the rule

$$b_i := \begin{cases} s_i & \text{if } i \in \mathrm{Lb}(J) \\ \\ t_i & \text{if } i \in I\backslash\mathrm{Lb}(J) \end{cases}$$

and is distinct from s. Since J has no minimal member, let alone a minimum, we have $J \subset I\backslash\mathrm{Lb}(J)$. Choose $i \in J$. We then have $b_i = t_i \neq s_i$, so that indeed $b \neq s$.

Let $i \in I$ be given. If $i \in \mathrm{Lb}(J)$ we have $\mathrm{Spr}(i) \subset \mathrm{Lb}(J)$, and therefore, by (82.8), $b_i = s_i = \phi_i(s|_{\mathrm{Spr}(i)}) = \phi_i(b|_{\mathrm{Spr}(i)})$. If, on the other hand, $i \in I\backslash\mathrm{Lb}(J)$, we may choose $j \in J$ such that j precedes i; since j is not a minimal member of J, we may choose $k \in J \cap \mathrm{Spr}(j) \subset J \cap \mathrm{Spr}(i) \subset I\backslash\mathrm{Lb}(J)$. By (82.8) we have $b_k = t_k \neq s_k$, hence $b|_{\mathrm{Spr}(i)} \neq s|_{\mathrm{Spr}(i)}$, hence $b_i = t_i = \phi_i(b|_{\mathrm{Spr}(i)})$. In either case, therefore, $b_i = \phi_i(b|_{\mathrm{Spr}(i)})$, and we conclude that $a := b$ satisfies (82.1). ■

In practice, there are situations where recursive definitions are required but Theorem 82B is not immediately applicable. We shall discuss two such situations in Theorems 82E and 82H.

In the first of these situations, the recursive rule (82.6) that determines a_i when the terms a_j for $j \in \mathrm{Spr}(i)$ are known may "break down" because ϕ_i is defined, not on the whole set $\bigtimes_{j\in\mathrm{Spr}(i)} A_j$, but only on a prescribed subset U_i. In extreme cases, when the recursion is meant to stop short of i, we may even have $U_i = \emptyset$. In all these cases, the aim is to obtain a family a whose index set is as large as possible while the family satisfies the recursion rule wherever it is defined. Sometimes it then turns out that the index set of a is I after all (e.g., in Theorem 82H).

82E. Theorem. (a): *Let the well-founded set I and the family of sets $(A_i \mid i \in I)$ be given. Let a family of sets of families $(U_i \mid i \in I)$ and a family of mappings $(\phi_i \mid i \in I)$ also be given, with $U_i \subset \bigtimes_{j\in\mathrm{Spr}(i)} A_j$ and $\phi_i \in \mathrm{Map}(U_i, A_i)$ for all $i \in I$. Then there is exactly one combination of subset K of I and family $a \in \bigtimes_{i\in K} A_i$ satisfying the conditions*

$$(82.9) \quad i \in K \;\Leftrightarrow\; (\mathrm{Spr}(i) \subset K \text{ and } a|_{\mathrm{Spr}(i)} \in U_i) \qquad \text{for all } i \in I,$$

$$(82.10) \quad a_i = \phi_i(a|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in K.$$

Moreover, if J is a subset of I such that $\mathrm{Spr}(i) \subset J$ for all $i \in J$, then $b \in \bigtimes_{i\in J} A_i$ satisfies

$$(82.11) \quad b|_{\mathrm{Spr}(i)} \in U_i \text{ and } b_i = \phi_i(b|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in J$$

if and only if $J \subset K$ and $b = a|_J$.

(b): *Under the additional assumption that I is well-ordered, either $K = I$ or there exists exactly one $k \in I$ such that $K = \mathrm{Spr}(k) = [\![\min I, k[\![$.*

Proof. Proof of (a). 1. We choose an object, say ω, such that $\omega \notin \bigcup_{i\in I} A_i$ (think of ω as "trash"). We define the family of sets $(A'_i \mid i \in I)$ by $A'_i := A_i \cup \{\omega\}$ for every $i \in I$, and the family of mappings $(\phi'_i \mid i \in I)$, with $\phi'_i \in \mathrm{Map}(\bigtimes_{j\in \mathrm{Spr}(i)} A'_j, A'_i)$ for each $i \in I$, by the rule

$$\phi'_i(u) := \begin{cases} \phi_i(u) & \text{if } u \in U_i \\ \\ \omega & \text{otherwise} \end{cases} \qquad \text{for all } i \in I \text{ and } u \in \bigtimes_{j\in \mathrm{Spr}(i)} A'_j. \tag{82.12}$$

By Theorem 82B there is exactly one family $a' \in \bigtimes_{i\in I} A'_i$ such that

$$a'_i = \phi'_i(a'|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in I.$$

From (82.12), we find that

$$a'_i = \begin{cases} \phi_i(a'|_{\mathrm{Spr}(i)}) & \text{if } a'|_{\mathrm{Spr}(i)} \in U_i \\ \\ \omega & \text{otherwise} \end{cases} \qquad \text{for all } i \in I. \tag{82.13}$$

We set $L := \{i \in I \mid a'_i \neq \omega\}$.

2. Let the subset K of I and the family $a \in \bigtimes_{i\in K} A_i$ be given and assume that (82.9) and (82.10) hold. Then $\mathrm{Spr}(i) \subset K$ for all $i \in K$, and using (82.12),

$$a|_{\mathrm{Spr}(i)} \in U_i \quad \text{and} \quad a_i = \phi'_i(a|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in K.$$

By Theorem 82B (last part of statement) we must have $a = a'|_K$. Since no term of a is ω, we must have $K \subset L$. Suppose that $K \neq L$, and choose a minimal member m of $L\backslash K$. Since $m \in L$, we get from (82.13) that $a'|_{\mathrm{Spr}(m)} \in U_m \subset \bigtimes_{j\in \mathrm{Spr}(m)} A_j$, so that $\mathrm{Spr}(m) \subset L$; since m is a minimal member of $L\backslash K$, we must have $\mathrm{Spr}(m) \subset K$, and therefore $a|_{\mathrm{Spr}(m)} = a'|_{\mathrm{Spr}(m)} \in U_m$. By (82.9) we deduce that $m \in K$, a contradiction. Therefore we must have $K = L$ and $a = a'|_L$.

Conversely, it follows at once from (82.13) that $K := L$ and $a := a'|_L$ satisfy (82.9) and (82.10). This completes the proof of the existence and uniqueness of K and a satisfying (82.9) and (82.10).

3. Let J be a subset of I such that $\mathrm{Spr}(i) \subset J$ for all $i \in J$, and let $b \in \bigtimes_{i\in J} A_i$ be given. In view of (82.12), b satisfies (82.11) if and only if

$$b_i = \phi'_i(b|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in J.$$

It now follows, exactly as in the beginning of Part 2 of this proof, that this requires $J \subset L = K$ and $b = a'|_J = a|_J$. Conversely, if $J \subset K$ it follows at once from (82.10) that $b := a|_J$ satisfies (82.11).

Proof of (b). This is an immediate consequence of (82.9) and the following lemma. ∎

82F. Lemma. *Let the well-ordered set I and the subset K of I be given. The following statements are equivalent:*

(i): $\mathrm{Spr}(i) \subset K$ *for every* $i \in K$.

(ii): $K = I$, *or else* $K = \mathrm{Spr}(\min(I \backslash K))$.

(iii): $K = I$ *or* $K = \mathrm{Spr}(k)$ *for some* $k \in I$.

(iv): *if* $K \neq I$, *there is exactly one* $k \in I$ *such that* $K = \mathrm{Spr}(k)$.

Proof. The implications (ii) ⇒ (iii) and (iv) ⇒ (i) are trivial, and so is (iii) ⇒ (iv) since I is totally ordered (Proposition 64A). It remains to prove the implication (i) ⇒ (ii). Suppose that (i) holds and that $K \neq I$, and set $k := \min(I \backslash K)$. Obviously, $\mathrm{Spr}(k) \subset K$. We have to prove the reverse inclusion. Let $i \in K$ be given. From (i) we have $\mathrm{Spr}(i) \subset K$. Since $k \notin K$ we infer that $k \neq i$ and $k \notin \mathrm{Spr}(i)$, i.e., k does not precede i. Since I is totally ordered, we conclude that i strictly precedes k, i.e., $i \in \mathrm{Spr}(k)$. Since $i \in K$ was arbitrary, we have proved the desired reverse inclusion $K \subset \mathrm{Spr}(k)$, and therefore $K = \mathrm{Spr}(k)$, as asserted by (ii). ∎

82G. Example*. Without going into details, we mention that Theorem 82E applies to all recursive algorithms with "stopping rules", such as the Euclidean algorithm for finding the greatest common divisor of two natural numbers, or a computer program for approximating a solution of a polynomial equation by Newton's method. ∎

Another situation in which Theorem 82B is not sufficient as it stands is this: the recursive rule, instead of specifying a_i itself when the a_j for $j \in \mathrm{Spr}(i)$ are known, merely specifies a set from which a_i is to be *chosen*. The problem is further complicated by the possibility that this set is not guaranteed to be non-empty, unless the a_j for $j \in \mathrm{Spr}(i)$ already satisfy the recursion rule themselves.

•82H. Theorem. *Let the well-founded set I and the family of sets $(A_i \mid i \in I)$ be given. Let a family of set-valued mappings $(\Phi_i \mid i \in I)$ also be given with $\Phi_i \in \mathrm{Map}(\bigtimes_{j \in \mathrm{Spr}(i)} A_j, \mathfrak{P}(A_i))$ for all $i \in I$. Assume that*

$$(\forall j \in \mathrm{Spr}(i), \quad u_j \in \Phi_j(u|_{\mathrm{Spr}(j)})) \;\Rightarrow\; \Phi_i(u) \neq \emptyset \quad \text{for all } i \in I \text{ and } u \in \bigtimes_{j \in \mathrm{Spr}(i)} A_j. \tag{82.14}$$

Then there exists a family $a \in \bigtimes_{i \in I} A_i$ such that

$$a_i \in \Phi_i(a|_{\mathrm{Spr}(i)}) \quad \text{for all } i \in I. \tag{82.15}$$

Proof. 1. •For each $i \in I$ and each non-empty subset B of A_i we choose $\gamma_i(B) \in B$. (This is equivalent to •choosing a member of $\bigtimes_{(i,B) \in \Xi} B$, where $\Xi := \dot{\bigcup_{i \in I}} \mathfrak{P}^{\times}(A_i)$.)

We define the family $(U_i \mid i \in I)$ by

$$U_i := \{u \in \bigtimes_{j \in \mathrm{Spr}(i)} A_j \mid \forall j \in \mathrm{Spr}(i), \quad u_j \in \Phi_j(u|_{\mathrm{Spr}(j)})\} \quad \text{for all } i \in I. \tag{82.16}$$

We can rephrase (82.14) as

$$\Phi_i(u) \neq \emptyset \qquad \text{for all } i \in I \text{ and } u \in U_i.$$

We may therefore define the family of mappings $(\phi_i \mid i \in I)$ with $\phi_i \in \mathrm{Map}(U_i, A_i)$ for all $i \in I$ by the rule

$$\phi_i(u) := \gamma_i(\Phi_i(u)) \in \Phi_i(u) \qquad \text{for all } i \in I \text{ and } u \in U_i. \tag{82.17}$$

We now have the assumptions of Theorem 82E. We therefore obtain a subset K of I and a family $a \in \bigtimes_{i \in K} A_i$ satisfying (82.9) and (82.10).

2. We now show by induction that $K = I$. Let $i \in I$ be given and suppose that $\mathrm{Spr}(i) \subset K$ (induction hypothesis). For every $j \in \mathrm{Spr}(i)$ we have $j \in K$ and, by (82.10) and (82.17), $a_j = \phi_j(a|_{\mathrm{Spr}(j)}) \in \Phi_j(a|_{\mathrm{Spr}(j)})$. Therefore $a|_{\mathrm{Spr}(i)} \in U_i$, by (82.16). It then follows by (82.9) that $i \in K$. This completes the induction step. We have shown that $K = I$ and, by (82.10) and (82.17), that a satisfies (82.15). ■

The following terminology is used in connection with Theorem 82H. Given the well-founded set I and the families $(A_i \mid i \in I)$ and $(\Phi_i \mid i \in I)$ and assuming that (82.14) is satisfied, a family $a \in \bigtimes_{i \in I} A_i$ that satisfies (82.15) is said to be *chosen recursively by the rule* (82.15).

This page intentionally left blank

Chapter 9

THE NATURAL NUMBERS

91. Principles of counting

The natural or counting numbers, 1, 2, 3, etc., are so deeply embedded in our social and intellectual culture that it may seem strange to assert that they require more than cursory analysis. A closer examination of their history and function reveals, however, that the art of counting is the product of a complicated cultural development.

Some of the ideas about numbers that seem to us obvious are shown by the historic, anthropological, and linguistic evidence not to have been obvious at all. Among these often imperfectly realized ideas are: the use of a common system of numbers for counting, regardless of the nature of the objects counted; the perception that quantitative judgments about collections of objects are reducible to counting; the notion that numbers can be recorded by symbols that reflect the process of counting. (For some of the evidence on these matters, as well as much more about numbers, an interesting source is K. Menninger, *Zahlwort und Ziffer*, translated as *Number Words and Number Symbols* (M.I.T. Press).)

The most difficult idea to grasp in all its implications appears to have been the innocent-looking insight that counting goes on and on: there always is a next number. Philosophical discussions about the "potential" or "actual" infiniteness of the set of all numbers have lasted well into the twentieth century of our era. Archimedes ($Αρχιμήδης$, d. 212 B.C.E.), in his $Ψαμμίτης$ (*The Sand-Reckoner*) struggled to make plain a related matter. To simplify his point somewhat: he showed that some huge collection — all the grains of sand in the entire world — was neither "infinite", as some claimed, nor, as others averred, "not infinite, yet greater than any nameable number", but was, on the contrary, comfortably within the scope of a well-designed system of counting. The point for us is that valid conclusions can be drawn concerning natural numbers no matter how great; even so great that they are wholly beyond effective recording, not to mention attainment through actual counting. This suggests the need for a firmer framework than can be constructed by inspecting the few small numbers that are within our immediate experience.

We shall base our understanding of the natural numbers on some elementary insights concerning the process of counting, and on nothing else. This will therefore

be a thoroughly "ordinal" view of the natural numbers, as opposed to a "cardinal" view. The contrast is perhaps best illustrated by a sketch of the definition of the operation of addition of natural numbers. The "ordinal" approach says that $m + n$ is obtained by "counting m" from 0 to m, and then "counting n more" from m to $m + n$. The "cardinal" approach would consist in realizing $m + n$ by taking a set "with m members" and another, disjoint from it, "with n members", to obtain their union, a set "with $m + n$ members". There appears to be a consensus to the effect that the "ordinal" approach, one version of which we adopt, is more easily formulated and leads to usable results faster. Some of the flavor of the "cardinal" approach will be found in Chapter 10.

The insights concerning counting that we consider basic are: that counting has a beginning (which, for various good reasons, we take to be 0 rather than 1); that counting proceeds by going from each number to the next number; that a number, once counted, is never counted again; and that every number is counted "eventually".

To formalize these insights, we define a **counting system** to be a set N endowed with structure by the prescription of a member 0 of N, called **zero**, and of a mapping seq: $N \to N$, called the **successor-mapping**, subject to the following conditions:

(Count I): $0 \notin \text{Rng seq}$.

(Count II): seq *is injective*.

(Count III): $\forall S \in \mathfrak{P}(N), \quad (0 \in S \textit{ and } \text{seq}_{>}(S) \subset S) \;\Rightarrow\; S = N$.

We shall show in Section 95 that counting systems are all alike in everything that matters. The question regarding the *existence* of counting systems belongs to the foundational aspects of mathematics, and we shall not discuss it: we take it for granted, or agreed, that counting systems exist. Actually, we shall adopt, or pretend to adopt, the naive view that one specific counting system is revealed to us, or singled out by us, as the **Natural-Number System**, to be denoted by $\mathbb{N}$, and its members to be known as the **natural numbers**.

91A. Remark. Since everything we do with this Natural-Number System is based on (Count I), (Count II), (Count III), and on nothing else, every conclusion we reach about it will be valid for every other counting system as well. ■

We thus have a set $\mathbb{N}$, whose members are called (**natural**) **numbers**; a member 0 of $\mathbb{N}$, called the number **zero**; and a mapping $\text{seq} : \mathbb{N} \to \mathbb{N}$, called the **successor-mapping**, satisfying the conditions

($\mathbb{N}$I): $0 \notin \text{Rng seq}$.

($\mathbb{N}$II): seq *is injective*.

($\mathbb{N}$III): $\forall S \in \mathfrak{P}(\mathbb{N}), \quad (0 \in S \textit{ and } \text{seq}_{>}(S) \subset S) \;\Rightarrow\; S = \mathbb{N}$.

The conditions ($\mathbb{N}$I), ($\mathbb{N}$II), ($\mathbb{N}$III) are known as the *Peano Axioms*. They are so called in honor of Giuseppe Peano (1858-1932), who in 1889 proposed essentially the same conditions as a foundation for a systematic account of the natural numbers. (As Peano explained elsewhere, however, these conditions were originally due to Julius Wilhelm Richard Dedekind (1831-1916).) Condition ($\mathbb{N}$III) is known as the *Induction Axiom*, since it underlies a scheme for "proofs by induction" (see Inductive-Proof Scheme 91C).

The value seq$n$ of the successor-mapping at $n \in \mathbb{N}$ is called the **successor of** n. It is useful to have names for some natural numbers. We define the numbers **one**, **two**, and **four** to be $1 := \text{seq}0, 2 := \text{seq}1$, and $4 := \text{seq seq}2$, respectively.

We set $\mathbb{N}^\times := \mathbb{N}\backslash\{0\}$, and obtain our first consequence of the Peano Axioms.

91B. Proposition. $\mathbb{N}^\times = \text{Rng seq}$.

Proof. Set $S := \{0\} \cup \text{Rng seq} \subset \mathbb{N}$. Then $0 \in S$ and $\text{seq}_>(S) \subset \text{Rng seq} \subset S$. By (ℕIII) we must have $\{0\} \cup \text{Rng seq} = S = \mathbb{N}$, whence $\mathbb{N}^\times \subset \text{Rng seq}$. But by (ℕI) we have $\text{Rng seq} \subset \mathbb{N}^\times$, and therefore equality must hold. ■

91C. Inductive-Proof Scheme. The Induction Axiom (ℕIII) is most frequently used as follows. Let $P(\ )$ denote a predicate describing a property that a natural number may have. We then have the implication

$$(91.1) \qquad (P(0) \text{ and } (\forall n \in \mathbb{N},\ \ P(n) \ \Rightarrow\ P(\text{seq}n))) \ \Rightarrow\ (\forall n \in \mathbb{N},\ \ P(n)),$$

which follows upon applying (ℕIII) to the subset $S := \{n \in \mathbb{N} \mid P(n)\}$ of $\mathbb{N}$.

This provides a scheme for a proof of the assertion that $P(n)$ holds for all $n \in \mathbb{N}$: one *proves* that $P(0)$ holds *and* that

$$(91.2) \qquad P(n) \ \Rightarrow\ P(\text{seq}n) \qquad \text{for all } n \in \mathbb{N},$$

and then applies (91.1) to reach the desired conclusion. The proof of (91.2) is called the *induction step*, and the statement "$P(n)$" that occurs in it is called the *induction hypothesis*. A proof according to this scheme is called a *proof by induction*.

(This terminology does not clash with that introduced in Section 81: it will be seen in Section 93 that the present scheme is actually a special case of the other, as already suggested in Example 81C.) ■

92. Order

We recall that with the mapping $\mathrm{seq} : \mathbb{N} \to \mathbb{N}$ we can associate the *functional relation* $\overset{\mathrm{seq}}{\longmapsto}$ in $\mathbb{N}$, defined by

$$\forall m, n \in \mathbb{N}, \quad m \overset{\mathrm{seq}}{\longmapsto} n \ :\Leftrightarrow \ n = \mathrm{seq} m.$$

It follows from (NII) that this relation satisfies

$$\forall m, n \in \mathbb{N}, \quad m \overset{\mathrm{seq}}{\longmapsto} \mathrm{seq} n \ \Leftrightarrow \ m = n. \tag{92.1}$$

We define the relation $<$ in $\mathbb{N}$ to be the *transitive closure* of the relation $\overset{\mathrm{seq}}{\longmapsto}$ (Section 73), i.e., the narrowest among the transitive relations in $\mathbb{N}$ that are broader than $\overset{\mathrm{seq}}{\longmapsto}$. For our present purposes it will be enough to recall that this relation $<$ is transitive and satisfies

$$\forall m, n \in \mathbb{N}, \quad m < n \ \Leftrightarrow \ (m \overset{\mathrm{seq}}{\longmapsto} n \ \text{ or } \ (\exists p \in \mathbb{N}, \ \ m < p \text{ and } p \overset{\mathrm{seq}}{\longmapsto} n)) \tag{92.2}$$

(Proposition 73H).

From (92.2) it follows that $m < n$ can only hold if $n \in \mathrm{Rng}\,\mathrm{seq}$. By (NI) we therefore have

$$\forall m, n \in \mathbb{N}, \quad m < n \ \Rightarrow \ n \neq 0. \tag{92.3}$$

From (92.1) and (92.2) we obtain

$$\forall m, n \in \mathbb{N}, \quad m < \mathrm{seq} n \ \Leftrightarrow \ (m = n \text{ or } m < n). \tag{92.4}$$

All the information we need about $<$ beyond its transitivity is contained in (92.3) and (92.4). This will be less surprising after consideration of the following proposition.

92A. Proposition. *There is exactly one relation ρ in $\mathbb{N}$ such that*

$$\forall m, n \in \mathbb{N}, \quad m \ \rho \ n \ \Rightarrow \ n \neq 0 \tag{92.5}$$

$$\forall m, n \in \mathbb{N}, \quad m \ \rho \ \mathrm{seq} n \ \Leftrightarrow \ (m = n \text{ or } m \ \rho \ n); \tag{92.6}$$

namely, the relation $<$.

Proof. Suppose that ρ' and ρ'' are relations in $\mathbb{N}$ such that (92.5) and (92.6) hold with $\rho := \rho'$ and also with $\rho := \rho''$. We prove by induction that $P(n) :\Leftrightarrow (\forall m \in \mathbb{N}, \quad m \ \rho' \ n \ \Leftrightarrow \ m \ \rho'' \ n)$ holds for all $n \in \mathbb{N}$.

Now $P(0)$ holds, since both $m \ \rho' \ 0$ and $m \ \rho'' \ 0$ are ruled out by (92.5) for every $m \in \mathbb{N}$. Let $n \in \mathbb{N}$ be given and suppose that $P(n)$ holds. Then (92.6) yields the following chain of equivalences for every $m \in \mathbb{N}$:

$$m \ \rho' \ \mathrm{seq} n \ \Leftrightarrow \ (m = n \text{ or } m \ \rho' \ n) \ \Leftrightarrow \ (m = n \text{ or } m \ \rho'' \ n) \ \Leftrightarrow \ m \ \rho'' \ \mathrm{seq} n,$$

and hence $P(\mathrm{seq}n)$ holds. This completes the induction step. Thus $P(n)$ holds for all $n \in \mathbb{N}$, and this means $\rho' = \rho''$. There is thus at most one relation ρ in $\mathbb{N}$ satisfying (92.5) and (92.6); but from (92.3) and (92.4) it follows that $\rho := <$ does satisfy (92.5) and (92.6). ■

The next few results establish the essential properties of the relation $<$.

92B. Theorem. *The relation $<$ is a strict-order in $\mathbb{N}$.*

Proof. Since we know that $<$ is transitive, it remains to prove that it is irreflexive. We prove by induction that $(\mathrm{not}(n < n))$ holds for all $n \in \mathbb{N}$. From (92.3) we obtain $(\mathrm{not}(0 < 0))$. Let $n \in \mathbb{N}$ be given. By (92.4) we have $n < \mathrm{seq}n$; also, if $\mathrm{seq}n < \mathrm{seq}n$, then $\mathrm{seq}n = n$ or $\mathrm{seq}n < n$; since $<$ is transitive, this implies $n < n$. By contraposition, $(\mathrm{not}(n < n))$ implies $(\mathrm{not}(\mathrm{seq}n < \mathrm{seq}n))$. This completes the induction step, and with it the proof. ■

To the strict-order $<$ there corresponds, as usual, the (lax) order $\leq$, in accordance with Proposition 56A,(b). (Some prefer the symbol $\leqq$ instead of $\leq$; both forms are common.) From now on, the set $\mathbb{N}$ shall be regarded as ordered by $\leq$, unless a different order in $\mathbb{N}$ is explicitly specified. We note that (92.4) implies that $\mathrm{seq}n$ immediately follows n for each $n \in \mathbb{N}$.

We next show that the mapping $\mathrm{seq} : \mathbb{N} \to \mathbb{N}$ is strictly isotone, and a bit more.

92C. Proposition.

$$\forall m, n \in \mathbb{N}, \quad m < n \;\Leftrightarrow\; \mathrm{seq}m < \mathrm{seq}n.$$

Proof. Define the relation ρ in $\mathbb{N}$ by

$$\forall m, n \in \mathbb{N}, \quad m \;\rho\; n \;:\Leftrightarrow\; \mathrm{seq}m < \mathrm{seq}n.$$

By (92.4) we obtain

$$\forall m, n \in \mathbb{N}, \quad m \;\rho\; n \;\Leftrightarrow\; \mathrm{seq}m \leq n. \tag{92.7}$$

Since $m < \mathrm{seq}m$ and since $<$ is transitive, we find that $m \;\rho\; n$ implies $m < n$. In view of (92.3), this shows that ρ satisfies (92.5). From (92.7), (92.4), and (NII) we obtain

$$\forall m, n \in \mathbb{N}, \quad m \;\rho\; \mathrm{seq}n \;\Leftrightarrow\; \mathrm{seq}m \leq \mathrm{seq}n \;\Leftrightarrow\; (m = n \text{ or } m \;\rho\; n),$$

so that ρ satisfies (92.6). Proposition 92A then shows that ρ is $<$, as asserted. ■

92D. Theorem. *The set $\mathbb{N}$ is well-ordered by $\leq$, with $0 = \min\mathbb{N}$ and $1 = \min\mathbb{N}^\times$.*

Proof. 1. We first prove by induction that $0 \leq n$ for all $n \in \mathbb{N}$. Obviously $0 \leq 0$. Let $n \in \mathbb{N}$ be given. If $0 \leq n$, then (92.4) implies $0 \leq n < \mathrm{seq}n$. This completes the inductive proof. Therefore $0 = \min\mathbb{N}$. An immediate consequence is

$$0 \in A \;\Rightarrow\; 0 = \min A \qquad \text{for all } A \in \mathfrak{P}(\mathbb{N}). \tag{92.8}$$

Using Propositions 92C and 91B it further follows that

$$1 = \mathrm{seq}0 = \mathrm{seq}\,\min\mathbb{N} = \min\,\mathrm{seq}_>(\mathbb{N}) = \min\mathbb{N}^\times.$$

2. We now prove by induction that

$$P(n) \;:\Leftrightarrow\; (\forall A \in \mathfrak{P}(\mathbb{N}), \quad n \in A \;\Rightarrow\; (A \text{ has a minimum}))$$

holds for all $n \in \mathbb{N}$. Now $P(0)$ holds on account of (92.8). Let $n \in \mathbb{N}$ be given, and assume that $P(n)$ holds. Let A be a subset of $\mathbb{N}$ that contains seq$n$. Then $n \in A \cup \{n\}$; by the induction hypothesis we may set $m := \min(A \cup \{n\})$. If $m \in A$, then $m = \min A$. If $m \notin A$, then $n \notin A$ and $m = n$. Let $p \in A$ be given. Then $n = \min(A \cup \{n\}) < p$. By Proposition 92C, seq$n <$ seq$p$, and by (92.4), seq$n = p$ or seq$n <$ seq$p$. Since $p \in A$ was arbitrary, it follows that seq$n = \min A$. Thus A has a minimum in either case. Therefore $P(\text{seq}n)$ holds. This completes the induction step. We have shown that $P(n)$ holds for all $n \in \mathbb{N}$, which means that

$$n \in A \;\Rightarrow\; (A \text{ has a minimum}) \qquad \text{for all } A \in \mathfrak{P}(\mathbb{N}) \text{ and } n \in \mathbb{N}.$$

This is a complicated way of saying that every non-empty subset A of $\mathbb{N}$ has a minimum; in other words, that $\mathbb{N}$ is well-ordered. ■

It follows from Theorem 92D and from Proposition 64A that the ordered set $\mathbb{N}$ is totally ordered and well-founded.

We introduce some notation. For each $n \in \mathbb{N}$, we set $n^{\sqsubset} := \text{Spr}(n) = \{m \in \mathbb{N} \mid m < n\}$. Since $0 = \min\mathbb{N}$, the set $n^{\sqsubset}$ is precisely the order-interval $[\![0, n[\![$ of $\mathbb{N}$. We also define $n^{\sqsupset} := \text{seq}_{>}(n^{\sqsubset})$. In particular, $0^{\sqsubset} = 0^{\sqsupset} = \emptyset, 1^{\sqsubset} = \{0\},\ 2^{\sqsubset} = \{0, 1\},\ 1^{\sqsupset} = \{1\},\ 2^{\sqsupset} = \{1, 2\}$.

92E. Proposition. *Let $n \in \mathbb{N}$ be given. Then:*

(a): $(\text{seq}n)^{\sqsubset} = [\![0, n]\!]$, *so that* $n = \max(\text{seq}n)^{\sqsubset}$ and $n^{\sqsubset} = (\text{seq}n)^{\sqsubset}\backslash\{n\}$.

(b): $n^{\sqsupset} = (\text{seq}n)^{\sqsubset}\backslash\{0\} =]\!]0, n]\!]$. *If $n \neq 0$, then $n^{\sqsupset} = [\![1, n]\!]$.*

Proof. (a) is an immediate consequence of (92.4). From Propositions 92C and 91B we have

$$n^{\sqsupset} = \text{seq}_{>}(n^{\sqsubset}) = (\text{seq}n)^{\sqsubset} \cap \text{Rng seq} = (\text{seq}n)^{\sqsubset}\backslash\{0\}.$$

From (a) we have $(\text{seq}n)^{\sqsubset}\backslash\{0\} = [\![0, \text{n}]\!]\backslash\{0\} =]\!]0, n]\!]$. If $n \neq 0$, then $1 = \min\mathbb{N}^{\times} \in]\!]0, n]\!] \subset \mathbb{N}^{\times}$, and therefore $1 = \min]\!]0, n]\!]$, whence $]\!]0, n]\!] = [\![1, n]\!]$. This completes the proof of (b). ■

The following notation is convenient and suggestive. For given $m, n \in \mathbb{N}$ we set

$$m..n := \{k \in \mathbb{N} \mid m \leq k \leq n\} = \begin{cases} [\![m, n]\!] & \text{if } m \leq n \\ \\ \emptyset & \text{if } m > n. \end{cases} \tag{92.9}$$

This notation is designed to allow the formula "$k \in m..n$" to replace, with the least alteration, the often-encountered, but inappropriate, formula "$k = m, \ldots, n$" (the latter formula misuses the symbol $=$ and includes the uninterpreted ellipsis). We note that $m..n = n^{\sqsupset}\backslash m^{\sqsubset}$ for all $m, n \in \mathbb{N}$, except that $0..0 = \{0\}$; in particular, $1..n = n^{\sqsupset}$ for all $n \in \mathbb{N}$.

The set $\mathbb{N}$ itself has no maximum, since $n <$ seq$n$ for all $n \in \mathbb{N}$. Many subsets of $\mathbb{N}$, however, do have maxima.

92F. Proposition. *A subset of $\mathbb{N}$ has a maximum if (and only if) it is not empty and has an upper bound.*

Proof. We prove by induction that

$$P(n) \;:\Leftrightarrow\; (\forall A \in \mathfrak{P}(n^{\sqsubset}), \quad A \neq \emptyset \;\Rightarrow\; (A \text{ has a maximum}))$$

holds for all $n \in \mathbb{N}$. This will be sufficient: indeed, if m is an upper bound of A, then $A \subset (\mathrm{seq}m)^{\sqsubset}$. Now $P(0)$ holds vacuously, since $0^{\sqsubset} = \emptyset$. Let $n \in \mathbb{N}$ be given and assume that $P(n)$ holds. If A is a non-empty subset of $(\mathrm{seq}n)^{\sqsubset}$, then Proposition 92E,(a) leads to an alternative: either $\max(\mathrm{seq}n)^{\sqsubset} = n \in A$, and then $n = \max A$, or else $A \subset (\mathrm{seq}n)^{\sqsubset} \backslash \{n\} = n^{\sqsubset}$, and then A has a maximum by the induction hypothesis; in either case A has a maximum. Therefore $P(\mathrm{seq}n)$ holds, and the induction step is complete. ■

We conclude this section with three results on sequences in ordered sets. We recall that a **sequence** is a family whose index set is $\mathbb{N}$ or $\mathbb{N}^{\times}$.

92G. Proposition. *Let a set D ordered by $\prec$ and a sequence $a \in D^{\mathbb{N}}$ be given. Then a is [strictly] isotone if (and only if) $a_n \prec a_{\mathrm{seq}n}$ $[a_n \precneqq a_{\mathrm{seq}n}]$ for all $n \in \mathbb{N}$.*

Proof. The relation ρ in $\mathbb{N}$ defined by the rule

$$\forall m, n \in \mathbb{N}, \quad m \;\rho\; n \;:\Leftrightarrow\; a_m \prec a_n \;[a_m \precneqq a_n],$$

is transitive, and is broader than $\overset{\mathrm{seq}}{\longmapsto}$. It is therefore broader than $<$, the transitive closure of $\overset{\mathrm{seq}}{\longmapsto}$. Thus,

$$\forall m, n \in \mathbb{N}, \quad m < n \;\Rightarrow\; m \;\rho\; n \;\Leftrightarrow\; a_m \prec a_n \;[a_m \precneqq a_n]. \;\blacksquare$$

92H. Proposition. (Principle of Descent). *Let a well-founded ordered set $(D; \prec)$ be given.*

(a): *For every antitone sequence s in D there exists $m \in \mathbb{N}$ such that $s_n = s_m$ for all $n \in \mathbb{N} \backslash m^{\sqsubset}$.*

(b): *There is no strictly antitone sequence in D.*

Proof. Let the antitone sequence s in D be given. We may choose a minimal member d of $\mathrm{Rng}s$, and may further choose $m \in \mathbb{N}$ such that $s_m = d$. For every $n \in \mathbb{N} \backslash m^{\sqsubset}$ we have $m \leq n$, and hence $s_n \prec s_m = d$; since d is a minimal member of $\mathrm{Rng}s$, this implies $s_n = d = s_m$. In particular, $s_{\mathrm{seq}m} = s_m$, so that s is not strictly antitone (Proposition 92G). ■

92I. Proposition. *If $a \in \mathbb{N}^{\mathbb{N}}$ is a strictly isotone sequence, then $n \leq a_n$ for all $n \in \mathbb{N}$.*

Proof. We have $0 \leq a_0$. Let $n \in \mathbb{N}$ be given. If $n \leq a_n$, then $n \leq a_n < a_{\mathrm{seq}n}$, and Proposition 92C and (92.4) yield $\mathrm{seq}n < \mathrm{seq}a_{\mathrm{seq}n}$, and hence $\mathrm{seq}n \leq a_{\mathrm{seq}n}$. This completes the proof by induction. ■

93. General induction and recursive definitions

93A. General Inductive-Proof Scheme. Since $\mathbb{N}$ is well-ordered (Theorem 92D), we have available to us the Inductive-Proof Scheme 81B. We note that the induction step now reads

$$\forall n \in \mathbb{N}, \quad (\forall m \in n^{\sqsubset}, \quad P(m)) \;\Rightarrow\; P(n) \tag{93.1}$$

If it is necessary to distinguish a proof using the induction step (93.1) from one that uses $P(0)$ and (91.2) according to Inductive-Proof Scheme 91C, we shall refer to the former as a *proof by general induction*, and to the latter as a *proof by special induction*. We note that the latter is actually a particular case of the former, in the following way. We examine the number 0 and the numbers in $\mathbb{N}^{\times} = \operatorname{Rng}\operatorname{seq}$ separately. On the one hand, the assertion $P(0)$ trivially implies the assertion $(\forall m \in 0^{\sqsubset}, P(m)) \;\Rightarrow\; P(0)$ (they are in fact equivalent). On the other hand, for each $n \in \mathbb{N}$ we have $n \in (\operatorname{seq} n)^{\sqsubset}$ and the assertion $P(n) \;\Rightarrow\; P(\operatorname{seq} n)$ therefore implies the assertion $(\forall m \in (\operatorname{seq} n)^{\sqsubset},\; P(m)) \;\Rightarrow\; P(\operatorname{seq} n)$. ■

As regards recursive definitions on the well-ordered index set $\mathbb{N}$, we sometimes have to rely on the full force Theorems 82B, 82E, and 82H, with $I := \mathbb{N}$. There is, however, a more usual and more special pattern of recursive definition, derived from the general one and related to it just as proofs by special induction are related to proofs by general induction. (This pattern was anticipated in Example 82C.)

93B. Theorem. *Let the sequence of sets $(A_n \mid n \in \mathbb{N})$, the member z of A_0, and the sequence of mappings $(h_n \mid n \in \mathbb{N}) \in \bigtimes_{n\in\mathbb{N}} \operatorname{Map}(A_n, A_{\operatorname{seq} n})$ be given. Then there is exactly one sequence $a \in \bigtimes_{n\in\mathbb{N}} A_n$ such that*

$$a_0 = z \tag{93.2}$$

$$a_{\operatorname{seq} n} = h_n(a_n) \quad \text{for all } n \in \mathbb{N}. \tag{93.3}$$

Moreover, if $m \in \mathbb{N}$, then $b \in \bigtimes_{n\in(\operatorname{seq} m)^{\sqsubset}} A_n$ satisfies

$$b_0 = z$$

$$b_{\operatorname{seq} n} = h_n(b_n) \quad \text{for all } n \in m^{\sqsubset}$$

if and only if $b = a|_{(\operatorname{seq} m)^{\sqsubset}}$.

Proof. For each $n \in \mathbb{N}^{\times}$ we have $n = \operatorname{seq}\max n^{\sqsubset}$, by Propositions 91B and 92E,(a). The assertion therefore follows from Theorem 82B, with $I := \mathbb{N}$ and with the sequence of mappings $(\phi_n \mid n \in \mathbb{N})$ defined by the rule

$$\phi_n(u) := \begin{cases} z & \text{if } n = 0 \\ h_{\max n^{\sqsubset}}(u_{\max n^{\sqsubset}}) & \text{if } n \in \mathbb{N}^{\times} \end{cases} \quad \text{for all } u \in \bigtimes_{m\in n^{\sqsubset}} A_m. \tag{93.4}$$

■

The unique sequence a that satisfies (93.2) and (93.3) is said to be *defined recursively by the rules*

$$\begin{aligned} a_0 &:= z \\ a_{\mathrm{seq}n} &:= h_n(a_n) \quad \textit{for all } n \in \mathbb{N}. \end{aligned}$$

93C. Theorem. *Let the sequences of sets* $(A_n \mid n \in \mathbb{N})$ *and* $(B_n \mid n \in \mathbb{N})$ *be given, with* $B_n \subset A_n$ *for all* $n \in \mathbb{N}$. *Let the member* z *of* A_0 *and the sequence of mappings* $(h_n \mid n \in \mathbb{N}) \in \bigtimes_{n\in\mathbb{N}} \mathrm{Map}(B_n, A_{\mathrm{seq}n})$ *be given. Then either there is exactly one sequence* $a \in \bigtimes_{n\in\mathbb{N}} B_n$ *such that* (93.2) *and* (93.3) *hold, or else there is exactly one combination of a number* $k \in \mathbb{N}$ *and a list* $a \in \bigtimes_{n\in(\mathrm{seq}k)^\sqsubset} A_n$ *such that*

$$a_0 = z$$

$$a_n \in B_n \ \textit{ and } \ a_{\mathrm{seq}n} = h_n(a_n) \quad \textit{for all } n \in k^\sqsubset$$

$$a_k \in A_k \backslash B_k.$$

Proof. This follows from Theorem 82E just as Theorem 93B above followed from Theorem 82B, with $I := \mathbb{N}$ and $(\phi_n \mid n \in \mathbb{N})$ defined by (93.4), and $(U_n \mid n \in \mathbb{N})$ defined by

$$U_0 := \{\emptyset\} = \bigtimes_{m\in 0^\sqsubset} A_m$$

$$U_{\mathrm{seq}n} := \{u \in \bigtimes_{m\in(\mathrm{seq}n)^\sqsubset} A_m \mid u_n \in B_n\} \quad \text{for all } n \in \mathbb{N}. \ \blacksquare$$

•93D. Theorem. *Let the sequence of sets* $(A_n \mid n \in \mathbb{N})$, *the subset* Z *of* A_0, *and the sequence of set-valued mappings* $(H_n \mid n \in \mathbb{N}) \in \bigtimes_{n\in\mathbb{N}} \mathrm{Map}(A_n, \mathfrak{P}(A_{\mathrm{seq}n}))$ *be given. Assume that* $Z \neq \emptyset$ *and that*

$$\begin{aligned} &(u_0 \in Z \textit{ and } (\forall m \in n^\sqsubset, \ \ u_{\mathrm{seq}m} \in H_m(u_m))) \ \Rightarrow H_n(u_n) \neq \emptyset \\ &\qquad \textit{for all } n \in \mathbb{N} \textit{ and } u \in \bigtimes_{m\in(\mathrm{seq}n)^\sqsubset} A_m. \end{aligned} \tag{93.5}$$

Then there exists a sequence $a \in \bigtimes_{n\in\mathbb{N}} A_n$ *such that*

$$a_0 \in Z \tag{93.6}$$

$$a_{\mathrm{seq}n} \in H_n(a_n) \quad \textit{for all } n \in \mathbb{N}. \tag{93.7}$$

Proof. This follows from •Theorem 82H, exactly as Theorem 93B above followed from Theorem 82B. ■

A sequence a that satisfies (93.6) and (93.7) is said to be *chosen recursively by the rules* (93.6) *and* (93.7).

As an application of Theorem 93D, we complement the Principle of Descent (Proposition 92H) with a converse.

•93E. Corollary. *An ordered set is well-founded if and only if there is no strictly antitone sequence in it.*

Proof. The "only if" part is Proposition 92H,(b). To prove the "if" part, we show that in a given ordered set D that is not well-founded a strictly antitone sequence can be chosen recursively.

We may choose a non-empty subset A of D such that A has no minimal members. We apply •Theorem 93D with $A_n := A$ for all $n \in \mathbb{N}$, $Z := A$, and $H_n(x) := A \cap \mathrm{Spr}(x)$ for all $n \in \mathbb{N}$ and $x \in A$. Since A is not empty and has no minimal members, we have $Z \neq \emptyset$ and $H_n(x) \neq \emptyset$ for all $n \in \mathbb{N}$ and $x \in A$. By •Theorem 93D, there is a sequence $a \in A^{\mathbb{N}} \subset D^{\mathbb{N}}$ such that $a_0 \in A$ and $a_{\mathrm{seq}n} \in A \cap \mathrm{Spr}(a_n)$ for all $n \in \mathbb{N}$. Since $a_{\mathrm{seq}n}$ strictly precedes a_n for every $n \in \mathbb{N}$, it follows from Proposition 92G that a is strictly antitone. ■

94. Iteration

For our immediate ends (up to and including Section 97) we shall require only one special case of recursive definition. Since the reader might still feel uncomfortable with such definitions and their justification, and this special case is so plausible that it may temporarily be granted without checking the proof, we state this case explicitly.

94A. Theorem. *Let the set D and the mapping $f\colon D \to D$ be given. Then there is exactly one sequence of mappings $(g_n \mid n \in \mathbb{N})$ in $\mathrm{Map}(D, D)$ such that*

$$g_0 = 1_D$$

$$g_{\mathrm{seq}n} = f \circ g_n \quad \text{for all } n \in \mathbb{N}.$$

Proof. This is a special case of Theorem 93B, with $A_n := \mathrm{Map}(D, D)$ for all $n \in \mathbb{N}$, $z := 1_D$, and $h_n(u) := f \circ u$ for all $n \in \mathbb{N}$ and $u \in \mathrm{Map}(D, D)$. ■

The unique sequence provided by Theorem 94A is denoted by $(f^{\circ n} \mid n \in \mathbb{N})$ and called the **sequence of iterates of** f; it is defined recursively by the rules

$$f^{\circ 0} := 1_D \tag{94.1}$$

$$f^{\circ \mathrm{seq}n} := f \circ f^{\circ n} \quad \text{for all } n \in \mathbb{N}. \tag{94.2}$$

For each $n \in \mathbb{N}$, the term $f^{\circ n}$ is called the n**th iterate of** f. We note that $f^{\circ 1} = f$ and $f^{\circ 2} = f \circ f$.

94B. Corollary. *Let the set D, the member z of D, and the mapping $f\colon D \to D$ be given. Then there is exactly one sequence $a \in D^{\mathbb{N}}$ such that*

$$a_0 = z \quad \text{and} \quad a \circ \mathrm{seq} = f \circ a, \tag{94.3}$$

namely $a := (f^{\circ n}(z) \mid n \in \mathbb{N})$.

Proof. Although this is a direct consequence of Theorem 93B and (94.1) and (94.2), we can also obtain it from Theorem 94A. Let a be a sequence satisfying (94.3). Then $a_0 = z = f^{\circ 0}(z)$. If $a_n = f^{\circ n}(z)$, then $a_{\mathrm{seq}n} = f(a_n) = f(f^{\circ n}(z)) = f^{\circ \mathrm{seq}n}(z)$. We have proved by induction that $a_n = f^{\circ n}(z)$ for all $n \in \mathbb{N}$. Hence there is at most one sequence a satisfying (94.3). But $a := (f^{\circ n}(z) \mid n \in \mathbb{N})$ obviously does satisfy (94.3). ■

94C. Corollary. $\mathrm{seq}^{\circ n}(0) = n$ *for all* $n \in \mathbb{N}$.

Proof. Apply Corollary 94B to $D := \mathbb{N}, z := 0$, and $f := \mathrm{seq}$. ■

In the following propositions, we collect some useful facts about iterates.

94D. Proposition. *Let the set D and the mappings $f, g \in \mathrm{Map}(D, D)$ be given, and assume that f and g commute. Then the following rules hold:*

$$1_D{}^{\circ n} = 1_D \text{ for all } n \in \mathbb{N} \tag{94.4}$$

$$f^{\circ m} \text{ and } g \text{ commute for all } m \in \mathbb{N} \tag{94.5}$$

(94.6) $f^{\circ m}$ *and* $g^{\circ n}$ *commute for all* $m, n \in \mathbb{N}$

(94.7) $(f \circ g)^{\circ n} = f^{\circ n} \circ g^{\circ n}$ *for all* $n \in \mathbb{N}$

(94.8) $(f^{\circ m})^{\circ n} = (f^{\circ n})^{\circ m}$ *for all* $m, n \in \mathbb{N}$

(94.9) *if* f *is invertible, then* $f^{\circ n}$ *is invertible and* $(f^{\circ n})^{\leftarrow} = (f^{\leftarrow})^{\circ n}$ *for all* $n \in \mathbb{N}$.

Proof. (94.4) and (94.5) are proved by induction. (94.6) follows from two successive applications of (94.5). (94.7) is proved by induction, using (94.5). (94.8) is proved by induction, using (94.4) and (94.7). (94.9) follows directly from (94.4) and (94.7). ■

94E. Proposition. *Let the set* D *and the mapping* $f\colon D \to D$ *be given. If* f *is injective, every iterate of* f *is injective. If* f *is surjective, every iterate of* f *is surjective.*

Proof. Assume that f is injective [surjective]. Now $f^{\circ 0} = 1_D$ is injective [surjective]; and if $f^{\circ n}$ is injective [surjective], then so is $f^{\circ \mathrm{seq}n} = f \circ f^{\circ n}$ (Proposition 32B.L,(a) [Proposition 32B.R,(a)]). This completes the inductive proof. ■

▼ In the light of Remark 91A, it is interesting to note that the conclusion of Corollary 94B (omitting the explicit form of a) provides a *characterization* of counting systems, as we proceed to show. This fact was observed by Francis William Lawvere (b. 1937), who used this property to *define* what is here called a counting system.

94F. Theorem. *Let the set* N*, the member* 0 *of* N*, and the mapping* $\mathrm{seq} : N \to N$ *be given. The following statements are equivalent:*

(i): N *is a counting system with* 0 *as zero and* seq *as successor-mapping.*

(ii): *for every set* D*, every member* z *of* D*, and every mapping* $f\colon D \to D$ *there exists exactly one mapping* $\phi : N \to D$ *such that* $\phi(0) = z$ *and* $\phi \circ \mathrm{seq} = f \circ \phi$.

Proof. (i) ⇒ (ii). This follows from Corollary 94B (using the language of mappings rather than that of families/sequences) and Remark 91A.

(ii) ⇒ (i). 1. We first apply the assumption to $D := N$, $z := 0$, $f := \mathrm{seq}$, and conclude that if a mapping $\psi : N \to N$ satisfies $\psi(0) = 0$ and $\psi \circ \mathrm{seq} = \mathrm{seq} \circ \psi$, then $\psi = 1_N$.

2. Choose distinct objects a and b (e.g., $a := \emptyset$ and $b := \{\emptyset\}$), and set $D := \{a, b\}, z := a$, $f := b_{D \to D}$. Consider the only $\phi : N \to D$ such that $\phi(0) = z = a$ and $\phi \circ \mathrm{seq} = f \circ \phi = b_{N \to D}$. Then $\phi(0) = a \neq b$, but $\phi_{>}(\mathrm{Rng\ seq}) \subset \{b\}$. Therefore $0 \notin \mathrm{Rng\ seq}$, and (Count I) holds.

3. Define the mappings

$$\pi := (m, n) \mapsto m : N \times N \to N \quad \text{and} \quad \pi' := (m, n) \mapsto n : N \times N \to N,$$

so that $\xi = (\pi(\xi), \pi'(\xi))$ for all $\xi \in N \times N$.

Set $D := N \times N$, $z := (0,0) \in D$, and $f := (m,n) \mapsto (\mathrm{seq}m, m) : D \to D$, and consider the only mapping $\phi : N \to D$ such that $\phi(0) = z = (0,0)$ and $\phi \circ \mathrm{seq} = f \circ \phi$. We observe that

$$(\phi \circ \mathrm{seq})(n) = (f \circ \phi)(n) = f((\pi(\phi(n)), \pi'(\phi(n)))) = (\mathrm{seq}\pi(\phi(n)), \pi(\phi(n))) \quad \text{for all } n \in N. \tag{94.10}$$

We have $(\pi \circ \phi)(0) = \pi(\phi(0)) = \pi((0,0)) = 0$ and, by (94.10),

$$((\pi \circ \phi) \circ \mathrm{seq})(n) = \pi((\mathrm{seq}\pi(\phi(n)), \pi(\phi(n)))) = \mathrm{seq}\pi(\phi(n)) = (\mathrm{seq} \circ (\pi \circ \phi))(n) \quad \text{for all } n \in N.$$

It follows from Part 1 applied to $\psi := \pi \circ \phi$ that

$$\pi \circ \phi = 1_N. \tag{94.11}$$

From (94.10) and (94.11) we obtain $(\phi \circ \mathrm{seq})(n) = (\mathrm{seq}n, n)$ and hence

$$(\pi' \circ \phi \circ \mathrm{seq})(n) = \pi'((\mathrm{seq}n, n)) = n \qquad \text{for all } n \in N.$$

Therefore $(\pi' \circ \phi) \circ \mathrm{seq} = 1_N$, and hence seq is injective; i.e., (Count II) holds.

4. Let $S \in \mathfrak{P}(N)$ be given, and assume that $0 \in S$ and $\mathrm{seq}_>(S) \subset S$. Applying the assumption to $D := S$, $z := 0$, and $f := \mathrm{seq}|^S_S$, we conclude that there is exactly one $\phi\colon N \to S$ such that $\phi(0) = 0$ and $\phi \circ \mathrm{seq} = (\mathrm{seq}|^S_S) \circ \phi$. Then $(1_{S \subset N} \circ \phi)(0) = \phi(0) = 0$ and

$$(1_{S \subset N} \circ \phi) \circ \mathrm{seq} = 1_{S \subset N} \circ (\mathrm{seq}|^S_S) \circ \phi = \mathrm{seq} \circ (1_{S \subset N} \circ \phi).$$

By Part 1 applied to $\psi := 1_{S \subset N} \circ \phi$ we have $1_{S \subset N} \circ \phi = 1_N$. Therefore $1_{S \subset N}$ is surjective, and hence $S = N$. We conclude that (Count III) holds. ■

▲

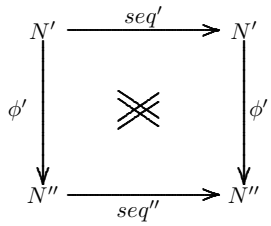
95. Essential uniqueness of counting systems

In this section we digress briefly from our study of the Natural-Number System to dispose of an issue suggested in Section 91, and perhaps allay some of the uneasiness that may be felt about singling out one specific counting system as *the* Natural-Number System. We shall show that any two counting systems are in fact *naturally isomorphic*, in that there is *exactly one* invertible mapping from one to the other such that both this mapping and its inverse preserve the structure consisting of the zero and the successor-mapping.

In the following theorem, N' and N'' are counting systems. We distinguish their zeros by denoting them by $0'$ and $0''$, and their successor-mappings by denoting them by seq' and seq'', respectively.

95A. Theorem. *Let the counting systems N' and N'' be given. There exists exactly one mapping $\phi' : N' \to N''$ such that $\phi'(0') = 0''$ and $\phi' \circ \text{seq}' = \text{seq}'' \circ \phi'$. There exists exactly one mapping $\phi'' : N'' \to N'$ such that $\phi''(0'') = 0'$ and $\phi'' \circ \text{seq}'' = \text{seq}' \circ \phi''$. Each of the mappings ϕ' and ϕ'' is the inverse of the other.*

Proof. In view of Remark 91A, Corollary 94B is valid with N', $0'$, seq' instead of $\mathbb{N}$, 0, seq, respectively. We apply it with $D := N''$, $z := 0''$, $f := \text{seq}''$, and deduce that there is exactly one mapping $\phi' : N' \to N''$ such that $\phi'(0') = 0''$ and $\phi' \circ \text{seq}' = \text{seq}'' \circ \phi'$.



Interchanging the counting systems N' and N'' in the preceding argument, we deduce that there also is exactly one mapping $\phi'' : N'' \to N'$ such that $\phi''(0'') = 0'$ and $\phi'' \circ \text{seq}'' = \text{seq}' \circ \phi''$.

Repeating the argument with appropriate choices, we further deduce that there is exactly one mapping $\psi' : N' \to N'$ such that $\psi'(0') = 0'$ and $\psi' \circ \text{seq}' = \text{seq}' \circ \psi'$. However, on the one hand $1_{N'}(0') = 0'$ and $1_{N'} \circ \text{seq}' = \text{seq}' \circ 1_{N'}$, and on the other hand $(\phi'' \circ \phi')(0') = \phi''(0'') = 0'$ and $\phi'' \circ \phi' \circ \text{seq}' = \phi'' \circ \text{seq}'' \circ \phi' = \text{seq}' \circ \phi'' \circ \phi'$. Therefore $\phi'' \circ \phi' = \psi' = 1_{N'}$.

Interchanging N' and N'', ϕ' and ϕ'' in the preceding argument, we conclude that $\phi' \circ \phi'' = 1_{N''}$. ■

95B. Example. For the purpose of this example only, a collection of sets is said to be **saturated** if it contains $\emptyset$ and also contains, for each of its member sets S, the set $S \cup \{S\}$. It is clear that the intersection of every nonempty collection of saturated collections is a saturated collection. Now suppose that there exists *some* saturated

collection. It follows easily that there is exactly one saturated collection $\mathcal{N}$ that is the *smallest*, in that it is included in every saturated collection. We claim that $\mathcal{N}$, with $\emptyset$ as *zero*, and $(S \mapsto S \cup \{S\}) : \mathcal{N} \to \mathcal{N}$ as the *successor-mapping*, is a counting system. (Count I) is trivially satisfied, since $S \in S \cup \{S\}$ for every set S. (Count III) is satisfied precisely because $\mathcal{N}$ is the smallest saturated collection. To prove that (Count II) holds, we observe first that $\{S \in \mathcal{N} \mid S \subset \mathfrak{P}(S)\}$ is a saturated collection included in $\mathcal{N}$, and therefore

$$S \subset \mathfrak{P}(S) \qquad \text{for all } S \in \mathcal{N}. \tag{95.1}$$

Now suppose that $S, T \in \mathcal{N}$ satisfy $S \cup \{S\} = T \cup \{T\}$. It follows that $T \in S \cup \{S\}$, i.e., $T = S$ or $T \in S$. From (95.1) it then follows that $T \subset S$. Interchanging S and T in this argument, we infer that $S \subset T$, and therefore conclude that $S = T$. This shows that (Count II) holds.

Theorem 95A now informs us that there is exactly one mapping $\phi : \mathbb{N} \to \mathcal{N}$ such that $\phi(0) = \emptyset$ and $\phi(\mathrm{seq}n) = \phi(n) \cup \{\phi(n)\}$ for all $n \in \mathbb{N}$, and that this mapping is invertible.

The counting system $\mathcal{N}$ has some exotic properties, derived from its construction, and not from (Count I), (Count II), (Count III). One of these is (95.1); another is the fact that the strict-order $<$ (the transitive closure of the successor-relation in $\mathcal{N}$) satisfies the equivalence

$$\forall S, T \in \mathcal{N}, \quad S < T \;\Leftrightarrow\; S \in T,$$

and therefore $S^{\sqsubset} := [\![\emptyset, S[\![= S$ for every $S \in \mathcal{N}$.

Despite these and other peculiarities, the absolute uniqueness of a smallest saturated collection of sets has proved for many mathematicians an irresistible temptation to declare the counting system $\mathcal{N}$ to be *the* Natural-Number System itself. This turns out to be technically attractive in the specialized branch of mathematics called axiomatic set theory. (The assumption that there exists *some* saturated collection of sets is, in this theory, a foundational agreement called the Axiom of Infinity.) For our purposes, however, this identification is an unnecessary and perhaps objectionable departure from intuition. ■

▼ For another approach to constructing a counting system under an apparently
▲ weaker foundational agreement, see Section 106, and in particular Theorem 106L.

96. Addition and subtraction

Our next aim is to define the familiar arithmetical operations in $\mathbb{N}$ and to verify their elementary properties.

In Section 91 we proposed to define *addition* by "counting on". Accordingly, we define the mapping $((m,n) \mapsto m+n) : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, called (the **operation of**) **addition** (**of natural numbers**), by the rule

$$m + n := \mathrm{seq}^{\circ n}(m) \qquad \text{for all } m, n \in \mathbb{N}. \tag{96.1}$$

The symbol + is read "**plus**", and $m+n$ is called the **sum of** m **and** n (m and n are the **summands**), and is said to be obtained by **adding** m **and** n or **adding** n **to** m.

From Corollary 94C we have $m = \mathrm{seq}^{\circ m}(0)$ for all $m \in \mathbb{N}$, and therefore

$$m + n = (\mathrm{seq}^{\circ n} \circ \mathrm{seq}^{\circ m})(0) \qquad \text{for all } m, n \in \mathbb{N}. \tag{96.2}$$

Combining (96.1) and (96.2) we find

$$(m+n)+p = \mathrm{seq}^{\circ p}(m+n) = (\mathrm{seq}^{\circ p} \circ \mathrm{seq}^{\circ n} \circ \mathrm{seq}^{\circ m})(0) \quad \text{for all } m, n, p \in \mathbb{N}. \tag{96.3}$$

96A. Proposition. *Addition satisfies the following rules, valid for all* $m, n, p \in \mathbb{N}$*:*

$$m + \mathrm{seq}n = \mathrm{seq}(m+n) \tag{96.4}$$

$$m + n = 0 \;\Leftrightarrow\; m = n = 0 \tag{96.5}$$

$$n + 0 = n \tag{96.6}$$

$$n + 1 = \mathrm{seq}n \tag{96.7}$$

$$m + n = n + m \qquad (\textit{commutative law}) \tag{96.8}$$

$$(m+n)+p = m+(n+p) \qquad (\textit{associative law}) \tag{96.9}$$

$$m + p = n + p \;\Rightarrow\; m = n \qquad (\textit{cancellation law}). \tag{96.10}$$

Proof. (96.4), (96.6), (96.7) are immediate consequences of (96.1). (96.5) follows from (96.4) and Proposition 91B.

From (94.6) applied to $D := \mathbb{N}$ and $f := g := \mathrm{seq}$, we learn that all iterates of seq commute. From (96.2) we therefore get the commutative law (96.8); and from (96.3) we get

$$(m+n)+p = (n+p)+m \qquad \text{for all } m, n, p \in \mathbb{N}.$$

Applying the commutative law to the right-hand side, we obtain the associative law (96.9).

By (ℕII), seq is injective. By Proposition 94E, every iterate of seq is injective. Applying this and (96.1), we obtain the cancellation law (96.10). ■

Because of the associative law (96.9), we may unambiguously write $m + n + p$; and a similar license is in effect for sums of more summands.

96B. Proposition. *Let the set D and the mapping $f\colon D \to D$ be given. Then:*

$$f^{\circ(m+n)} = f^{\circ n} \circ f^{\circ m} \qquad \text{for all } m, n \in \mathbb{N}. \tag{96.11}$$

Proof. By (96.6), $f^{\circ(m+0)} = f^{\circ m} = f^{\circ 0} \circ f^{\circ m}$. If $f^{\circ(m+n)} = f^{\circ n} \circ f^{\circ m}$, then (96.4) implies $f^{\circ(m+\mathrm{seq}n)} = f^{\circ \mathrm{seq}(m+n)} = f \circ f^{\circ(m+n)} = f \circ f^{\circ n} \circ f^{\circ m} = f^{\circ \mathrm{seq}n} \circ f^{\circ m}$. This completes the inductive proof of (96.11). ■

The relationship between addition and order is based on the following result.

96C. Proposition.

$$\forall m, n \in \mathbb{N}, \quad m < n \;\Leftrightarrow\; (\exists p \in \mathbb{N}^{\times}, \quad n = m + p).$$

Proof. We define the relation ρ in $\mathbb{N}$ by the rule

$$\forall m, n \in \mathbb{N}, \quad m \;\rho\; n \;:\Leftrightarrow\; (\exists p \in \mathbb{N}^{\times}, \quad n = m + p). \tag{96.12}$$

From (96.12), Proposition 91B, and (96.4) we have

$$\begin{aligned} \forall m, n \in \mathbb{N}, \quad m \;\rho\; n \;&\Leftrightarrow\; (\exists p \in \mathbb{N}, \quad n = m + \mathrm{seq}p) \qquad \Leftrightarrow \\ &\Leftrightarrow\; (\exists p \in \mathbb{N}, \quad n = \mathrm{seq}(m + p)). \end{aligned} \tag{96.13}$$

From (96.13) and (ℕI) it follows that $m \;\rho\; n$ requires $n \neq 0$, so that ρ satisfies (92.5). From (96.13), (ℕII), and (96.12) we have the following chain of equivalences for all $m, n \in \mathbb{N}$:

$$\begin{aligned} m \;\rho\; \mathrm{seq}n \;&\Leftrightarrow\; (\exists p \in \mathbb{N}, \quad \mathrm{seq}n = \mathrm{seq}(m + p)) \;\Leftrightarrow\; (\exists p \in \mathbb{N}, \quad n = m + p) \;\Leftrightarrow \\ &\Leftrightarrow\; (m = n \text{ or } (\exists p \in \mathbb{N}^{\times}, \quad n = m + p)) \;\Leftrightarrow\; (m = n \text{ or } m \;\rho\; n). \end{aligned}$$

Therefore ρ satisfies (92.6). Proposition 92A then shows that ρ is $<$, as asserted. ■

We now consider the following equation, for given $m, n \in \mathbb{N}$:

$$?p \in \mathbb{N}, \quad m + p = n. \tag{96.14}$$

Proposition 96C and the cancellation law allow us to solve the existence and uniqueness problems for (96.14).

96D. Theorem. *For given $m, n \in \mathbb{N}$, the equation* (96.14) *has at most one solution. It has exactly one solution if and only if $m \leq n$.*

Proof. The cancellation law (96.10) shows that (96.14) has at most one solution. The number 0 is a solution if and only if $m = n$. Proposition 96C shows that there is a solution in $\mathbb{N}^\times$ if and only if $m < n$. ■

When $m \le n$, the unique solution of (96.14) is denoted by $n - m$, read "n **minus** m", and called the **difference of** n **and** m. We have thus defined a mapping $((n, m) \mapsto n - m) : \mathrm{Gr}(\ge) \to \mathbb{N}$, called (the **operation of**) **subtraction** (**of natural numbers**), by the rule

$$n - m :\in \{p \in \mathbb{N} \mid m + p = n\} \qquad \text{for all } (n, m) \in \mathrm{Gr}(\ge).$$

$n - m$ is said to be obtained by **subtracting** m (the **subtrahend**) **from** n (the **minuend**).

96E. Proposition. *Addition and subtraction satisfy the following rules, valid for all $m, n, p \in \mathbb{N}$:*

$$n - 0 = n \qquad \mathrm{seq}n - 1 = n$$

$$\mathrm{seq}(n - 1) = n \quad \textit{if } n \in \mathbb{N}^\times$$

$$(m + n) - p = m + (n - p) \quad \textit{if } p \le n$$

$$(m + n) - p = m - (p - n) \quad \textit{if } n \le p \le m + n$$

$$m - (n + p) = (m - n) - p \quad \textit{if } n + p \le m.$$

The usual rules about omitting parentheses when additions and subtractions occur together will be observed: operations not otherwise given priority by parentheses are performed from left to right. Thus $m + n - p - q = ((m + n) - p) - q$.

96F. Proposition. *Addition, subtraction, and order satisfy the following rules (monotonicity laws), valid for all $m, n, p \in \mathbb{N}$:*

$$m < n \;\Leftrightarrow\; m + p < n + p$$

$$m < n \;\Leftrightarrow\; m - p < n - p \quad \textit{if } \; p \le m$$

$$m < n \;\Leftrightarrow\; p - n < p - m \quad \textit{if } \; n \le p.$$

Proof. For every $q \in \mathbb{N}^\times$ we have

$$m + q = n \;\Leftrightarrow\; m + p + q = n + p$$

$$m + q = n \;\Leftrightarrow\; m + q - p = n - p \;\Leftrightarrow\; m - p + q = n - p \quad \text{if } p \le m$$

$$m + q = n \;\Leftrightarrow\; m = n - q \;\Leftrightarrow\; p - (n - q) = p - m \;\Leftrightarrow\; p - n + q = p - m \quad \text{if } n \le p.$$

The assertions then respectively follow by Proposition 96C. ■

We mention some useful notational conventions. If A and B are subsets of $\mathbb{N}$, we denote the image of $A \times B$ under the operation of addition by $A + B$. Thus,

$$A + B := \{m + n \mid m \in A,\ n \in B\}.$$

In the same way, if $A \times B \subset \mathrm{Gr}(\geq)$, we denote the image of $A \times B$ under the operation of subtraction by $A - B$, so that

$$A - B := \{m - n \mid m \in A,\ n \in B\}.$$

If either one of A or B is a singleton of the form $\{p\}$, it is customary to omit the braces if confusion is unlikely; for instance, one writes $p + B$ instead of $\{p\} + B$. Examples of these notations are $\mathbb{N} + \mathbb{N} = \mathbb{N}$; $\mathbb{N}^\times = 1 + \mathbb{N}$; $n^\sqcap - 1 = n^\sqsubset$ for each $n \in \mathbb{N}$; $\mathbb{N}^\times + \mathbb{N}^\times = 2 + \mathbb{N}$.

To conclude this section, we record an application of Proposition 96B to a characterization of the transitive closure of a relation.

96G. Proposition. *Let the set D and the relation ρ in D be given, and let τ be the transitive closure of ρ. Then*

$$\forall x, y \in D,\ x\ \tau\ y\ \Leftrightarrow\ (\exists n \in \mathbb{N}^\times,\ y \in (\rho_>)^{\circ n}(\{x\})).$$

Proof. We define the relation σ in D by the rule

$$\forall x, y \in D,\ x\ \sigma\ y\ :\Leftrightarrow\ (\exists n \in \mathbb{N}^\times,\ y \in (\rho_>)^{\circ n}(\{x\})).$$

For all $x, y \in D$, we have $x\ \rho\ y$ if and only if $y \in \rho_>(\{x\}) = (\rho_>)^{\circ 1}(\{x\})$, and hence σ is broader than ρ. If $x, y, z \in D$ satisfy $x\ \sigma\ y$ and $y\ \sigma\ z$, we may choose $m, n \in \mathbb{N}^\times$ such that $y \in (\rho_>)^{\circ m}(\{x\})$ and $z \in (\rho_>)^{\circ n}(\{y\})$. By (96.5) we have $m + n \in \mathbb{N}^\times$. By Proposition 96B we have

$$z \in (\rho_>)^{\circ n}(\{y\}) \subset (\rho_>)^{\circ n}((\rho_>)^{\circ m}(\{x\})) = (\rho_>)^{\circ(m+n)}(\{x\}),$$

so that $x\ \sigma\ z$. Thus σ is transitive. Since τ is the transitive closure of ρ, we conclude that σ is broader than τ.

Let $x \in D$ be given. Since ρ is narrower than τ, we have $(\rho_>)^{\circ \mathrm{seq}0}(\{x\}) = \rho_>(\{x\}) \subset \tau_>(\{x\})$. If $n \in \mathbb{N}$ is such that $(\rho_>)^{\circ \mathrm{seq}n}(\{x\}) \subset \tau_>(\{x\})$, then $(\rho_>)^{\circ \mathrm{seq}\ \mathrm{seq}n}(\{x\}) = \rho_>((\rho_>)^{\circ \mathrm{seq}n}(\{x\})) \subset \rho_>(\tau_>(\{x\})) \subset \tau_>(\tau_>(\{x\})) \subset \tau_>(\{x\})$, since τ is transitive. We have proved by induction that $(\rho_>)^{\circ \mathrm{seq}n}(\{x\}) \subset \tau_>(\{x\})$ for all $n \in \mathbb{N}$, so that $\sigma_>(\{x\}) \subset \tau_>(\{x\})$. Since $x \in D$ was arbitrary, this implies that σ is narrower than τ. Since σ was also broader than τ, we conclude that $\sigma = \tau$. ∎

97. Multiplication and division

We shall define *multiplication* by iterated addition. We accordingly define the mapping $((m,n) \mapsto m \cdot n)) : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, called (the **operation of**) **multiplication** (**of natural numbers**), by the rule

$$m \cdot n := (\mathrm{seq}^{\circ m})^{\circ n}(0) \qquad \text{for all } m,n \in \mathbb{N}. \tag{97.1}$$

The symbol $\cdot$ is read "**times**" or "**by**", and $m \cdot n$ is called the **product of** m **and** n (the **factors**), and is said to be obtained by **multiplying** m **and** (or **by**) n. The multiplication sign $\times$ is sometimes used instead of $\cdot$ for special emphasis, but is best avoided. Most frequently, the symbol $\cdot$ is omitted altogether, the names of the factors being merely juxtaposed; but it is retained when juxtaposition would be confusing (thus $56 \neq 5 \cdot 6 = 30$). The usual rules about omitting parentheses will be observed; multiplication has priority over addition and subtraction.

Applying (94.6) twice and using (96.1) and (97.1), we find

$$\begin{aligned} (\mathrm{seq}^{\circ p})^{\circ n}(m) &= ((\mathrm{seq}^{\circ p})^{\circ n} \circ \mathrm{seq}^{\circ m})(0) = (\mathrm{seq}^{\circ m} \circ (\mathrm{seq}^{\circ p})^{\circ n})(0) = \\ &= \mathrm{seq}^{\circ m}(pn) = pn + m \qquad \text{for all } m,n,p \in \mathbb{N}. \end{aligned} \tag{97.2}$$

97A. Proposition. *Multiplication and addition satisfy the following rules, valid for all* $m,n,p \in \mathbb{N}$*:*

$$m \cdot \mathrm{seq} n = mn + m \tag{97.3}$$

$$n \cdot 0 = 0 \tag{97.4}$$

$$n \cdot 1 = n \tag{97.5}$$

$$mn = nm \qquad (\textit{commutative law}) \tag{97.6}$$

$$p(m+n) = pm + pn \qquad (\textit{distributive law}) \tag{97.7}$$

$$m,n \in \mathbb{N}^\times \;\Leftrightarrow\; mn \in \mathbb{N}^\times \tag{97.8}$$

$$mn = 1 \;\Leftrightarrow\; m = n = 1. \tag{97.9}$$

Proof. (97.3), (97.4), (97.5) are trivial consequences of the definitions. The commutative law follows from (97.1) and (94.8). To prove the distributive law we apply Proposition 96B together with (97.1), (96.1), and (97.2):

$$p(m+n) = (\mathrm{seq}^{\circ p})^{\circ (m+n)}(0) = ((\mathrm{seq}^{\circ p})^{\circ n} \circ (\mathrm{seq}^{\circ p})^{\circ m})(0) = (\mathrm{seq}^{\circ p})^{\circ n}(pm) = pm + pn.$$

To prove (97.8), assume that $m, n \in \mathbb{N}^\times$. Using Propositions 96E and 96C, we find

$$mn = m \cdot \mathrm{seq}(n-1) = m(n-1) + m = m + m(n-1) \geq m > 0,$$

so that $mn \in \mathbb{N}^\times$. The reverse implication follows from (97.4) and the commutative law. To prove (97.9), assume that $mn = 1$. By (97.8) we must have $m, n \in \mathbb{N}^\times$. Therefore

$$1 = mn = m(n-1) + m = m + m(n-1) \geq m \geq 1$$

$$1 = mn = nm = n(m-1) + n = n + n(m-1) \geq n \geq 1,$$

which implies $m = n = 1$. The reverse implication follows from (97.5) and the commutative law. ■

97B. Proposition. *Multiplication, subtraction, and order satisfy the following rules, valid for all $m, n \in \mathbb{N}$ and $p \in \mathbb{N}^\times$:*

(97.10) $$m < n \;\Leftrightarrow\; pm < pn \qquad (\textit{isotonicity law})$$

(97.11) $$m = n \;\Leftrightarrow\; pm = pn \qquad (\textit{cancellation law})$$

(97.12) $$p(n-m) = pn - pm \;\; \textit{if } m \leq n \qquad (\textit{distributive law}).$$

Proof. If $m < n$, then $n - m \in \mathbb{N}^\times$. By the distributive law (97.7), we have $pm + p(n-m) = p(m + (n-m)) = pn$, which proves (97.12) (the case $m = n$ is trivial). By (97.8) we have $p(n-m) \in \mathbb{N}^\times$, and Proposition 96C shows that $pm < pn$. We have just shown that the mapping $(k \mapsto pk) : \mathbb{N} \to \mathbb{N}$ is strictly isotone. Since $\mathbb{N}$ is totally ordered, it follows that the mapping is injective — which proves (97.11) — and that the equivalence (97.10) holds (Remark 62A,(a)). ■

97C. Proposition. *Let the set D and the mapping $f\colon D \to D$ be given. Then*

$$f^{\circ mn} = (f^{\circ m})^{\circ n} \qquad \textit{for all } m, n \in \mathbb{N}.$$

Proof. $f^{\circ(m\cdot 0)} = f^{\circ 0} = 1_D = (f^{\circ m})^{\circ 0}$. If $f^{\circ mn} = (f^{\circ m})^{\circ n}$, then (97.3) and Proposition 96B yield $f^{\circ(m\cdot \mathrm{seq}n)} = f^{\circ(mn+m)} = f^{\circ m} \circ f^{\circ mn} = f^{\circ m} \circ (f^{\circ m})^{\circ n} = (f^{\circ m})^{\circ \mathrm{seq}n}$. This completes the inductive proof. ■

97D. Corollary. *Multiplication satisfies the following rule (associative law) valid for all $m, n, p \in \mathbb{N}$:*

$$(mn)p = m(np).$$

Proof. Using (97.1) and Proposition 97C, we find

$$(mn)p = (\mathrm{seq}^{\circ mn})^{\circ p}(0) = ((\mathrm{seq}^{\circ m})^{\circ n})^{\circ p}(0) = (\mathrm{seq}^{\circ m})^{\circ np})(0) = m(np). \blacksquare$$

Because of the associative law, we may unambiguously write mnp, without parentheses; and a similar license is in effect for products of more factors.

For images under the operation of multiplication we use a notation similar to that previously introduced for addition and subtraction. Thus, if A and B are subsets of $\mathbb{N}$, we define

$$AB := \{mn \mid m \in A,\ n \in B\}.$$

It is also customary to write pB and Ap instead of $\{p\}B$ and $A\{p\}$, and we shall usually do so. For instance, the range of the familiar arithmetic progression with initial term m and difference p is $m + p\mathbb{N}$.

For each $p \in \mathbb{N}$ we define the mapping $p\cdot : \mathbb{N} \to \mathbb{N}$, called **multiplication by** p, by the rule

$$p\cdot(n) := pn \qquad \text{for all } n \in \mathbb{N}.$$

On account of (97.4), (97.5), and the commutative and associative laws, these mappings satisfy the rules

$$0\cdot = 0_{\mathbb{N}\to\mathbb{N}} \qquad\qquad 1\cdot = 1_{\mathbb{N}} \tag{97.13}$$

$$(p\cdot) \circ (q\cdot) = pq\cdot \qquad \text{for all } p, q \in \mathbb{N}, \tag{97.14}$$

and by (97.10) $p\cdot$ is strictly isotone for every $p \in \mathbb{N}^\times$. We also have $\operatorname{Rng}(p\cdot) = p\mathbb{N}$ for all $p \in \mathbb{N}$.

We can use these mappings to define *exponentiation* by iterated multiplication. Accordingly, we define the mapping $((m, n) \mapsto m^n) : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, called (the **operation of**) **exponentiation** (**of natural numbers**), by the rule

$$m^n := (m\cdot)^{\circ n}(1) \qquad \text{for all } m, n \in \mathbb{N}.$$

The symbol m^n is read "m **to the** (**power**) n" or "m **to the** n**th** (**power**)", and m^n is said to be obtained by **raising** m (the **base**) **to the power** n (the **exponent**).

97E. Proposition. *Exponentiation, multiplication, and addition satisfy the following rules, valid for all $m, n, p \in \mathbb{N}$:*

$$m^0 = 1 \qquad m^1 = m \qquad m^2 = mm \qquad 1^n = 1$$

$$0^n = 0 \qquad \textit{if } n \in \mathbb{N}^\times$$

$$(mn)^p = m^p n^p$$

$$p^{m+n} = p^m p^n \qquad p^{mn} = (p^m)^n$$

$$n < m^n \qquad \textit{if } m > 1.$$

It is useful to note that

$$4 = 2 + 2 = 2 \cdot 2 = 2^2. \tag{97.15}$$

We now turn to the concept of *division*. For a given pair $(m, n) \in \mathbb{N} \times \mathbb{N}$, we consider the problem

$$?q \in \mathbb{N}, \qquad qn \le m \text{ and } m - qn < n. \tag{97.16}$$

97F. Theorem. (97.16) *has no solution for any pair* (m,n) *with* $n = 0$. *For each pair* $(m,n) \in \mathbb{N} \times \mathbb{N}^\times$, *there is exactly one solution of* (97.16), *namely* $\max\{p \in \mathbb{N} \mid pn \leq m\}$.

Proof. The statement concerning the case $n = 0$ holds trivially. Assume now that $n \in \mathbb{N}^\times$. The set $\{p \in \mathbb{N} \mid pn \leq m\}$ contains 0, and m is an upper bound: indeed, $1 \leq n$ and therefore $pn \leq m$ implies $pn \leq m \cdot 1 \leq mn$, whence $p \leq m$ (Proposition 97B). By Proposition 92F this set has a maximum. Define $q_0 := \max\{p \in \mathbb{N} \mid pn \leq m\}$.

We now use the rules in Propositions 96F and 97B. If q is a solution of (97.16), we must have $qn \leq m$, and therefore $q \leq q_0$. On the other hand, we have $m - qn < n$, and therefore $q_0 n \leq m < qn + n = (\mathrm{seq}q)n$. This implies $q_0 < \mathrm{seq}q$, and this in turn is equivalent to $q_0 \leq q$. We conclude that q must be q_0, so that there is at most one solution of (97.16).

However, $q_0 n \leq m < (q_0+1)n$, and therefore $m - q_0 n < (q_0+1)n - q_0 n = 1 \cdot n = n$, so that q_0 is indeed a solution of (97.16). ∎

For every pair $(m,n) \in \mathbb{N} \times \mathbb{N}^\times$, the unique solution $q := \max\{p \in \mathbb{N} \mid pn \leq m\}$ of (97.16) and the corresponding number $m - qn$ are called the **quotient** and the **remainder** (or **residue**), respectively, **of the division of** m (the **dividend**) **by** n (the **divisor**). When the remainder is 0, it is customary to denote the quotient by m/n or $\frac{m}{n}$. (Theorem 97F rules out "division by 0".)

Let the number $m \in \mathbb{N}$ be given. Since $2^\sqsubset = \{0,1\}$, the remainder of the division of m by 2 is either 0 or 1. If the remainder is 0, m is said to be **even**; if the remainder is 1, m is said to be **odd**. Thus m is even if and only if $m \in 2\mathbb{N}$, and m is odd if and only if $m \in \mathbb{N}\backslash(2\mathbb{N}) = 2\mathbb{N} + 1$.

98. Divisors and multiples

We record some additional terminology relating to multiplication. Let the natural numbers m, n be given. If $n = mp$ for some $p \in \mathbb{N}$, we say that m **divides** n, and that n is **divisible by** m; and m is called a **divisor of** n, and n a **multiple of** m. (The clash between this use of the term *divisor* and the use associated with division (Theorem 97F) is unlikely to lead to confusion.) Thus m divides n if and only if $n \in m\mathbb{N}$. The relation "divides" and its reverse, "is a multiple of", were already mentioned in several examples in Sections 55, 56, 61, 62, and 64, anticipating some of the results to be proved here and adding some comments.

98A. Proposition. *The relation "divides" is an order in* $\mathbb{N}$. 1 *divides every natural number, and every natural number divides* 0.

Proof. The reflexivity and the properties of 0 and 1 follow from (97.4), (97.5), and the commutative law (97.6). The transitivity follows from the associative law. It remains to prove the antisymmetry. If m divides n and n divides m, we may choose $p, q \in \mathbb{N}$ such that $mp = n$ and $nq = m$. Then $nqp = mp = n = n \cdot 1$. If $n = 0$, then $m = 0 \cdot q = 0$, so that $m = n$. If $n \in \mathbb{N}^\times$, the cancellation law yields $qp = 1$, and (97.9) requires $q = p = 1$. Therefore $m = n \cdot 1 = n$ in this case also. ■

To distinguish the set $\mathbb{N}$ ordered by "divides" from the set $\mathbb{N}$ ordered by $\leq$, special terminology is used whenever possible. Instead of *upper bound, lower bound, supremum, infimum*, the specific terms **common multiple**, **common divisor**, **least common multiple**, **greatest common divisor** are used, and the symbols sup and inf are replaced by lcm and gcd, respectively. According to Proposition 98A, the numbers 1 and 0 are, respectively, the minimum and the maximum of $\mathbb{N}$ ordered by "divides".

98B. Lemma. *The mapping* $1_{\mathbb{N}^\times}$ *is a strictly isotone mapping from* $\mathbb{N}^\times$ *ordered by "divides" to* $\mathbb{N}^\times$ *ordered by* $\leq$.

Proof. If $m, n \in \mathbb{N}^\times$ and $p \in \mathbb{N}$ satisfies $n = mp$, we must have $p \in \mathbb{N}^\times$, by (97.8), and therefore $1 \leq p$. By the isotonicity law, $m = m \cdot 1 \leq mp = n$. Therefore $1_{\mathbb{N}^\times}$ is isotone; since it is injective, it is strictly isotone. ■

98C. Lemma. *The collection* $\{p\mathbb{N} \mid p \in \mathbb{N}\} \subset \mathfrak{P}(\mathbb{N})$ *is intersection-stable.*

Proof. Let the subset A of $\mathbb{N}$ be given, and set $S := \bigcap_{p \in A}^{\mathbb{N}} p\mathbb{N}$. By Proposition 98A, we have $0 \in S$. If $S = \{0\}$, then $S = 0 \cdot \mathbb{N}$. We may therefore assume that $S \cap \mathbb{N}^\times \neq \emptyset$ (which implies $0 \notin A$), and set $d := \min(S \cap \mathbb{N}^\times)$. We claim that $S = d\mathbb{N}$.

For every $p \in A$, we have $d \in p\mathbb{N}$ and therefore, by the associative law, also $d\mathbb{N} \subset p\mathbb{N}$. Therefore $d\mathbb{N} \subset S$.

To prove the reverse inclusion, let $n \in S$ be given, and let q be the quotient of the division of n by d. Let $p \in A$ be given. Since $n, d \in p\mathbb{N}$, it follows from Proposition 97B that $n - qd \in p\mathbb{N}$. Since $p \in A$ was arbitrary, we have shown that $n - qd \in S$. But the definition of q implies that $n - qd < d$. This is compatible with the definition of d only if $n - qd = 0$. But then $n \in d\mathbb{N}$. Since $n \in S$ was arbitrary, we have proved that $S \subset d\mathbb{N}$, and our claim that $S = d\mathbb{N}$ is established. ■

98D. Theorem. *The set* $\mathbb{N}$ *ordered by the relation "divides" is completely ordered and well-founded.*

Proof. 1. The set $\mathbb{N}^\times$ is well-ordered, hence well-founded by $\leq$. It follows from Lemma 98B and from Proposition 64B that $\mathbb{N}^\times$ ordered by "divides" is well-founded. It follows immediately that $\mathbb{N} = \mathbb{N}^\times \cup \{0\}$ ordered by "divides" is also well-founded. (Actually, 0 is a minimal member of $\{0\}$, and of no other subset of $\mathbb{N}$.)

2. For every subset A of $\mathbb{N}$, the set of common multiples of A is $\bigcap_{n\in A}{}^{\mathbb{N}} n\mathbb{N}$. By Lemma 98C, this set is $d\mathbb{N}$ for a suitable $d \in \mathbb{N}$. But d is a member of, and divides every member of, $d\mathbb{N}$; therefore d is the least common multiple of A. By Proposition 71B↑, it follows that $\mathbb{N}$ is completely ordered by the order "divides". ■

98E. The Euclidean Algorithm. The purpose of the procedure described here is to find, for given $m, p \in \mathbb{N}^\times$, the greatest common divisor of $\{m, p\}$. The success of the procedure rests on the following observation: *If* $(m, p) \in \mathbb{N} \times \mathbb{N}^\times$, *and if* r *is the remainder of the division of* m *by* p, *then* $\gcd\{m, p\} = \gcd\{p, r\}$. If q is the quotient of the same division, we have $r = m - qp$ and $m = r + qp$. It follows from these, together with the associative and distributive laws, that every common divisor of m and p also divides r, and every common divisor of p and r also divides m. This establishes the claimed observation.

Let us use the notation $\mathrm{rem}(u, v)$ for the remainder of the division of u by v (so that we have a mapping $\mathrm{rem} : \mathbb{N} \times \mathbb{N}^\times \to \mathbb{N}$). Informally, the algorithm proceeds by attempting to construct recursively a sequence $(p_n \mid n \in \mathbb{N})$ in which $p_0 := p, p_1 := \mathrm{rem}(m, p)$, and $p_{n+1} := \mathrm{rem}(p_{n-1}, p_n)$ for $n > 1$. This construction breaks off when, for some $k \in \mathbb{N}^\times$, we have $p_{k+1} = 0$. When it does, we have $p_k = \gcd\{m, p\}$.

More formally, this is an instance of a recursive definition with a rule that "breaks down". We formulate it as an application of Theorem 93C, and there is no formal reason to exclude the possibility that m or p might be 0.

We specify the data in the assumption of Theorem 93C as follows: $A_n := \mathbb{N} \times \mathbb{N}$ and $B_n := \mathbb{N} \times \mathbb{N}^\times$ for all $n \in \mathbb{N}$; $z := (m, p) \in \mathbb{N} \times \mathbb{N}$, and $h_n((u, v)) := (v, \mathrm{rem}(u, v))$ for all $n \in \mathbb{N}$ and $(u, v) \in \mathbb{N} \times \mathbb{N}^\times$. According to Theorem 93C there are now two alternatives. In the first, there is a sequence $((m_n, p_n) \mid n \in \mathbb{N})$ in $\mathbb{N} \times \mathbb{N}^\times$ such that $(m_0, p_0) = (m, p)$ and

$$(m_{\mathrm{seq}n}, p_{\mathrm{seq}n}) = (p_n, \mathrm{rem}(m_n, p_n)) \qquad \text{for all } n \in \mathbb{N}.$$

Now $p_{\mathrm{seq}n} = \mathrm{rem}(m_n, p_n) < p_n$ for all $n \in \mathbb{N}$. By Proposition 92G, the sequence $(p_n \mid n \in \mathbb{N})$ in $\mathbb{N}^\times$ is strictly antitone; but this is ruled out by the Principle of Descent (Proposition 92H). Therefore this alternative is excluded, and we are left with the other: there is a unique number $k \in \mathbb{N}$ and *list* $((m_n, p_n) \mid n \in (\mathrm{seq}k)^\sqsubset)$ such that

$$(m_0, p_0) = (m, p)$$

$$p_n \in \mathbb{N}^\times \text{ and } (m_{\mathrm{seq}n}, p_{\mathrm{seq}n}) = (p_n, \mathrm{rem}(m_n, p_n)) \qquad \text{for all } n \in k^\sqsubset$$

$$p_k \in \mathbb{N} \backslash \mathbb{N}^\times = \{0\}.$$

Now $\gcd\{m_{\mathrm{seq}n}, p_{\mathrm{seq}n}\} = \gcd\{p_n, \mathrm{rem}(m_n, p_n)\} = \gcd\{m_n, p_n\}$ for all $n \in k^{\sqsubset}$, according to our earlier observation. It then follows at once by induction (adapted trivially to the index set $(\mathrm{seq}k)^{\sqsubset}$) that $\gcd\{m_n, p_n\} = \gcd\{m_0, p_0\} = \gcd\{m, p\}$ for all $n \in (\mathrm{seq}k)^{\sqsubset}$. In particular, $m_k = \gcd\{m_k, 0\} = \gcd\{m_k, p_k\} = \gcd\{m, p\}$. If $k = 0$, we are in the trivial case $p = 0$, and verify that $\gcd\{m, p\} = \gcd\{m, 0\} = m = m_0$; if $k \neq 0$, we have $\gcd\{m, p\} = m_k = p_{k-1}$. ■

We shall not pursue the discussion of divisibility and the structure of $\mathbb{N}$ ordered by "divides" at this time. This is a major topic, including such matters as *prime numbers* (i.e., the minimal members of $\mathbb{N}^{\times}\backslash\{1\} = 2 + \mathbb{N}$ for that order) and prime factorization. It will be studied in a broader algebraic context.

Chapter 10

FINITE SETS

101. Finite sets and their cardinals

We recall (Section 33) that the sets S and T are said to be **equinumerous**, and that S is said to be **equinumerous to** T, if there is an invertible mapping from S to T or, equivalently, from T to S.

A set S is said to be **finite** if there exists an injection from S to $\mathbb{N}$ whose range has an upper bound or, equivalently, if there exists an injection from S to $n^\sqsubset$ for some $n \in \mathbb{N}$. A set is said to be **infinite** if it is not finite. A family is said to be **finite** or **infinite** according as its index set is finite or infinite.

If S is a finite set, the set $\{n \in \mathbb{N} \mid$ there is an injection from S to $n^\sqsubset\}$ is not empty; we define the **cardinal** (**number**) **of** S to be the minimum of this set, and denote it by $\#S$; thus

$$\#S := \min\{n \in \mathbb{N} \mid \text{there is an injection from } S \text{ to } n^\sqsubset\}.$$

(Other notations in use for the cardinal of S are $|S|$ and the unwieldy, obsolescent $\bar{\bar{S}}$.)

For every set S, we denote by $\mathfrak{F}(S)$ the collection of all finite subsets of S, i.e.,

$$\mathfrak{F}(S) := \{A \in \mathfrak{P}(S) \mid A \text{ is finite}\};$$

we set $\mathfrak{F}^\times(S) := \mathfrak{F}(S)\backslash\{\emptyset\}$, and for every $n \in \mathbb{N}$ we denote by $\mathfrak{F}_n(S)$ the collection of all finite subsets of S whose cardinal is n, i.e.,

$$\mathfrak{F}_n(S) := \{A \in \mathfrak{F}(S) \mid \#A = n\} \qquad \text{for all } n \in \mathbb{N}.$$

101A. Remark. A set S is finite and its cardinal is 0, 1, or 2 if and only if S is, respectively, the empty set, a singleton, or a doubleton. ■

101B. Theorem. *Let the finite set S be given. Every injection from S to $(\#S)^\sqsubset$ is invertible. The sets S and $(\#S)^\sqsubset$ are equinumerous.*

Proof. Let $n \in \mathbb{N}$ be given, and assume that there is an injection $\phi : S \to n^\sqsubset$ that is not surjective. Then $n^\sqsubset \neq \emptyset$, so that $n \in \mathbb{N}^\times$. Moreover, $n^\sqsubset\backslash\mathrm{Rng}\phi \neq \emptyset$, so we may set $p := \min(n^\sqsubset\backslash\mathrm{Rng}\phi)$, and define the mapping $\omega : n^\sqsubset \to n^\sqsubset$ by the rule

$$\omega(k) := \begin{cases} n-1 & \text{if } k = p \\ p & \text{if } k = n-1 \\ k & \text{if } k \in n^{\sqsubset} \backslash \{p, n-1\}. \end{cases}$$

Thus $\omega \circ \omega = 1_{n^{\sqsubset}}$, so that ω is invertible, and $\omega \circ \phi$ is injective. But $\mathrm{Rng}(\omega \circ \phi) \subset n^{\sqsubset} \backslash \{\omega(p)\} = n^{\sqsubset} \backslash \{n-1\} = (n-1)^{\sqsubset}$. Therefore $(\omega \circ \phi)|^{(n-1)^{\sqsubset}} : S \to (n-1)^{\sqsubset}$ is injective, and $\#S \leq n-1 < n$.

By contraposition, every injection from S to $(\#S)^{\sqsubset}$ must be surjective, and hence invertible. By the definition of $\#S$, such an injection exists; it follows that S and $(\#S)^{\sqsubset}$ are equinumerous. ■

101C. Remark. Theorem 101B shows that a set is finite if and only if it is equinumerous to $n^{\sqsubset}$ for some $n \in \mathbb{N}$. This property is often used to *define* the concept of finite set. See also Corollary 101J. ■

The following corollary justifies the term *equinumerous* ("having equal number"), at least for finite sets.

101D. Corollary. *Let the sets S and T be given. If S is finite, then S and T are equinumerous if and only if T is also finite and $\#S = \#T$.*

101E. Proposition. *Let the sets S and T be given, and assume that T is finite. Then there exists an injection from S to T if and only if S is finite and $\#S \leq \#T$. If $f\colon S \to T$ is an injection, then $\#S = \#T$ if and only if f is invertible.*

Proof. We choose an invertible mapping $\phi : T \to (\#T)^{\sqsubset}$, as we may by Theorem 101B. Let the injection $f\colon S \to T$ be given. Then $\phi \circ f\colon S \to (\#T)^{\sqsubset}$ is an injection. If such an injection f exists, it follows that S is finite and $\#S \leq \#T$. Moreover, if $\#S = \#T$, Theorem 101B shows that $\phi \circ f$ is invertible, and hence that f is invertible too. On the other hand, if f is invertible, then $\#S = \#T$ by Corollary 101D.

Assume, conversely, that S is finite and $\#S \leq \#T$, and choose an invertible mapping $\psi : S \to (\#S)^{\sqsubset}$, as we may by Theorem 101B. Then the mapping $\phi^{\leftarrow} \circ 1_{(\#S)^{\sqsubset} \subset (\#T)^{\sqsubset}} \circ \psi\colon S \to T$ is injective. ■

101F. Corollary. *Let the finite set T be given. Every subset S of T is finite and satisfies $\#S \leq \#T$; moreover, $\#S = \#T$ if and only if $S = T$.*

101G. Proposition. *Let the sets S and T and the surjection $f\colon S \to T$ be given. If S is finite, then T is finite and $\#S \geq \#T$. Moreover, $\#S = \#T$ if and only if f is invertible.*

Proof. Since f is surjective, we may choose a right-inverse $g : T \to S$ of f. (Here is a construction of g that avoids an appeal to the Axiom of Choice: choose a bijection $\psi\colon (\#S)^{\sqsubset} \to S$, as one may by Theorem 101B; and define g by the rule $g(y) := \psi(\min(f \circ \psi)^{<}(\{y\}))$ for all $y \in T$.) Then g is injective. Applying Proposition 101E to $T,\ S,\ g$ instead of $S,\ T,\ f$, we find that T is finite and that $\#S \geq \#T$; and also that $\#S = \#T$ holds if and only if g is invertible. But g is invertible if and only if its left-inverse f is invertible. ■

101H. Corollary (Pigeonhole Principle). *Let the equinumerous finite sets S*

and T and the mapping $f\colon S \to T$ be given. The following statements are equivalent:

(i): *f is injective.*

(ii): *f is surjective.*

(iii): *f is invertible.*

101I. Proposition. *Let the set S and the number $n \in \mathbb{N}$ be given. Then S and $n^\sqsubset$ are equinumerous if and only if S is finite and $n = \#S$.*

Proof. The "if" part follows from Theorem 101B. To prove the "only if" part, suppose that S and $n^\sqsubset$ are equinumerous. Then S is finite and $\#S \leq n$, so that $(\#S)^\sqsubset \subset n^\sqsubset$. Since S and $(\#S)^\sqsubset$ are equinumerous by Theorem 101B, we have $\#((\#S)^\sqsubset) = \#S = \#(n^\sqsubset)$. By Corollary 101F we must have $(\#S)^\sqsubset = n^\sqsubset$, and therefore $\#S = n$. ■

101J. Corollary. *Let the set S be given. The set $\{n \in \mathbb{N} \mid S \text{ and } n^\sqsubset \text{ are equinumerous}\}$ is empty if S is infinite and is the singleton $\{\#S\}$ if S is finite.*

101K. Corollary. *For every $n \in \mathbb{N}$, the sets $n^\sqsubset$ and $n^\sqsupset$ are finite, and $\#(n^\sqsubset) = \#(n^\sqsupset) = n$. The set $\mathbb{N}$ is infinite.*

Proof. The assertion concerning $n^\sqsubset$ follows from Proposition 101I with $S := n^\sqsubset$. Now the mapping $\mathrm{seq}\colon \mathbb{N} \to \mathbb{N}$ is injective and $n^\sqsupset := \mathrm{seq}_>(n^\sqsubset)$. Therefore $\mathrm{seq}|_{n^\sqsubset}^{n^\sqsupset}$ is bijective, and $n^\sqsubset$ and $n^\sqsupset$ are equinumerous. The assertion concerning $n^\sqsupset$ now follows from Corollary 101D. For every $n \in \mathbb{N}$, the set $\mathbb{N}$ includes the set $n^\sqsubset$, whose cardinal is n. By Proposition 101E, $\mathbb{N}$ is infinite. ■

101L. Remark. One often encounters a finite set being regarded routinely as the range of a list. This practice is usually unnecessary and indeed distracting; exceptionally, however, it becomes desirable. Now a finite set S is equinumerous to $(\#S)^\sqsubset$ and to $(\#S)^\sqsupset$ (Theorem 101B and Corollary 101K). For every bijection $\phi : (\#S)^\sqsubset \to S$ or $\phi : (\#S)^\sqsupset \to S$, the list $(\phi(m) \mid m \in (\#S)^\sqsubset)$ or $(\phi(m) \mid m \in (\#S)^\sqsupset)$, of length $\#S$, is called a **listing of** S. ■

101M. Proposition. *Let the set S and $s \in S$ be given. If $S\backslash\{s\}$ is finite, then S is finite and $\#S = \mathrm{seq}\#(S\backslash\{s\})$.*

Proof. By Theorem 101B we may choose a bijection $\phi : S\backslash\{s\} \to n^\sqsubset$, where $n := \#(S\backslash\{s\})$. We then define $\psi : S \to (\mathrm{seq}n)^\sqsubset$ by the rule

$$\psi(x) := \begin{cases} \phi(x) & \text{if } x \in S\backslash\{s\} \\ \\ n & \text{if } x = s. \end{cases}$$

Then ψ is bijective. By Proposition 101I, S is finite and $\#S = \mathrm{seq}n = \mathrm{seq}\#(S\backslash\{s\})$. ■

101N. Proposition. *Let the set S be given. The set $\{\#T \mid T \in \mathfrak{F}(S)\}$ is $(\mathrm{seq}\#S)^\sqsubset$ if S is finite and $\mathbb{N}$ if S is infinite.*

Proof. Suppose that S is finite. By Theorem 101B we may choose a bijection $\phi\colon S \to (\#S)^\sqsubset$. For every $n \in (\mathrm{seq}\#S)^\sqsubset$, the mapping $\phi|_{\phi^<(n^\sqsubset)}^{n^\sqsubset}$ is a bijection from $T := \phi^<(n^\sqsubset) \subset S$ to $n^\sqsubset$. By Proposition 101I we conclude that T is finite and that $\#T = n$. Since $n \in (\mathrm{seq}\#S)^\sqsubset$ was arbitrary, we have proved that $\{\#T \mid T \in \mathfrak{F}(S)\} \supset S \supset (\mathrm{seq}\#S)^\sqsubset$. The reverse inclusion follows from Corollary 101F.

Suppose now that S is infinite. Then $0 \in \{\#T \mid T \in \mathfrak{F}(S)\}$, since $\#\emptyset = 0$. If $n \in \{\#T \mid T \in \mathfrak{F}(S)\}$ is given, we may choose $A \in \mathfrak{F}(S)$ such that $\#A = n$. Since S is infinite, we have $S \backslash A \neq \emptyset$, and we may choose $s \in S \backslash A$. Then $A = (A \cup \{s\}) \backslash \{s\}$. By Proposition 101M, $A \cup \{s\} \in \mathfrak{F}(S)$ and

$$\mathrm{seq}n = \mathrm{seq}\#A = \#(A \cup \{s\}) \in \{\#T \mid T \in \mathfrak{F}(S)\}.$$

By (ℕIII) we conclude that $\{\#T \mid T \in \mathfrak{F}(S)\} = \mathbb{N}$. ■

102. Induction

102A. Theorem. *Every collection of finite sets, ordered by inclusion, is well-founded.*

Proof. Let $\mathcal{F}$ be a collection of finite sets, and let the non-empty subcollection $\mathcal{A}$ of $\mathcal{F}$ be given. Define $p := \min\{\#S \mid S \in \mathcal{A}\}$, and choose $A \in \mathcal{A}$ such that $\#A = p$. We claim that A is a minimal member of $\mathcal{A}$. If $S \in \mathcal{A}$ satisfies $S \subset A$, then Corollary 101F implies $\#S \leq \#A$; but the definition of p implies $\#A = p \leq \#S$. Therefore $\#S = \#A$, and by Corollary 101F we must have $S = A$. This establishes our claim. ■

102B. Theorem. *A set S is finite if and only if $\mathfrak{P}(S)$, ordered by inclusion, is well-founded.*

Proof. If S is finite, then Corollary 101F implies that $\mathfrak{P}(S) = \mathfrak{F}(S)$. By Theorem 102A, $\mathfrak{P}(S)$, ordered by inclusion, is then well-founded.

Assume, conversely, that $\mathfrak{P}(S)$, ordered by inclusion is well-founded. If T is an infinite subset of S, we have $T \neq \emptyset$ (by Remark 101A), and we may choose $t \in T$. By Proposition 101M, $T\backslash\{t\}$ is also infinite. The collection of all infinite subsets of S has therefore no minimal member, and must be empty. In particular, S itself is finite. ■

102C. Remark. The characterization of finite sets by Theorem 102B does not involve the natural numbers at all. It could be used to *define* the concept of finite sets; this has been done occasionally (notably by Wacław Sierpiński (1882-1969)), and has some foundational advantages. We shall take this matter up in Section 106. ■

Theorem 102A of course implies the availability of inductive-proof schemes and recursive definitions (Chapter 8) for collections of finite sets. It is convenient to describe more in detail the inductive-proof schemes for the collections of the form $\mathfrak{F}(S)$.

102D. Inductive-Proof Schemes. Let the set S be given. By Theorem 102A, the collection $\mathfrak{F}(S)$, ordered by inclusion, is well-founded. Since every subset of a finite set is finite (Corollary 101F), the collection of members of $\mathfrak{F}(S)$ that strictly precede a given $A \in \mathfrak{F}(S)$ is the collection of its proper subsets, i.e., $\mathrm{Spr}(A) = \mathfrak{P}(A)\backslash\{A\}$. If $P(\ \ )$ is a predicate describing a property that members of $\mathfrak{F}(S)$ may have, the standard scheme (Inductive-Proof Scheme 81B) for proving that $P(A)$ holds for all $A \in \mathfrak{F}(S)$ has the following *induction step*:

$$(\forall B \in \mathfrak{P}(A)\backslash\{A\}, \quad P(B)) \ \Rightarrow \ P(A) \qquad \textit{for all } A \in \mathfrak{F}(S). \tag{102.1}$$

In many cases, a stronger induction step (which yields a weaker proof scheme) is more convenient: *To prove that $P(A)$ holds for all $A \in \mathfrak{F}(S)$ it is enough to prove that the following assertions hold:*

$$P(\emptyset) \tag{102.2}$$

$$P(A\backslash\{a\}) \ \Rightarrow P(A) \qquad \textit{for all } A \in \mathfrak{F}^{\times}(S) \textit{ and all } a \in A. \tag{102.3}$$

Indeed, (102.2) and (102.3) together imply (102.1). When a distinction is necessary, a proof using the induction step (102.1) will be called a *proof by general induction*, and a proof using (102.2) and the *induction step* (102.3) will be called a *proof by special induction.* ■

102E. Inductive Proof Schemes. It is often desirable to establish a certain property for *all* finite sets by an inductive proof. It is not appropriate to attempt to do this by applying the Inductive-Proof Schemes 102D to the "set of all objects"; a different strategy for applying them is required. Suppose, then, that $P(\ \ )$ is a predicate describing a property that finite sets may have. We assert that *to prove that $P(S)$ holds for every finite set S it is enough to prove*

(102.4) *For every finite set A,* $\quad(\forall B \in \mathfrak{P}(A)\backslash\{A\}, \quad P(B)) \ \Rightarrow\ P(A).$

Indeed, suppose that (102.4) holds, and let the finite set S be given. Then (102.1) holds, and therefore $P(A)$ holds for all $A \in \mathfrak{F}(S)$; in particular, $P(S)$ holds. Proofs by this scheme are again called *proofs by general induction.* If one uses (102.2) and (102.3) instead of (102.1), one obtains the corresponding scheme for *proofs by special induction: To prove that $P(S)$ holds for every finite set S, it is enough to prove that the following assertions hold:*

(102.5) $\qquad P(\emptyset)$

(102.6) $\quad P(A\backslash\{a\}) \ \Rightarrow\ P(A)$ *for every non-empty finite set A and all $a \in A$.*

We record a simple and useful instance of proof by special induction.

102F. Proposition. *Let the finite set S and the non-empty nest $\mathcal{N}$ be given. Then $S \subset \bigcup\mathcal{N}$ if and only if $S \subset A$ for some $A \in \mathcal{N}$.*

Proof. The "if" part of the statement is trivial. We prove the "only if" part by special induction in $\mathfrak{F}(\bigcup\mathcal{N})$. If $S = \emptyset$, we may choose a member A of the non-empty collection $\mathcal{N}$ and find $S = \emptyset \subset A$.

Let $S \in \mathfrak{F}^{\times}(\bigcup\mathcal{N})$ and $s \in S$ be given, and assume that $S\backslash\{s\} \subset A$ for a suitable $A \in \mathcal{N}$. Since $s \in S \subset \bigcup\mathcal{N}$, we may choose $B \in \mathcal{N}$ such that $s \in B$. Then $S = (S\backslash\{s\}) \cup \{s\} \subset A \cup B$ and $A \cup B \in \mathcal{N}$, since $\mathcal{N}$ is a nest. This completes the induction step. ■

103. Operations with finite sets

In this section we show that certain sets constructed from finite sets are again finite, and that their cardinals can be determined or estimated from the cardinals of the given sets.

103A. Theorem. *Let the finite sets S and T be given. If S and T are disjoint, then $S \cup T$ is a finite set, and $\#(S \cup T) = \#S + \#T$.*

Proof. Let the finite set T be given. We prove by special induction that $P(S)$ holds for all finite sets S, where

$$P(S) \;:\Leftrightarrow\; (S \cap T = \emptyset \;\Rightarrow\; (S \cup T \text{ is finite and } \#(S \cup T) = \#S + \#T)).$$

Since $\emptyset \cup T = T$ and $\#\emptyset = 0$, we see that $P(\emptyset)$ holds.

Let the non-empty finite set S be given, and let $s \in S$ be such that $P(S\backslash\{s\})$ holds. If $S \cap T = \emptyset$, then $(S\backslash\{s\}) \cap T = \emptyset$ and $(S \cup T)\backslash\{s\} = (S\backslash\{s\}) \cup T$. Since $S\backslash\{s\}$ is finite by Corollary 101F, the induction hypothesis implies that $(S \cup T)\backslash\{s\}$ is finite and that $\#((S \cup T)\backslash\{s\}) = \#(S\backslash\{s\}) + \#T$. By Proposition 101M it follows that $S \cup T$ is finite and that $\#(S \cup T) = \#((S \cup T)\backslash\{s\}) + 1 = \#(S\backslash\{s\}) + 1 + \#T = \#S + \#T$. Therefore $P(S)$ holds. This completes the induction step, and $P(S)$ holds for all finite sets S.

Since T was an arbitrary finite set, the assertion is proved. ∎

103B. Corollary. *Let the finite sets S and T be given. Then $S \cup T$ and $S \cap T$ are finite sets, and*

$$\#(S \cup T) + \#(S \cap T) = \#S + \#T. \tag{103.1}$$

Proof. By Corollary 101F, $S \cap T$ and $S\backslash T$ are finite sets. By Theorem 103A, $S \cup T = (S\backslash T) \cup T$ is a finite set, and

$$\#S = \#(S\backslash T) + \#(S \cap T) \qquad \text{and} \qquad \#(S\backslash T) + \#T = \#(S \cup T).$$

Combining these equalities, we obtain (103.1). ∎

The results that follow have a quantitative aspect that depends on the concepts of the *sum* and the *product* of a finite family of natural numbers. These concepts will only be formally introduced in Section 112. Since they are, however, easily accessible to our intuition, we have preferred to anticipate their introduction and use them here, instead of deferring the quantitative parts of our results (as we might) and having to go over essentially the same proof again. We suppose it to be known that for each finite family of natural numbers $(n_i \mid i \in I)$ we are provided with natural numbers $\sum_{i \in I} n_i$ and $\prod_{i \in I} n_i$, and that these satisfy the following conditions (for definitions and proofs, see Section 112, especially Remark 112C,(b)):

$$\sum_{\imath \in I} n_i = 0 \;\text{ and }\; \prod_{i \in I} n_i = 1 \;\text{ if } I = \emptyset. \tag{103.2}$$

(103.3) $\sum_{i\in I} n_i = \sum_{i\in I\setminus\{j\}} n_i + n_j$ and $\prod_{i\in I} n_i = (\prod_{i\in I\setminus\{j\}} n_i)\cdot n_j$ if $I \neq \emptyset$ and $j \in I$.

103C. Lemma. *Let the finite set I and the natural number n be given. Then $\sum_{i\in I} n = n(\#I)$ and $\prod_{i\in I} n = n^{\#I}$.*

Proof. We prove by special induction that $\sum_{i\in J} n = n(\#J)$ and $\prod_{i\in J} n = n^{\#J}$ for all $J \in \mathfrak{F}(I)$, and hence in particular for $J := I$. By (103.2) we have $\sum_{i\in\emptyset} n = 0 = n\cdot 0 = n(\#\emptyset)$ and $\prod_{i\in\emptyset} n = 1 = n^0 = n^{\#\emptyset}$. Let $J \in \mathfrak{F}^\times(I)$ and $j \in J$ be given, and assume that $\sum_{i\in J\setminus\{j\}} n = n(\#(J\setminus\{j\}))$ and $\prod_{i\in J\setminus\{j\}} n = n^{\#(J\setminus\{j\})}$. Then (103.3) and Proposition 101M yield

$$\sum_{i\in J} n = \sum_{i\in J\setminus\{j\}} n + n = n(\#(J\setminus\{j\})) + n = n(\#J\setminus\{j\}) + 1) = n(\#J)$$

$$\prod_{i\in J} n = (\prod_{i\in J\setminus\{j\}} n)\cdot n = n^{\#(J\setminus\{j\})}\cdot n = n^{\#(J\setminus\{j\})+1} = n^{\#J}.$$

This completes the induction step. ■

103D. Theorem. *Let the finite family of finite sets $(T_i \mid i \in I)$ be given. Then $\dot\bigcup_{i\in I} T_i$ is a finite set and $\#(\dot\bigcup_{i\in I} T_i) = \sum_{i\in I} \#T_i$.*

Proof. We prove by special induction that

$$P(J) \;:\Leftrightarrow\; (\dot\bigcup_{i\in J} T_i \text{ is finite and } \#(\dot\bigcup_{i\in J} T_i) = \sum_{i\in J} \#T_i)$$

holds for all $J \in \mathfrak{P}(I) = \mathfrak{F}(I)$. We have $\dot\bigcup_{i\in\emptyset} T_i = \emptyset$ and $\sum_{i\in\emptyset} \#T_i = 0 = \#\emptyset$, so that $P(\emptyset)$ holds.

Let the non-empty subset J of I be given and let $j \in J$ be such that $P(J\setminus\{j\})$ holds. We observe that $\dot\bigcup_{i\in J\setminus\{j\}} T_i$ and $\dot\bigcup_{i\in\{j\}} T_i$ are disjoint and that their union is $\dot\bigcup_{i\in J} T_i$. We further observe that the mapping $(t \mapsto (j,t)) : T_j \to \dot\bigcup_{i\in\{j\}} T_i$ is bijective, so that, by Corollary 101D, $\dot\bigcup_{i\in\{j\}} T_i$ is finite and its cardinal is $\#T_j$. Applying the induction hypothesis and Theorem 103A with $S := \dot\bigcup_{i\in J\setminus\{j\}} T_i$ and $T := \dot\bigcup_{i\in\{j\}} T_i$, we

conclude that $\dot{\bigcup_{i\in J}} T_i$ is finite and that

$$\#(\dot{\bigcup_{i\in J}} T_i) = \#(\dot{\bigcup_{i\in J\setminus\{j\}}} T_i) + \#(\dot{\bigcup_{i\in\{j\}}} T_i) = \sum_{i\in J\setminus\{j\}} \#T_i + \#T_j = \sum_{i\in J} \#T_i.$$

We have shown that $P(J)$ holds. This completes the induction step.

Therefore $P(J)$ holds for all $J \in \mathfrak{P}(I)$. In particular, $P(I)$, which is the assertion of this theorem, also holds. ■

103E. Corollary. *Let the finite family of finite sets $(T_i \mid i \in I)$ be given. Then $\bigcup_{i\in I} T_i$ is a finite set, and $\#(\bigcup_{i\in I} T_i) \le \sum_{i\in I} \#T_i$; equality holds if and only if the family is disjoint. In particular, $\#S = \sum_{s\in S} 1$ for every finite set S.*

Proof. The mapping $((i,t) \mapsto t) : \dot{\bigcup_{i\in I}} T_i \to \bigcup_{i\in I} T_i$ is surjective, and it is bijective if and only if the family is disjoint. The assertion now follows from Theorem 103D and Proposition 101G. The particular case follows by setting $I := S$ and $T_s := \{s\}$ for every $s \in S$. ■

Remark. A more precise determination of $\#(\bigcup_{i\in I} T_i)$ is given by a generalization of Corollary 103B, known as the Inclusion-Exclusion Principle; it will be the subject of Section 116.

103F. Corollary. *The union of every finite collection of finite sets is a finite set.*

103G. Corollary. *Let the finite sets S and T be given. Then $S \times T$ is a finite set, and $\#(S \times T) = (\#S)(\#T)$.*

Proof. Apply Theorem 103D to $I := S$ and $T_s := T$ for all $s \in S$, and observe that $S \times T = \dot{\bigcup_{s\in S}} T$. Then $S \times T$ is a finite set and, using Lemma 103C,

$$\#(S \times T) = \#(\dot{\bigcup_{s\in S}} T) = \sum_{s\in S} \#T = (\#S)(\#T). \blacksquare$$

103H. Theorem. *Let the finite family of finite sets $(T_i \mid i \in I)$ be given. Then $\bigtimes_{i\in I} T_i$ is a finite set and $\#(\bigtimes_{i\in I} T_i) = \prod_{i\in I} \#T_i$.*

Proof. We prove by special induction that

$$P(J) \;:\Leftrightarrow\; (\bigtimes_{i\in J} T_i \text{ is finite and } \#(\bigtimes_{i\in J} T_i) = \prod_{i\in J} \#T_i)$$

holds for all $J \in \mathfrak{P}(I) = \mathfrak{F}(I)$. We have $\bigtimes_{i\in\emptyset} T_i = \{\emptyset\}$ and $\prod_{i\in\emptyset} \#T_i = 1 = \#\{\emptyset\}$, so that $P(\emptyset)$ holds.

Let the non-empty subset J of I be given and let $j \in J$ be such that $P(J\backslash\{j\})$ holds. By Proposition 44A, the sets $\bigtimes_{i\in J} T_i$ and $(\bigtimes_{i\in J\backslash\{j\}} T_i) \times T_j$ are equinumerous. By Corollaries 101D and 103G and the induction hypothesis, we conclude that $\bigtimes_{i\in J} T_i$ is finite and that

$$\#(\bigtimes_{i\in J} T_i) = (\#(\bigtimes_{i\in J\backslash\{j\}} T_i))(\#T_j) = (\prod_{i\in J\backslash\{j\}} \#T_i)(\#T_j) = \prod_{i\in J} \#T_i.$$

We have shown that $P(J)$ holds. This completes the induction step.

Therefore $P(J)$ holds for all $J \in \mathfrak{P}(I)$. In particular, $P(I)$, which is the assertion of this theorem, also holds. ■

103I. Corollary. *Let the finite sets S and T be given. Then the equinumerous sets T^S and $\mathrm{Map}(S,T)$ are finite, and $\#(T^S) = \#(\mathrm{Map}(S,T)) = (\#T)^{\#S}$.*

Proof. Apply Theorem 103H with $I := S$ and $T_s := T$ for all $s \in S$, and use Lemma 103C. ■

103J. Corollary. *A set S is finite if and only if $\mathfrak{P}(S)$ is finite. In that case, $\#\mathfrak{P}(S) = 2^{\#S}$.*

Proof. The mapping $(s \mapsto \{s\}) : S \to \mathfrak{P}(S)$ is injective. Therefore S is finite if $\mathfrak{P}(S)$ is finite (Proposition 101E).

The mapping $(A \mapsto \chi_{A\subset S}) : \mathfrak{P}(S) \to (2^{\sqsubset})^S$ is bijective: its inverse is $(\phi \mapsto \phi^<(\{1\})) : (2^{\sqsubset})^S \to \mathfrak{P}(S)$. If S is finite, Corollary 103I and the fact that $\#(2^{\sqsubset}) = 2$ (Remark 101A or Corollary 101K) imply that $\mathfrak{P}(S)$ is finite and $\#\mathfrak{P}(S) = \#((2^{\sqsubset})^S) = 2^{\#S}$. ■

103K. Remarks. (a): Let the family of sets $(T_i \mid i \in I)$ be given. From the preceding results and some trivial observations we obtain the following conclusions. $\dot{\bigcup}_{i\in I} T_i$ is finite if and only if the set $\{i \in I \mid T_i \neq \emptyset\}$ is finite and T_i is finite for every $i \in I$. $\bigtimes_{i\in I} T_i$ is finite if and •only if either $T_i = \emptyset$ for some $i \in I$, or else the set $\{i \in I \mid T_i$ is not a singleton$\}$ is finite and T_i is finite for every $i \in I$.

(b): Let the sets S and T be given. Then $S \cup T$ is finite if and only if S and T are both finite. $S \times T$ is finite if and only if $S = \emptyset$ or $T = \emptyset$ or S and T are both finite. The equinumerous sets T^S and $\mathrm{Map}(S,T)$ are finite if and only if $T = \emptyset$ or T is a singleton or $S = \emptyset$ or S and T are both finite. ■

The following important result is included in this section although it deals with sets that are not necessarily finite.

103L. Theorem (Principle of Finite Choice). *The Cartesian product of a finite family of non-empty sets is non-empty.*

Proof. The proof by special induction is entirely analogous to the proof of Theorem 103H. It uses the fact that if $S \neq \emptyset$ and $T \neq \emptyset$, then $S \times T \neq \emptyset$. ■

104. Factorials and binomial coefficients

In this section we continue the study of finite sets constructed from other finite sets, and the determination of their cardinals.

If S and T are finite sets, then $\mathrm{Inv}(S,T) := \{f \in \mathrm{Map}(S,T) \mid f \text{ is invertible}\} \subset \mathrm{Map}(S,T)$ is finite, by Corollary 103I. In particular, $\mathrm{Perm}(S) := \mathrm{Inv}(S,S)$, the set of all **permutations of** S, is finite for every finite set S; and $\mathrm{Perm}(S) \neq \emptyset$, since $1_S \in \mathrm{Perm}(S)$. On the other hand, the subset $\mathfrak{F}_m(S)$ of $\mathfrak{F}(S)$ is finite for every finite set S and every $m \in \mathbb{N}$, by Corollary 103J.

104A. Proposition. *Let the finite sets S and T be given.*

(a): *If $\#S \neq \#T$, then* $\mathrm{Inv}(S,T) = \emptyset$. *If $\#S = \#T$, then* $\#\mathrm{Perm}(S) = \#\mathrm{Inv}(S,T) = \#\mathrm{Perm}(T) \neq 0$.

(b): *If $\#S = \#T$, then $\#\mathfrak{F}_m(S) = \#\mathfrak{F}_m(T)$ for all $m \in \mathbb{N}$.*

Proof. The first assertion in (a) is the trivial part of Corollary 101D. We now assume that $\#S = \#T$, and may therefore choose an invertible mapping $h : S \to T$. Each of the mappings $(f \mapsto h^{\leftarrow} \circ f) : \mathrm{Inv}(S,T) \to \mathrm{Perm}(S)$ and $(g \mapsto h \circ g) : \mathrm{Perm}(S) \to \mathrm{Inv}(S,T)$ is the inverse of the other; the same is true of $(f \mapsto f \circ h^{\leftarrow}) : \mathrm{Inv}(S,T) \to \mathrm{Perm}(T)$ and $(g \mapsto g \circ h) : \mathrm{Perm}(T) \to \mathrm{Inv}(S,T)$. Thus $\mathrm{Perm}(S)$, $\mathrm{Inv}(S,T)$, and $\mathrm{Perm}(T)$ are equinumerous, and therefore have the same cardinal.

For every subset A of S and every subset B of T, the mappings $h|_A^{h_>(A)}$ and $h|_{h^<(B)}^B$ are bijective. Therefore $(h_>)_>(\mathfrak{F}_m(S)) = \mathfrak{F}_m(T)$, and $h_>|_{\mathfrak{F}_m(T)}^{\mathfrak{F}_m(S)}$ is bijective for each $m \in \mathbb{N}$. We conclude that $\mathfrak{F}_m(S)$ and $\mathfrak{F}_m(T)$ are equinumerous for each $m \in \mathbb{N}$.■

104B. Proposition. *Let the finite set S be given. Then:*

$$\#\mathfrak{F}_0(S) = 1 \qquad\qquad \#\mathfrak{F}_1(S) = \#S \tag{104.1}$$

$$\mathfrak{F}_m(S) = \emptyset \;\Leftrightarrow\; m > \#S \qquad \text{for all } m \in \mathbb{N} \tag{104.2}$$

$$m + n = \#S \;\Rightarrow\; \#\mathfrak{F}_m(S) = \#\mathfrak{F}_n(S) \qquad \text{for all } m, n \in \mathbb{N} \tag{104.3}$$

$$\sum_{m \in (\mathrm{seq}\#S)^{\sqsubset}} \#\mathfrak{F}_m(S) = 2^{\#S}. \tag{104.4}$$

Proof. $\mathfrak{F}_0(S) = \{\emptyset\}$, and the mapping $(x \mapsto \{x\}) : S \to \mathfrak{F}_1(S)$ is bijective. These facts imply that (104.1) holds. (104.2) is a rephrasing of part of Proposition 101N.

Let $m, n \in \mathbb{N}$ be given, and assume that $m + n = \#S$. Consider the complementation mapping $C_S : \mathfrak{P}(S) \to \mathfrak{P}(S)$. By Theorem 103A we have $\#A + \#C_S(A) = \#A + \#(S \backslash A) = \#S = m + n$ for all $A \in \mathfrak{P}(S)$; by the cancellation law it follows that $A \in \mathfrak{F}_m(S)$ if and only if $C_S(A) \in \mathfrak{F}_n(S)$. Since C_S is invertible, we conclude that $C_S|_{\mathfrak{F}_m(S)}^{\mathfrak{F}_n(S)}$ is invertible, and therefore $\mathfrak{F}_m(S)$ and $\mathfrak{F}_n(S)$ are equinumerous. This proves (104.3).

The list of collections $(\mathfrak{F}_m(S) \mid m \in (\mathrm{seq}\#S)^\sqsubset)$ is obviously disjoint, and by (104.2) we have $\bigcup_{m\in(\mathrm{seq}\#S)^\sqsubset} \mathfrak{F}_m(S) = \mathfrak{F}(S) = \mathfrak{P}(S)$. By Corollaries 103E and 103J we conclude that

$$\sum_{m\in(\mathrm{seq}\#S)^\sqsubset} \#\mathfrak{F}_m(S) = \#\Big(\bigcup_{m\in(\mathrm{seq}\#S)^\sqsubset} \mathfrak{F}_m(S)\Big) = \#\mathfrak{P}(S) = 2^{\#S}. \quad \blacksquare$$

104C. Theorem. *Let the finite set S and the subset A of S be given. Then*

$$\#\mathrm{Perm}(S) = (\#\mathfrak{F}_{\#A}(S))(\#\mathrm{Perm}(A))(\#\mathrm{Perm}(S\backslash A)). \tag{104.5}$$

Proof. For each $B \in \mathfrak{F}_{\#A}(S)$ we define the set of permutations $Q_B := \{\pi \in \mathrm{Perm}(S) \mid \pi_>(A) = B\}$. We observe that the family $(Q_B \mid B \in \mathfrak{F}_{\#A}(S))$ is disjoint. For every $\pi \in \mathrm{Perm}(S)$, the mapping $\pi|_A^{\pi_>(A)}$ is bijective, so that $\pi_>(A) \in \mathfrak{F}_{\#A}(S)$, and $\pi \in Q_{\pi_>(A)}$. Therefore $\bigcup_{B\in\mathfrak{F}_{\#A}(S)} Q_B = \mathrm{Perm}(S)$. It follows from Corollary 103E that

$$\#\mathrm{Perm}(S) = \sum_{B\in\mathfrak{F}_{\#A}(S)} \#Q_B. \tag{104.6}$$

Let $B \in \mathfrak{F}_{\#A}(S)$ be given. Then the mapping $(\pi \mapsto (\pi|_A^B,\ \pi|_{S\backslash A}^{S\backslash B})) : Q_B \to \mathrm{Inv}(A,B) \times \mathrm{Inv}(S\backslash A, S\backslash B)$ is obviously bijective. Since $\#A = \#B$, Theorem 103A shows that $\#(S\backslash A) = \#(S\backslash B)$. Therefore we find, using Corollary 103G and Proposition 104A, that

$$\#Q_B = (\#\mathrm{Inv}(A,B))(\#\mathrm{Inv}(S\backslash A, S\backslash B)) = (\#\mathrm{Perm}(A))(\#\mathrm{Perm}(S\backslash A)).$$

Here $B \in \mathfrak{F}_{\#A}(S)$ was arbitrary. Substituting this into (104.6) and applying Lemma 103C we obtain (104.5). $\blacksquare$

We now define the sequence $(n! \mid n \in \mathbb{N})$ in $\mathbb{N}^\times$ by the rule

$$n! := \#\mathrm{Perm}(n^\sqsubset) \qquad \text{for all } n \in \mathbb{N},$$

and the matrix $(\binom{n}{m} \mid (n,m) \in \mathbb{N}\times\mathbb{N})$ in $\mathbb{N}$ by the rule

$$\binom{n}{m} := \#\mathfrak{F}_m(n^\sqsubset) \qquad \text{for all } (n,m) \in \mathbb{N}\times\mathbb{N}.$$

The number $n!$ is called the **factorial of** n, and the numbers $\binom{n}{m}$ are called **binomial coefficients** (for reasons to be explained at a later time). By Propositions 104A and 101I we have

$$\#\mathrm{Perm}(S) = (\#S)! \qquad \text{for every finite set } S \tag{104.7}$$

$$\#\mathfrak{F}_m(S) = \binom{\#S}{m} \qquad \text{for every finite set } S \text{ and all } m \in \mathbb{N}. \tag{104.8}$$

104D. Proposition. *The sequence* $(n! \mid n \in \mathbb{N})$ *satisfies the following rules:*

$$0! = 1, \qquad \textit{and} \qquad (\mathrm{seq}n)! = n! \cdot \mathrm{seq}n \qquad \textit{for all } n \in \mathbb{N} \tag{104.9}$$

$$n! = \prod_{m \in n^\sqsupset} m \qquad \textit{for all } n \in \mathbb{N}. \tag{104.10}$$

Proof. We have $0! = \#\mathrm{Perm}(\emptyset) = \#\{1_\emptyset\} = 1$. For given $n \in \mathbb{N}$, we apply Theorem 104C with $S := (\mathrm{seq}n)^\sqsubset$ and $A := \{n\}$. We have $(\mathrm{seq}n)^\sqsubset \backslash \{n\} = n^\sqsubset$, $\mathfrak{F}_{\#\{n\}}((\mathrm{seq}n)^\sqsubset) = \#\mathfrak{F}_1((\mathrm{seq}n)^\sqsubset) = \#(\mathrm{seq}n)^\sqsubset = \mathrm{seq}n$ (by (104.1) and Corollary 101K), and $\#\mathrm{Perm}(\{n\}) = \#\{1_{\{n\}}\} = 1$. Consequently (104.5) yields

$$(\mathrm{seq}n)! = \#\mathrm{Perm}((\mathrm{seq}n)^\sqsubset) = (\mathrm{seq}n) \cdot 1 \cdot \#\mathrm{Perm}(n^\sqsubset) = (\mathrm{seq}n) \cdot n!.$$

This completes the proof of (104.9).

For each $n \in \mathbb{N}$ we have $(\mathrm{seq}n)^\sqsupset \backslash \{\mathrm{seq}n\} = n^\sqsupset$. From (103.2) and (103.3) we have

$$\prod_{m \in 0^\sqsupset} m = \prod_{m \in \emptyset} m = 1, \quad \text{and} \quad \prod_{m \in (\mathrm{seq}n)^\sqsupset} m = \Big(\prod_{m \in n^\sqsupset} m\Big) \cdot \mathrm{seq}n \qquad \text{for all } n \in \mathbb{N}.$$

Comparing this with (104.9) yields an inductive proof of (104.10). ∎

104E. Proposition. *The binomial coefficients satisfy the following rules:*

$$\binom{n}{m} = 0 \;\Leftrightarrow\; m > n \qquad \textit{for all } n, m \in \mathbb{N} \tag{104.11}$$

$$\binom{m+n}{m} = \binom{m+n}{n} \qquad \textit{for all } m, n \in \mathbb{N} \tag{104.12}$$

$$\sum_{m \in (\mathrm{seq}n)^\sqsubset} \binom{n}{m} = 2^n \qquad \textit{for all } n \in \mathbb{N} \tag{104.13}$$

$$\binom{m+n}{n} m!n! = (m+n)! \qquad \textit{for all } m, n \in \mathbb{N}. \tag{104.14}$$

Proof. (104.11), (104.12), (104.13) follow from (104.2), (104.3), (104.4), respectively, with a suitable choice of S in each case. To prove (104.14), we apply Theorem 104C with the choices $S := (m+n)^\sqsubset$, $A := m^\sqsubset$, using the observation that

$\#((m+n)^{\sqsubset}\backslash m^{\sqsubset}) = n$ by Theorem 103A to show that $\#\mathrm{Perm}((m+n)^{\sqsubset}\backslash m^{\sqsubset}) = n!$ with the help of (104.7). ■

Formula (104.14) shows that $m!n!$ divides $(m+n)!$ for all $m, n \in \mathbb{N}$ and we may write the familiar formula

$$\binom{m+n}{m} = \frac{(m+n)!}{m!n!} \qquad \text{for all } m, n \in \mathbb{N} \tag{104.15}$$

(from which we can again deduce (104.12)).

It is useful to have relations among binomial coefficients that permit their recursive calculation — just as (104.9) may be used to calculate factorials recursively. These relationships are given by the following proposition. In so far as they refer to the non-zero binomial coefficients, they were recorded in Europe as early as 1527 by Petrus Apianus (Peter Bienewitz, 1495-1552); the resulting triangular array is printed on the title-page of his *Arithmetica*. The introduction of this triangular array is occasionally attributed to Michael Stifel (1487-1567), to Niccolò Tartaglia (Niccolò Fontana, *c.* 1499-1557), or — quite erroneously — to Blaise Pascal (1623-1662) under the name "Pascal's Triangle". It should be noted, however, that the same triangular array appears in the block-printed treatise *Sìyuán yùjiàn* (*Precious Mirror of the Four Elements*), dated 1303, by Zhū Shìjié (*fl.* 1300), and can be traced back, by attribution, to Jiǎ Xiàn and Liú Rǔxié (*fl. c.* 1100), as published in the latter's (now lost) *Rújī shìsuǒ* (*Piling-up Powers and Unlocking Coefficients*).

104F. Theorem. *The binomial coefficients satisfy the following rules for all* $m, n \in \mathbb{N}$*:*

$$\binom{n}{0} = 1 \qquad\qquad \binom{0}{\mathrm{seq}m} = 0 \tag{104.16}$$

$$\binom{\mathrm{seq}n}{\mathrm{seq}m} = \binom{n}{m} + \binom{n}{\mathrm{seq}m}. \tag{104.17}$$

Proof. A proof can be obtained from (104.11), (104.15), and (104.9). We prefer to give a proof that illuminates the meaning of (104.17) directly in terms of the cardinal numbers that define the binomial coefficients. (104.16) is an immediate consequence of (104.1) and (104.2), since $0 < \mathrm{seq}m$ for all $m \in \mathbb{N}$.

Let $m, n \in \mathbb{N}$ be given. Define the collection $\mathcal{H} := \{A \in \mathfrak{F}_{\mathrm{seq}m}((\mathrm{seq}n)^{\sqsubset}) \mid n \in A\}$. Then each of the mappings $(A \mapsto A\backslash\{n\}) : \mathcal{H} \to \mathfrak{F}_m(n^{\sqsubset})$ and $(B \mapsto B \cup \{n\}) : \mathfrak{F}_m(n^{\sqsubset}) \to \mathcal{H}$ is the inverse of the other (here we have used Proposition 101M). Hence $\#\mathcal{H} = \mathfrak{F}_m(n^{\sqsubset})$. On the other hand,

$$\begin{aligned}\mathfrak{F}_{\mathrm{seq}m}((\mathrm{seq}n)^{\sqsubset})\backslash\mathcal{H} &= \{A \in \mathfrak{F}_{\mathrm{seq}m}((\mathrm{seq}n)^{\sqsubset}) \mid n \notin A\} = \\ &= \mathfrak{F}_{\mathrm{seq}m}((\mathrm{seq}n)^{\sqsubset}\backslash\{n\}) = \mathfrak{F}_{\mathrm{seq}m}(n^{\sqsubset}).\end{aligned}$$

By Theorem 103A, we have

$$\binom{\mathrm{seq}n}{\mathrm{seq}m} = \#\mathfrak{F}_{\mathrm{seq}m}((\mathrm{seq}n)^{\sqsubset}) = \#\mathcal{H} + \#(\mathfrak{F}_{\mathrm{seq}m}((\mathrm{seq}n)^{\sqsubset})\backslash\mathcal{H}) =$$

$$= \#\mathfrak{F}_m(n^{\sqsubset}) + \#\mathfrak{F}_{\mathrm{seq}m}(n^{\sqsubset}) = \binom{n}{m} + \binom{n}{\mathrm{seq}m}. \blacksquare$$

▼ As an application of the results of this section, we determine, for given finite sets S and T, the cardinal of the finite set $\mathrm{Inj}(S,T) := \{f \in \mathrm{Map}(S,T) \mid f \text{ is injective}\} \subset \mathrm{Map}(S,T)$.

104G. Proposition. *Let the finite sets S and T be given. If $\#S > \#T$, then $\mathrm{Inj}(S,T) = \emptyset$. If $\#S \leq \#T$, then $\#\mathrm{Inj}(S,T) = (\#T)!/(\#T - \#S)!$.*

Proof. The first part of the conclusion follows from Proposition 101E. Assume that $\#S \leq \#T$. For each $B \in \mathfrak{F}_{\#S}(T)$ we define the set $Q_B := \{f \in \mathrm{Inj}(S,T) \mid \mathrm{Rng}f = B\}$. Obviously, the family $(Q_B \mid B \in \mathfrak{F}_{\#S}(T))$ is disjoint. For every $f \in \mathrm{Inj}(S,T)$, the mapping $f|^{\mathrm{Rng}}$ is bijective, and therefore $\#\mathrm{Rng}f = \#S$ and $f \in Q_{\mathrm{Rng}f}$. Therefore $\bigcup_{B\in\mathfrak{F}_{\#S}(T)} Q_B = \mathrm{Inj}(S,T)$. By Corollary 103E, we have

$$\#\mathrm{Inj}(S,T) = \sum_{B\in\mathfrak{F}_{\#S}(T)} \#Q_B. \tag{104.18}$$

For each $B \in \mathfrak{F}_{\#S}(T)$, the mapping $(f \mapsto f|^B) : Q_B \to \mathrm{Inv}(S,B)$ is bijective, so that, by Proposition 104A,(a),

$$\#Q_B = \#\mathrm{Inv}(S,B) = \#\mathrm{Perm}(S).$$

Substituting this into (104.18) and using Lemma 103C, (104.7), (104,8), and (104.15), we find

▲ $$\mathrm{Inj}(S,T) = (\#\mathfrak{F}_{\#S}(T))(\#\mathrm{Perm}(S)) = \left(\begin{array}{c}\#T\\ \#S\end{array}\right)(\#S)! = (\#T)!/(\#T - \#S)!. \blacksquare$$

105. Orders in finite sets

105A. Proposition. *Every finite ordered set is well-founded.*

Proof. By Proposition 62E, every finite ordered set is order-isomorphic to a (finite) collection of finite sets ordered by inclusion. By Theorem 102A, the latter is well-founded; hence so is the former. ■

105B. Corollary. *In every finite totally ordered set, every non-empty subset has both a maximum and a minimum.*

Proof. By Proposition 105A, every finite totally ordered set is well-ordered (Proposition 64A); since the reverse order is also total, the set is also well-ordered by it. The assertion follows at once. ■

105C. Corollary. *Let the non-empty finite nest $\mathcal{N}$ be given. Then $\bigcup\mathcal{N} \in \mathcal{N}$ and $\bigcap\mathcal{N} \in \mathcal{N}$.*

105D. Corollary. *Let the subset S of $\mathbb{N}$ be given. The following statements are equivalent:*

(i): *S has an upper bound.*

(ii): *S is finite.*

(iii): *S is empty or S has a maximum.*

105E. Theorem. *Let the finite totally ordered set S be given. Then there is exactly one order-isomorphism from S to $(\#S)^\sqsubset$.*

Proof. We prove by general induction that

$$P(S) \quad :\Leftrightarrow \quad \begin{array}{l}\text{(For every total order } \prec \text{ in } S \text{ there is exactly one order-isomorphism}\\ \text{from } S \text{ ordered by } \prec \text{ to } (\#S)^\sqsubset \text{ ordered by } \leq)\end{array}$$

holds for all finite sets S.

Let the finite set S be given, and assume that $P(T)$ holds for all proper subsets T of S. Let S be totally ordered by $\prec$. If $S = \emptyset$, we have $(\#S)^\sqsubset = \emptyset$, and there is exactly one mapping from S to $(\#S)^\sqsubset$, namely $1_\emptyset$; and this mapping is trivially an order-isomorphism.

Assume, then, that $S \neq \emptyset$. By Corollary 105B, we may set $s := \max S$. By the induction hypothesis applied to the proper subset $S\backslash\{s\}$ of S, there is exactly one order-isomorphism ψ from $S\backslash\{s\}$ ordered by $\prec$ to $(\#(S\backslash\{s\}))^\sqsubset$ ordered by $\leq$. We recall that, by Proposition 101M, $\#S = \mathrm{seq}\#(S\backslash\{s\})$.

Suppose that $\phi : S \to (\#S)^\sqsubset$ is an order-isomorphism. Then $\phi(s) = \phi(\max S) = \max\phi_>(S) = \max(\#S)^\sqsubset = \max(\mathrm{seq}\#(S\backslash\{s\}))^\sqsubset = \#(S\backslash\{s\})$; therefore $\phi_>(S\backslash\{s\}) = (\mathrm{seq}\#(S\backslash\{s\}))^\sqsubset\backslash\{\#(S\backslash\{s\})\} = (\#(S\backslash\{s\}))^\sqsubset$. We conclude that $\phi|_{S\backslash\{s\}}^{(\#(S\backslash\{s\}))^\sqsubset}$ is an order-isomorphism, and must therefore be ψ. Hence ϕ must satisfy

$$\phi(x) = \begin{cases} \psi(x) & \text{if } x \in S\backslash\{s\} \\ \\ \#(S\backslash\{s\}) & \text{if } x = s. \end{cases} \tag{105.1}$$

Conversely, if $\phi : S \to (\#S)^\sqsubset$ is *defined* by (105.1), it follows easily from the

definitions of ψ and s that ϕ is an order-isomorphism. We have shown that $P(S)$ holds, both when S is empty and when S is not empty. This completes the induction step. ■

105F. Corollary. *Let the finite set S totally ordered by $\prec$ be given. A relation ρ in S is a total order if and only if there is a permutation π of S such that*

$$\forall x, y \in S, \quad x \prec y \;\Leftrightarrow\; \pi(x)\ \rho\ \pi(y). \tag{105.2}$$

Proof. The proof of the "if" part is a straightforward verification, and would remain valid even if the assumption that S is finite were removed. To prove the "only if" part, assume that ρ is a total order. By Theorem 105E we may choose order-isomorphisms ϕ from S ordered by $\prec$ to $(\#S)^\sqsubset$ ordered by $\leq$, and ψ from S ordered by ρ to $(\#S)^\sqsubset$ ordered by $\leq$. Then $\pi := \psi^{\leftarrow} \circ \phi$ is an order-isomorphism from S ordered by $\prec$ to S ordered by ρ, so that π is a permutation of S and satisfies (105.2). ■

105G. Theorem. *A set S is finite if and only if it can be totally ordered so that every non-empty subset of S has both a maximum and a minimum.*

Proof. Proof of the "only if" part. Let the finite set S be given. By Theorem 101B we may choose a bijection $\phi\colon S \to (\#S)^\sqsubset$. We then define the relation ρ in S by the rule

$$\forall x, y \in S, \quad x\ \rho\ y \;:\Leftrightarrow\; \phi(x) \leq \phi(y).$$

It is clear that ρ is a total order in S. The "only if" part of the assertion then follows from Corollary 105B.

Proof of the "if" part. Assume that the set S is totally ordered by $\prec$ and that every non-empty subset of S has both a maximum and a minimum. Consider the subset $U := \{x \in S \mid \mathrm{Ub}(\{x\}) \text{ is infinite}\}$ of S. Let $x \in U$ be given. Now $\mathrm{Ub}(\{x\})$ is infinite, hence $\mathrm{Ub}(\{x\})\backslash\{x\}$ is also infinite (Proposition 101M), and therefore not empty. This set then has a minimum, say y, and $\mathrm{Ub}(\{x\})\backslash\{x\} = \mathrm{Ub}(\{y\})$. Thus $x \precneqq y$ and $\mathrm{Ub}(\{y\})$ is infinite. Therefore $y \in U$ and x is not the maximum of U. Since $x \in U$ was arbitrary, we conclude that U has no maximum and is therefore empty. This means that $\mathrm{Ub}(\{x\})$ is finite for all $x \in S$.

If $S = \emptyset$, then S is finite. If $S \neq \emptyset$, then $S = \mathrm{Ub}(\{\min S\})$ is finite. ■

105H. Remark. Theorem 105G provides another characterization of finite sets that does not involve the natural numbers. It also has been used to *define* the concept of finite set. ■

▼ We end this section by examining the notion of a family *chosen recursively* (•Theorem 82H) when the well-founded index set is finite. It is found, unsurprisingly, that in this case no appeal to the •Axiom of Choice is needed, but since some of the sets involved may be infinite the proof is not an immediate consequence of the Principle of Finite Choice (Theorem 103L).

105I. Theorem. *Let the finite ordered set $(I; \prec)$ and the family of sets $(A_i \mid i \in I)$ be given. Let a family of set-valued mappings $(\Phi_i \mid i \in I)$ also be given, with $\Phi_i \in \mathrm{Map}(\bigtimes_{j \in \mathrm{Spr}(i)} A_j, \mathfrak{P}(A_i))$ for all $i \in I$. Assume that*

$$\forall j \in \mathrm{Spr}(i), \quad u_j \in \Phi_j(u|_{\mathrm{Spr}(j)}) \;\Rightarrow\; \Phi_i(u) \neq \emptyset$$

(105.3)

$$\text{for all } i \in I \text{ and all } u \in \bigtimes_{j\in\mathrm{Spr}(i)} A_j.$$

Then there exists $a \in \bigtimes_{i\in I} A_i$ *such that*

(105.4) $$a_i \in \Phi_i(a|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in I.$$

Proof. We shall prove by general induction (Inductive Proof Schemes 102E) that $P(I)$ holds for every finite set I, where

$$P(I) \;:\Leftrightarrow\; \text{For every } \prec \in \mathrm{Ord}(I), \text{the assertion of the present theorem is valid for the ordered set } (I;\prec).$$

Let the finite set I be given. Let $\prec \in \mathrm{Ord}(I)$ and the families $(A_i \mid i \in I)$ and $(\Phi_i \mid i \in I)$ be given as in the statement, and assume that (105.3) holds.

If $I = \emptyset$, there trivially exists a family $a \in \bigtimes_{i\in I} A_i = \{\emptyset\}$ (self-indexed), and (105.4) holds vacuously.

Assume then that $I \neq \emptyset$. By Proposition 105A applied to $(I;\succ)$ we may choose a maximal member m of $(I;\prec)$. We set $I' := I\backslash\{m\}$ and $\prec' := \prec|_{I'}$. For every $i \in I'$ the set of strict lower bounds of $\{i\}$ in $(I';\prec')$ is obviously the same as in $(I;\prec)$, namely $\mathrm{Spr}(i)$. The families $(A_i \mid i \in I')$ and $(\Phi_i \mid i \in I')$ therefore satisfy (105.3) with $(I;\prec)$ replaced by $(I';\prec')$.

By the induction hypothesis, $P(I')$ holds, and we may therefore choose $b \in \bigtimes_{i\in I'} A_i$ such that

(105.5) $$b_i \in \Phi_i(b|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in I'.$$

Since $\mathrm{Spr}(m) \subset I'$, this implies

$$b_i \in \Phi_i(b|_{\mathrm{Spr}(i)}) \qquad \text{for all } i \in \mathrm{Spr}(m);$$

together with (105.3) this yields $\Phi_m(b|_{\mathrm{Spr}(m)}) \neq \emptyset$. We may therefore choose

(105.6) $$z \in \Phi_m(b|_{\mathrm{Spr}(m)}).$$

We now define $a \in \bigtimes_{i\in I} A_i$ by

$$a_i := \begin{cases} b_i & \text{if } i \in I' = I\backslash\{m\} \\ z & \text{if } i = m. \end{cases}$$

By (105.5) and (105.6) it follows that a satisfies (105.4). This completes the induction step. ■

105J. Remark. Theorem 103L describes the special case of Theorem 105I in which for every $i \in I$ the mapping Φ_i is the constant mapping whose single value is the non-empty set A_i itself. ■

106. Finiteness without counting

In Theorem 102B it was shown that a set S is finite if and only if $\mathfrak{P}(S)$, ordered by inclusion, is well-founded; and in Remark 102C it was noted that this characterization, which contains no reference to natural numbers or to any counting system, could be used — and had indeed be used — to *define* the concept of a finite set. Inasmuch as the existence of a counting system (and hence of the Natural-Number System) rests on foundational agreements such as an Axiom of Infinity (see Section 95), such an alternative definition of finiteness would seem to have merit. It is contended that every result about finite sets that does not explicitly involve a counting system (such as one involving the cardinals of finite sets) can be derived from this alternative definition.

In this section we intend to illustrate this contention by deriving a result describing the "size-comparability" of finite sets, as well as the Pigeonhole Principle (Corollary 101H).

A second purpose of this section is to show that the existence of a counting system — which would then lead to identifying the alternative definition, via Theorem 102B, with that given in Section 101 — is actually *required* if the following very modest version of the Axiom of Infinity is accepted: *There exists a set that is not finite according to the alternative definition.*

In this section alone, therefore, a set S is said to be **finite** if $\mathfrak{P}(S)$, ordered by inclusion, is well-founded. A set that is not finite is said to be **infinite**. For every set S we define $\mathfrak{F}(S) := \{A \in \mathfrak{P}(S) \mid A \text{ is finite}\}$.

106A. Proposition (cf. Theorem 102A). *Every collection of finite sets, ordered by inclusion, is well-founded.*

Proof. Let $\mathcal{F}$ be a collection of finite sets, and let the non-empty subcollection $\mathcal{A}$ of $\mathcal{F}$ be given. Choose $S \in \mathcal{A}$. Since S is finite, we may choose a minimal member A of the non-empty subcollection $\mathcal{A} \cap \mathfrak{P}(S)$ of $\mathfrak{P}(S)$, ordered by inclusion. Then A is a minimal member of $\mathcal{A}$. ■

106B. Remark. It follows from Proposition 106A that inductive-proof schemes and recursive definitions (Chapter 8) are available for collections of finite sets. In particular, the Inductive-Proof Schemes 102D for collections of the form $\mathfrak{F}(S)$ may and shall be appropriated for the context of this section. ■

106C. Proposition (cf. Proposition 101E). *Let the sets A and B be given. If B is finite and there exists an injection from A to B, then A is finite. In particular, every subset of a finite set is finite.*

Proof. Choose an injection $f\colon A \to B$. Then $f_> : \mathfrak{P}(A) \to \mathfrak{P}(B)$ is injective and isotone with respect to inclusion (Propositions 34A.L and 23A). The conclusion follows from the definition of *finite* in this section. ■

106D. Theorem. *Let the finite set S and the set T be given. Then either there exists an injection from S to T, or T is finite and there exists an injection from T to S.*

Proof. Set $\mathcal{C} := \{A \in \mathfrak{P}(S) \mid \text{there exists an injection from } S\backslash A \text{ to } T\}$. Now $S \in \mathcal{C}$,

since $1_{\emptyset\subset T} : S\backslash S \to T$ is injective; hence $\mathcal{C} \neq \emptyset$. Since S is finite, we may choose a minimal member M of $\mathcal{C}$, and we may further choose an injective $f\colon S\backslash M \to T$. We distinguish two cases.

Case 1. $M = \emptyset$. Then f is an injection from S to T.

Case 2. $M \neq \emptyset$. We claim that $\mathrm{Rng}f = T$. Suppose not; we may then choose $u \in M$ and $v \in T\backslash\mathrm{Rng}f$, and define $g\colon (S\backslash M) \cup \{u\} \to T$ by

$$g(x) := \begin{cases} f(x) & \text{if } x \in S\backslash M \\ \\ v & \text{if } x = u. \end{cases}$$

Then g is obviously an injection from $S\backslash(M\backslash\{u\})$ to T; hence $M\backslash\{u\} \in \mathcal{C}$, contradicting the minimality of M. This establishes our claim; it follows that f is bijective. By Proposition 106C applied to the injection $(f^{\leftarrow})|^S : T \to S$ it follows that T is finite in this case. ■

106E. Corollary. *Let the finite sets S and T be given. Then exactly one of the following statements is valid:*

(a): *There exists an injection from S to T, but none from T to S.*

(b): *There exists an injection from T to S, but none from S to T.*

(c): *S and T are equinumerous.*

Proof. This is an immediate consequence of Theorem 106D and the Schröder-Bernstein Theorem (Theorem 75C). ■

Remark. For an analogous assertion for infinite sets see Section 175.

106F. Theorem. *Let the mapping f be given, and assume that* $\mathrm{Dom}f$ *is finite and that* $\mathrm{Dom}f \subset \mathrm{Rng}f$. *Then f is injective and* $\mathrm{Dom}f = \mathrm{Rng}f$.

Proof. We shall prove by special induction that $P(A)$ holds for all $A \in \mathfrak{F}(\mathrm{Dom}f)$, where

$$P(A) :\Leftrightarrow \text{ Every } g \in \mathrm{Map}(A, \mathrm{Cod}f) \text{ such that } A \subset \mathrm{Rng}g \text{ is injective and satisfies } \mathrm{Rng}g = A.$$

Once this is proved, $P(\mathrm{Dom}f)$ applied to f will yield the conclusion. $P(\emptyset)$ is valid, since $1_{\emptyset\subset\mathrm{Cod}f}$ is the only member of $\mathrm{Map}(\emptyset, \mathrm{Cod}f)$, and this mapping is injective and its range is $\emptyset$.

Let $A \in \mathfrak{F}^{\times}(\mathrm{Dom}f)$ and $a \in A$ be given, and assume that $P(A\backslash\{a\})$ holds.

Let $g \in \mathrm{Map}(A, \mathrm{Cod}f)$ be given, and assume that $A \subset \mathrm{Rng}g$. Define $\sigma \in \mathrm{Perm}(\mathrm{Cod}f)$ by the rule

$$\sigma(x) := \begin{cases} g(a) & \text{if } x = a \\ \\ a & \text{if } x = g(a) \\ \\ x & \text{if } x \in \mathrm{Cod}f\backslash\{a, g(a)\}, \end{cases}$$

and set $h := \sigma \circ g \in \mathrm{Map}(A, \mathrm{Cod}f)$. Then $\mathrm{Rng}h = \mathrm{Rng}g$ and $h(a) = a$. Therefore

$$\mathrm{Rng}(h|_{A\backslash\{a\}}) = h_{>}(A\backslash\{a\}) \supset \mathrm{Rng}h\backslash\{h(a)\} = \mathrm{Rng}g\backslash\{a\} \supset A\backslash\{a\} = \mathrm{Dom}(h|_{A\backslash\{a\}}).$$

By the induction hypothesis $P(A\backslash\{a\})$ we conclude that $h|_{A\backslash\{a\}}$ is injective and that $h_>(A\backslash\{a\}) = \mathrm{Rng}(h|_{A\backslash\{a\}}) = A\backslash\{a\}$. Using the fact that $h(a) = a$, we conclude that h is injective and that $\mathrm{Rng}h = A$. Then $g = \sigma \circ h$ is also injective, and $\mathrm{Rng}g = \mathrm{Rng}h = A$. This completes the induction step. ■

106G. Corollary. *Let the finite set S and the mapping $f\colon S \to S$ be given. The following statements are equivalent:*

(i): *f is injective.*

(ii): *f is surjective.*

(iii): *f is bijective.*

Proof. Obviously, (iii) implies (ii). It is an immediate consequence of Theorem 106F that (ii) implies (i).

To prove that (i) implies (iii), assume that f is injective. By Proposition 33E.L, we may choose a left-inverse $g\colon S \to S$ of f; then g is surjective. By Theorem 106F, g is also injective, hence invertible. But f is a right-inverse of g; by Proposition 33B we have $f = g^{\leftarrow}$, and hence f is also invertible, and therefore bijective. ■

106H. Corollary. (Pigeonhole Principle) (cf. Corollary 101H). *Let the equinumerous finite sets S and T and the mapping $f\colon S \to T$ be given. The following statements are equivalent:*

(i): *f is injective.*

(ii): *f is surjective.*

(iii): *f is invertible.*

Proof. Choose a bijection $h\colon T \to S$ and apply Corollary 106G to $h \circ f\colon S \to S$ instead of f. ■

The remainder of this section is devoted to the second purpose mentioned before, viz., to show how to construct a counting system from a given *infinite* set S. To help the reader understand the idea behind this construction — at the cost of some bending of the logic — we suggest that, if the Natural-Number System *were* available, we would identify each $n \in \mathbb{N}$ with the subcollection $\mathfrak{F}_n(S)$ (as defined in Section 101) of $\mathfrak{P}(S)$.

We begin with a simple observation about sets that are not necessarily finite.

106I. Lemma. *Let the non-empty sets A and B, and $a \in A$ and $b \in B$, be given. Then A and B are equinumerous if and only if $A\backslash\{a\}$ and $B\backslash\{b\}$ are equinumerous.*

106J. Proposition (cf. Proposition 101M). *Let the non-empty set A and $a \in A$ be given. If $A\backslash\{a\}$ is finite, then A is finite.*

Proof. Assume that $A\backslash\{a\}$ is finite. Let the non-empty collection $\mathcal{C}$ of subsets of A be given. We must show that $\mathcal{C}$ has a minimal member with respect to inclusion.

Now $\mathcal{C}' := \{B\backslash\{a\} \mid B \in \mathcal{C}\}$ is a non-empty collection of subsets of the finite set $A\backslash\{a\}$. We may therefore choose $K \in \mathcal{C}$ such that $K\backslash\{a\}$ is a minimal member of $\mathcal{C}'$.

Let $B \in \mathcal{C}$ be given, and assume that $B \subset K$. Then $B\backslash\{a\} \in \mathcal{C}'$ and $B\backslash\{a\} \subset K\backslash\{a\}$. Therefore $B\backslash\{a\} = K\backslash\{a\}$, and consequently $K\backslash\{a\} \subset B \subset K$, so that $B = K\backslash\{a\}$ or $B = K$. We conclude that either $K\backslash\{a\}$ or K is a minimal member of $\mathcal{C}$, according as $K\backslash\{a\}$ is a member of $\mathcal{C}$ or not. ■

From now on, we shall assume that an infinite set S has been given.

In $\mathfrak{F}(S)$ the relation

$$\forall A, B \in \mathfrak{F}(S), \quad A \sim B \quad :\Leftrightarrow \ A \text{ and } B \text{ are equinumerous}$$

is an equivalence relation. We denote the corresponding partition of $\mathfrak{F}(S)$ by N.

We next define the mapping $\sigma\colon \mathfrak{P}(\mathfrak{P}(S)) \to \mathfrak{P}(\mathfrak{P}(S))$ by

$$\sigma(\mathcal{C}) := \{A \in \mathfrak{P}(S) \mid \exists a \in A, A\backslash\{a\} \in \mathcal{C}\} \qquad \text{for all } \mathcal{C} \in \mathfrak{P}(\mathfrak{P}(S)).$$

106K. Lemma. (a): *σ is isotone with respect to inclusion in $\mathfrak{P}(\mathfrak{P}(S))$.*

(b): $\sigma_>(\mathfrak{P}(\mathfrak{F}(S))) \subset \mathfrak{P}(\mathfrak{F}(S))$.

(c): *For all $\mathcal{C} \in \mathfrak{P}(\mathfrak{F}(S))$, $\sigma(\mathcal{C}) \neq \emptyset$ if (and only if) $\mathcal{C} \neq \emptyset$.*

(d): $\sigma_>(\mathrm{N}) \subset \mathrm{N}$.

Proof of (a). This is trivial.

Proof of (b). Let $\mathcal{C} \in \mathfrak{P}(\mathfrak{F}(S))$ be given. Let $A \in \sigma(\mathcal{C})$ be given. We may choose $a \in A$ such that $A\backslash\{a\} \in \mathcal{C}$. Then $A\backslash\{a\}$ is finite. By Lemma 106J, A is also finite. Since $A \in \sigma(\mathcal{C})$ was arbitrary, we have $\sigma(\mathcal{C}) \in \mathfrak{P}(\mathfrak{F}(S))$.

Proof of (c). Let $\mathcal{C} \in \mathfrak{P}(\mathfrak{F}(S))$ be given, and assume that $\mathcal{C} \neq \emptyset$. Choose $F \in \mathcal{C}$. Since F is finite and S *is infinite*, we may choose $a \in S\backslash F$. Then $A := F \cup \{a\}$ satisfies $A\backslash\{a\} = F \in \mathcal{C}$, and thus $A \in \sigma(\mathcal{C})$. We conclude that $\sigma(\mathcal{C}) \neq \emptyset$.

Proof of (d). Let $\mathcal{N} \in \mathrm{N}$ be given. We must show that $\sigma(\mathcal{N}) \in \mathrm{N}$. By Parts (a), (b), (c) we have $\sigma(\mathcal{N}) \in \mathfrak{P}(\mathfrak{F}(S))$ and $\sigma(\mathcal{N}) \neq \emptyset$. We may choose $A \in \sigma(\mathcal{N})$ and $a \in A$ such that $A\backslash\{a\} \in \mathcal{N}$. It remains to show that

$$\forall B \in \mathfrak{F}(S), \quad B \sim A \ \Leftrightarrow \ B \in \sigma(\mathcal{N}).$$

Let $B \in \mathfrak{F}(S)$ be given. Assume first that $B \sim A$; then $B \neq \emptyset$, and we may choose $b \in B$. By Lemma 106I we have $B\backslash\{b\} \sim A\backslash\{a\}$, and therefore $B\backslash\{b\} \in \mathcal{N}$. It follows that $B \in \sigma(\mathcal{N})$.

Assume conversely, that $B \in \sigma(\mathcal{N})$. We may choose $b \in B$ such that $B\backslash\{b\} \in \mathcal{N}$. Then $B\backslash\{b\} \sim A\backslash\{a\}$, and by Lemma 106I we conclude that $B \sim A$. ■

By virtue of Lemma 106K,(d), we may define the mapping

$$\mathrm{Seq} := \sigma|^{\mathrm{N}}_{\mathrm{N}} : \mathrm{N} \to \mathrm{N}.$$

We also note that $\emptyset$ is finite and that $\{\emptyset\} \in \mathrm{N}$.

106L. Theorem. *The set* N*, endowed with structure by the prescription of* $\{\emptyset\}$ *as zero and* Seq *as successor-mapping, is a counting system.*

Proof. 1. Let $\mathcal{N} \in \mathrm{N}$ be given. By the definitions of σ and Seq, we have $\emptyset \notin \sigma(\mathcal{N}) = \mathrm{Seq}\mathcal{N}$. Therefore $\{\emptyset\} \notin \mathrm{Rng\ Seq}$. Thus (Count I) is satisfied.

2. Let $\mathcal{N}, \mathcal{N}' \in \mathrm{N}$ be given and assume that $\mathrm{Seq}\mathcal{N} = \mathrm{Seq}\mathcal{N}'$. By Lemma 106K,(c) and the definition of Seq, we may choose $A \in \mathrm{Seq}\mathcal{N} = \mathrm{Seq}\mathcal{N}'$ and $a, a' \in A$ such that $A\backslash\{a\} \in \mathcal{N}$ and $A\backslash\{a'\} \in \mathcal{N}'$. By Lemma 106I, we have $A\backslash\{a\} \sim A\backslash\{a'\}$, and

consequently $\mathcal{N} \cap \mathcal{N}' \neq \emptyset$, whence $\mathcal{N} = \mathcal{N}'$. We have proved that Seq is injective; thus (Count II) is satisfied.

3. Let the subset Σ of N be given, and assume that $\{\emptyset\} \in \Sigma$ and $\mathrm{Seq}_{>}(\Sigma) \subset \Sigma$. We claim that $\Sigma = \mathrm{N}$, thus establishing (Count III), and completing the proof.

We shall prove by special induction that $P(A)$ holds for all $A \in \mathfrak{F}(S)$, where

$$P(A) \;:\Leftrightarrow\; \exists \mathcal{C} \in \Sigma,\; A \in \mathcal{C}.$$

Now $P(\emptyset)$ holds, since $\{\emptyset\} \in \Sigma$ and $\emptyset \in \{\emptyset\}$. Let $A \in \mathfrak{F}^{\times}(S)$ and $a \in A$ be given, and assume that $P(A\backslash\{a\})$ holds. We may therefore determine $\mathcal{C} \in \Sigma$ such that $A\backslash\{a\} \in \mathcal{C}$. By the definitions of σ and Seq, we have $A \in \sigma(\mathcal{C}) = \mathrm{Seq}\mathcal{C}$. By the assumption on Σ we have $\mathrm{Seq}\mathcal{C} \in \Sigma$. Thus $P(A)$ holds, and the induction step is complete.

Now let $\mathcal{N} \in \mathrm{N}$ be given, and choose $A \in \mathcal{N}$. By $P(A)$ we may determine $\mathcal{C} \in \Sigma$ such that $A \in \mathcal{C}$. But $\mathcal{C} \in \mathrm{N}$ and $A \in \mathcal{C} \cap \mathcal{N}$. Therefore $\mathcal{N} = \mathcal{C} \in \Sigma$. Since $\mathcal{N} \in \mathrm{N}$ ▲ was arbitrary and $\Sigma \subset \mathrm{N}$, we conclude that $\Sigma = \mathrm{N}$. ■

This page intentionally left blank

Chapter 11

FINITE SUMS

111. Commutative monoids

In much of mathematics one encounters structures that involve a binary operation that is associative and commutative, and an element that acts "neutrally" in this operation; for instance, addition and 0 in $\mathbb{N}$, or multiplication and 1 in $\mathbb{N}$. Such structures are a special case of so-called *algebraic structures*, but here we shall consider them by themselves and only to the extent that is practical for our immediate applications. Some more complicated algebraic structures will be introduced in Chapter 13.

We define a **commutative monoid** (**written additively**) to be a set M endowed with structure by the prescription of a member 0 of M and a mapping $((x, y) \mapsto x+y) : M \times M \to M$, subject to the following conditions:

(CM1): $\forall x, y, z \in M,\ (x + y) + z = x + (y + z)$ (*associative law*)

(CM2): $\forall x, y \in M,\ x + y = y + x$ (*commutative law*)

(CM3): $\forall x \in M,\ x + 0 = x$ (*neutrality law*).

Of course (CM2) and (CM3) imply that $0 + x = x$ for all $x \in M$. The commutative monoid M is said to **have its zero isolated** if

$$\forall x, y \in M, \quad x + y = 0 \ \Rightarrow\ x = y = 0.$$

We adopt the "additive" notation in the definition, because it is the one most frequently occurring in practice. In this notation, 0 is called **zero**, the mapping $(x, y) \mapsto x+y$ is called **addition**, and $x+y$ is called the **sum of** x **and** y (and is read "x **plus** y"). Other notations are used when convenient: when $+$ is replaced by $\cdot$ or mere juxtaposition, and 0 by some other symbol (sometimes by 1), one often speaks of a commutative monoid **written multiplicatively**; in this case, the mapping $(x, y) \mapsto x \cdot y$ is called **multiplication**, and $x \cdot y$, usually denoted xy, is called the **product of** x **and** y, and the member of M that plays the part of 0 is called the **unity**.

111A. Examples. (a): Let the set S be given. Then $\mathfrak{P}(S)$ becomes a commutative monoid with each of the following choices for "zero" and "plus": $\emptyset$ and $\cup$; S and $\cap$; $\emptyset$ and $\triangle$ (symmetric difference). The first and second of these have their "zeros" isolated, but not the third (unless $S = \emptyset$).

(b)*: Let G be a commutative group, written additively. Then G is a commutative monoid with the group zero and the group addition.

(c): Let D be an ordered set that has a minimum and is such that every doubleton has a supremum. Then D is a commutative monoid with $\min D$ as "zero" and $(x, y) \mapsto \sup\{x, y\}$ as "addition". This monoid always has its "zero" isolated.

(d): The set $\mathbb{N}$ with the number 0 and the addition of natural numbers is a commutative monoid and has its zero isolated.

(e): The set $\mathbb{N}$ and the set $\mathbb{N}^\times$, each with the number 1 and the multiplication of natural numbers (adjusted to domain $\mathbb{N}^\times \times \mathbb{N}^\times$ and codomain $\mathbb{N}^\times$ in the latter case) are commutative monoids, written multiplicatively, and either has its unity isolated.

(f)*: The set $\mathbb{R}^\times$, the set $\mathbb{P}^\times$, and the set $1 + \mathbb{P}$, each with the number 1 and the multiplication of real numbers, suitably adjusted, are commutative monoids written multiplicatively; only the third has its unity isolated. ■

112. Finite sums

Let M be a commutative monoid, written additively. On account of the associativity and the commutativity of addition we have

$$\begin{aligned}(x+y)+z &= (y+x)+z = (z+x)+y = (x+z)+y = (y+z)+x = (z+y)+x = \\ &= x+(y+z) = x+(z+y) = y+(z+x) = y+(x+z) = z+(x+y) = \\ &= z+(y+x);\end{aligned}$$

informally, we may say "the sum of x and y and z does not depend on the listing of the summands or on the priority in performing the operations". It is to be expected that a similar comment is meaningful and valid for "more than three terms"; that is, for arbitrary finite families.

A precise formulation calls for a definition of the sum of a finite family in a commutative monoid or, in a slightly more general and more useful manner, the sum of an arbitrary family over a finite subset of its index set. The main purpose of this section and the following ones is to develop this definition and to establish useful properties of this concept.

The strategy for the definition is outlined as follows: we first define sets such as $\{(x+y)+z, \dots, z+(y+x)\}$ of "all possible sums of the family" and then show, using the associative and commutative laws for addition, that each one of these sets is in fact a singleton.

112A. Theorem. *Let M be a commutative monoid, written additively, and let the family $a \in M^I$ be given. Then there is exactly one family $\sigma \in M^{\mathfrak{F}(I)}$ such that*

$$\sigma_\emptyset = 0 \tag{112.1}$$

$$\sigma_J = \sigma_{J\setminus\{j\}} + a_j \quad \text{for all } J \in \mathfrak{F}^\times(I) \text{ and all } j \in J. \tag{112.2}$$

Proof. 1. The collection $\mathfrak{F}(I)$, ordered by inclusion, is well-founded (Theorem 102A). We may therefore, by Theorem 82B, define the family $S \in \mathfrak{P}(M)^{\mathfrak{F}(I)}$ recursively by the rules

$$S_\emptyset := \{0\} \tag{112.3}$$

$$S_J := \bigcup_{j\in J} \{z + a_j \mid z \in S_{J\setminus\{j\}}\} \quad \text{for all } J \in \mathfrak{F}^\times(I). \tag{112.4}$$

We prove by general induction (i.e., using the induction step (102.1)) that S_J is a singleton for every $J \in \mathfrak{F}(I)$. Let $J \in \mathfrak{F}(I)$ be given and assume that S_K is a singleton for all proper subsets K of J. If $J = \emptyset$, then (112.3) asserts that S_J is a singleton. Assume, therefore, that $J \neq \emptyset$. By the induction hypothesis, $S_{J\setminus\{j\}}$ is a singleton, say $\{v_j\}$, for each $j \in J$. Since $J \neq \emptyset$, we may choose $j' \in J$; then $u' := v_{j'} + a_{j'} \in S_J$. Let $u \in S_J$ be given; by (112.4), $u = v_j + a_j$ for a suitable $j \in J$.

If $j = j'$, then $v_j = v_{j'}$ and $u = u'$. If $j \neq j'$ we have $j \in J\backslash\{j'\}$, $j' \in J\backslash\{j\}$, and $(J\backslash\{j'\})\backslash\{j\} = J\backslash\{j,j'\} = (J\backslash\{j\})\backslash\{j'\}$. By the induction hypothesis, $S_{J\backslash\{j,j'\}}$ is a singleton, say $\{w\}$. By (112.4) applied to $J\backslash\{j\}$ and $J\backslash\{j'\}$ instead of J, we find $v_j = w + a_{j'}, v_{j'} = w + a_j$. Since addition is *associative* and *commutative*, we have in this case

$$u = v_j + a_j = (w + a_{j'}) + a_j = w + (a_{j'} + a_j) = w + (a_j + a_{j'}) = (w + a_j) + a_{j'} = v_{j'} + a_{j'} = u'.$$

In either case, therefore, we find $u = u'$. Since $u \in S_J$ was arbitrary, we conclude that $S_J = \{u'\}$ is a singleton. This completes the induction step. Hence S_J is a singleton for every $J \in \mathfrak{F}(I)$.

2. On account of what we have just proved, we may define the family $s \in M^{\mathfrak{F}(I)}$ by $s_J :\in S_J$ for all $J \in \mathfrak{F}(I)$. From (112.3) and (112.4) we then obtain

$$s_\emptyset \in S_\emptyset = \{0\} \tag{112.5}$$

$$s_{J\backslash\{j\}} + a_j \in S_J = \{s_J\} \quad \text{for all } J \in \mathfrak{F}^\times(I) \text{ and all } j \in J. \tag{112.6}$$

Therefore $\sigma := s$ satisfies (112.1) and (112.2).

On the other hand, suppose that $\sigma \in M^{\mathfrak{F}(I)}$ satisfies (112.1) and (112.2). We then prove by special induction that $\sigma_J = s_J$ for all $J \in \mathfrak{F}(I)$; this will end the proof. By (112.1) and (112.5) we have $\sigma_\emptyset = 0 = s_\emptyset$. Let $J \in \mathfrak{F}^\times(I)$ be given and let $j \in J$ be such that $\sigma_{J\backslash\{j\}} = s_{J\backslash\{j\}}$. By (112.2) and (112.6), $\sigma_J = \sigma_{J\backslash\{j\}} + a_j = s_{J\backslash\{j\}} + a_j = s_j$. This completes the induction step. ■

For every family $a \in M^I$ we now define

$$\sum_J a := \sigma_J \qquad \text{for all } J \in \mathfrak{F}(I),$$

where $\sigma \in M^{\mathfrak{F}(I)}$ is the unique family satisfying (112.1) and (112.2). We call $\sum_J a$ the **sum of** the family a **over** J.

If additive notation is not used, a symbol other than $\sum$ is customary. In particular, when the commutative monoid is written multiplicatively, the symbol $\prod$ and the term **product** instead of *sum* are usual.

We now rewrite (112.1) and (112.2) in the newly introduced notation:

$$\sum_\emptyset a = 0 \tag{112.7}$$

$$\sum_J a = \sum_{J\backslash\{j\}} a + a_j \quad \text{for all } J \in \mathfrak{F}^\times(I) \text{ and all } j \in J. \tag{112.8}$$

We also note the following consequences of (112.7) and (112.8):

$$\sum_{\{i\}} a = \sum_{\emptyset} a + a_i = 0 + a_i = a_i \qquad \text{for all } i \in I. \tag{112.9}$$

$$\sum_{\{i,i'\}} a = \sum_{\{i\}} a + a_{i'} = a_i + a_{i'} \qquad \text{for all } i, i' \in I \text{ such that } i \neq i'. \tag{112.10}$$

112B. Proposition. *Let M be a commutative monoid, written additively, and let the family $a \in M^I$ be given. Then*

$$\sum_J a = \sum_J (a|_K) \qquad \textit{for all } K \in \mathfrak{P}(I) \textit{ and all } J \in \mathfrak{F}(K).$$

Proof. Let $K \in \mathfrak{P}(I)$ be given and define $\sigma^K \in M^{\mathfrak{F}(K)}$ by ${\sigma^K}_J := \sum_J a$ for all $J \in \mathfrak{F}(K)$. We have ${\sigma^K}_\emptyset = 0$. For every $J \in \mathfrak{F}^\times(K)$ and every $j \in J$ we have, by (112.8),

$${\sigma^K}_{J\setminus\{j\}} + (a|_K)_j = \sum_{J\setminus\{j\}} a + a_j = \sum_J a = {\sigma^K}_J.$$

By Theorem 112A and the definition of sum applied to $a|_K$ instead of a, we find ${\sigma^K}_J = \sum_J (a|_K)$ for all $J \in \mathfrak{F}(K)$. ∎

112C. Remarks. (a): Proposition 112B allows us to use without ambiguity the notation

$$\sum_{j\in J} a_j = \sum_J a$$

for each family $a \in M^I$ and every $J \in \mathfrak{F}(I)$. This notation is convenient if no explicit name for the family a is available.

(b): If we apply this notation and formulas (112.7) and (112.8) to the commutative monoid $\mathbb{N}$ with the number 0 and addition of natural numbers on the one hand, and to the commutative monoid $\mathbb{N}$, written multiplicatively, with the number 1 and multiplication of natural numbers on the other, we obtain (103.2) and (103.3). ∎

112D. Proposition. *Let M be a commutative monoid, written additively, that has its zero isolated. Let the family $a \in M^I$ and the finite subset J of I be given. If $\sum_J a = 0$, then $a_j = 0$ for all $j \in J$.*

Proof. The assertion certainly holds if $J = \emptyset$. Assume that $J \neq \emptyset$, and let $j \in J$ be given. Then (112.8) shows that $\sum_{J\setminus\{j\}} a + a_j = 0$. Since M has its zero isolated, we conclude that $a_j = 0$. ∎

113. Sums of families with finite support

Let M be a commutative monoid, written additively. Let the family $a \in M^I$ be given. The set $\mathrm{Supp}a := a^<(M\backslash\{0\}) = I\backslash a^<(\{0\})$ is called the **support of** a.

113A. Proposition. *Let M be a commutative monoid, written additively. Let the family $a \in M^I$ be given. Then*

$$\sum_J a = \sum_{J\cap \mathrm{Supp}a} a \qquad \text{for all } J \in \mathfrak{F}(I).$$

Proof. Define the family $\sigma \in M^{\mathfrak{F}(I)}$ by the rule $\sigma_J := \sum_{J\cap\mathrm{Supp}a} a$ for all $J \in \mathfrak{F}(I)$. Then $\sigma_\emptyset = \sum_\emptyset a = 0$. Let $J \in \mathfrak{F}^\times(I)$ and $j \in J$ be given. If $j \notin \mathrm{Supp}a$, then $a_j = 0$ and

$$\sigma_j = \sum_{J\cap\mathrm{Supp}a} a = \sum_{(J\backslash\{j\})\cap\mathrm{Supp}a} a = \sigma_{J\backslash\{j\}} = \sigma_{J\backslash\{j\}} + a_j;$$

if, on the other hand, $j \in \mathrm{Supp}a$, then $(J\cap\mathrm{Supp}a)\backslash\{j\} = (J\backslash\{j\})\cap\mathrm{Supp}a$ and, using (112.8), we obtain

$$\sigma_J = \sum_{J\cap\mathrm{Supp}a} a = \sum_{(J\backslash\{j\})\cap\mathrm{Supp}a} a + a_j = \sigma_{J\backslash\{j\}} + a_j.$$

Thus σ satisfies (112.1) and (112.2). By Theorem 112A and the definition of sum, we have $\sigma_J = \sum_J a$ for all $J \in \mathfrak{F}(I)$, as asserted. ∎

It is desirable to extend the definition of $\sum_J a$ to infinite sets J in certain cases, namely when $\mathrm{Supp}a$ is finite. Proposition 113A suggests the form this definition should take. First, however, we shall show that what we are defining is a natural concept.

113B. Theorem. *Let M be a commutative monoid, written additively. Let the family $a \in M^I$ be given and assume that* $\mathrm{Supp}a$ *is finite. Let the family $\tau \in M^{\mathfrak{P}(I)}$ be given. The following statements are equivalent:*

(i): $\tau_J = \sum_{J\cap\mathrm{Supp}a} a$ *for all* $J \in \mathfrak{P}(I)$;

(ii): τ *satisfies*

(113.1) $\tau_J = 0$ *for all* $J \in \mathfrak{P}(I\backslash\mathrm{Supp}a)$

(113.2) $\tau_J = \tau_{J\backslash\{j\}} + a_j$ *for all* $J \in \mathfrak{P}^\times(I)$ *and all* $j \in J$;

(iii): τ *satisfies* (113.2) *and there exists a set* $K \in \mathfrak{F}(I)$ *such that*

(113.3) $\tau_J = 0$ *for all* $J \in \mathfrak{P}(I\backslash K)$.

Proof. (i) ⇒ (ii). The proof that τ, as defined by (i), satisfies (113.2) is the same, with minor adjustments, as the proof of Proposition 113A. If $J \in \mathfrak{P}(I\backslash \mathrm{Supp}a)$, we have $\tau_J = \sum_{\emptyset} a = 0$, so that τ also satisfies (113.1).

(ii) ⇒ (iii). This follows trivially by setting $K := \mathrm{Supp}a$.

(iii) ⇒ (i). Let the finite subset K of I be chosen so that τ satisfies (113.3). If $j \in \mathrm{Supp}a$, we have $\tau_\emptyset = 0$, by (113.3) and therefore, by (113.2), $\tau_{\{j\}} = \tau_\emptyset + a_j = a_j \neq 0$. Therefore (113.3) implies that $\{j\} \notin \mathfrak{P}(I\backslash K)$, i.e., that $j \in K$. Since $j \in \mathrm{Supp}a$ was arbitrary, we have proved that $\mathrm{Supp}a \subset K$. Therefore Proposition 113A, applied to $J \cap K$ instead of J, shows that

$$\sum_{J\cap \mathrm{Supp}a} a = \sum_{J\cap K} a \qquad \text{for all } J \in \mathfrak{P}(I).$$

To prove that (i) holds it will therefore be enough to prove that $\tau_J = \sum_{J\cap K} a$ for all $J \in \mathfrak{P}(I)$. We prove this by proving by special induction that

$$P(A) :\Leftrightarrow (\forall J \in \mathfrak{P}(I),\ J \cap K = A \ \Rightarrow\ \tau_J = \sum_A a)$$

holds for all $A \in \mathfrak{P}(K) = \mathfrak{F}(K)$.

If $J \in \mathfrak{P}(I)$ and $J \cap K = \emptyset$, then $J \in \mathfrak{P}(I\backslash K)$, and (113.3) implies $\tau_J = 0 = \sum_\emptyset a$, so that $P(\emptyset)$ holds. Let $A \in \mathfrak{F}^\times(K)$ be given and let $i \in A$ be such that $P(A\backslash\{i\})$ holds. Let $J \in \mathfrak{P}(I)$ be given. If $J \cap K = A$, then $(J\backslash\{i\}) \cap K = (J \cap K)\backslash\{i\} = A\backslash\{i\}$; by (113.2) and the induction hypothesis,

$$\tau_J = \tau_{J\backslash\{i\}} + a_i = \sum_{A\backslash\{i\}} a + a_i = \sum_A a;$$

we have shown that $P(A)$ holds. This completes the induction step. We have proved that $P(A)$ holds for all $A \in \mathfrak{F}(K) = \mathfrak{P}(K)$, so that $\tau_J = \sum_{J\cap K} a$ for all $J \in \mathfrak{P}(I)$, as we wished to show. ∎

For every family $a \in M^I$ such that $\mathrm{Supp}a$ is finite, we set $\sum_J{}' a := \sum_{J\cap \mathrm{Supp}a} a$ for all $J \in \mathfrak{P}(I)$. Proposition 113A shows that $\sum_J{}' a = \sum_J a$ for all families $a \in M^I$ and all $J \in \mathfrak{P}(I)$ for which both sides are defined (i.e., J and $\mathrm{Supp}a$ both finite); we may and shall therefore drop the "prime" in this definition, without risk of confusion. Thus $\sum_J a$ is defined for all $a \in M^I$ and $J \in \mathfrak{P}(I)$ such that *either* J *or* $\mathrm{Supp}a$ is finite, and always satisfies

$$\sum_J a = \sum_{J\cap \mathrm{Supp}a} a. \tag{113.4}$$

113C. Proposition. *Let M be a commutative monoid, written additively. Let the family $a \in M^I$ be given and assume that* Supp$a$ *is finite. Then*

$$\sum_J a = \sum_J (a|_K) \qquad \text{for all } K \in \mathfrak{P}(I) \text{ and } J \in \mathfrak{F}(K).$$

Proof. Obviously, $\mathrm{Supp}(a|_K) = K \cap \mathrm{Supp}a$ is finite, so the statement is meaningful. Moreover, $J \cap \mathrm{Supp}a = J \cap K \cap \mathrm{Supp}a = J \cap \mathrm{Supp}(a|_K)$. By Proposition 112B and (113.4) we have

$$\sum_J a = \sum_{J \cap \mathrm{Supp}a} a = \sum_{J \cap \mathrm{Supp}(a|_K)} a = \sum_{J \cap \mathrm{Supp}(a|_K)} (a|_K) = \sum_J (a|_K). \blacksquare$$

113D. Remark. Proposition 113C allows us to use without ambiguity the notation

$$\sum_{j \in J} a_j := \sum_J a$$

for each family $a \in M^I$ with Supp$a$ finite and for all $J \in \mathfrak{P}(I)$. (Cf. Remark 112C,(a).) ■

113E. Proposition. *Let M be a commutative monoid, written additively, that has its zero isolated. Let the family $a \in M^I$ and the subset J of I be given. If* Supp$a$ *is finite and $\sum_J a = 0$, then $a_j = 0$ for all $j \in J$.*

Proof. This follows immediately from (113.4) and Proposition 112D. ■

114. Repeated and double sums

This section is devoted to some tedious but necessary "bookkeeping" aspects of dealing with sums. They are all versions of the following general associative law.

114A. Theorem. *Let M be a commutative monoid, written additively, and let the family $a \in M^I$ and the mapping $\omega : I \to K$ be given. For every finite subset J of I, and for every subset J of I if* Supp$a$ *is finite, the support of the family* $(\sum_{J\cap\omega^<(\{k\})} a \mid k \in K) \in M^K$ *is finite, and*

$$\sum_J a = \sum_{k\in K} \sum_{J\cap\omega^<(\{k\})} a = \sum_{k\in\omega_>(J)} \sum_{J\cap\omega^<(\{k\})} a.$$

Proof. 1. We assume for the time being that Supp$a$ is finite. For each subset J of I we set $b^J := (\sum_{J\cap\omega^<(\{k\})} a \mid k \in K) \in M^K$. If $k \in \text{Supp}b^J$ we have $\sum_{J\cap\omega^<(\{k\})} a \neq 0$. By Theorem 113B this implies $J \cap \omega^<(\{k\}) \cap \text{Supp}a \neq \emptyset$, i.e., $k \in \omega_>(J \cap \text{Supp}a)$. We conclude that

$$\text{Supp}b^J \subset \omega_>(J \cap \text{Supp}a) \qquad \text{for all } J \in (I). \tag{114.1}$$

Since Supp$a$ is finite, so is Suppb^J, as asserted (Corollary 101F and Proposition 101G). We are to prove that

$$\sum_J a = \sum_K b^J = \sum_{\omega_>(J)} b^J \qquad \text{for all } J \in \mathfrak{P}(I). \tag{114.2}$$

By (114.1) we have Supp$b^J \subset \omega_>(J)$, so that the second equality in (114.2) follows from (113.4).

2. We define the family $\tau := (\sum_K b^J \mid J \in \mathfrak{P}(I)) \in M^{\mathfrak{P}(I)}$. If $J \cap \text{Supp}a = \emptyset$, then Supp$b^J = \emptyset$, by (114.1), whence $\tau_J = \sum_K b^J = \sum_{K\cap\text{Supp}b^J} b^J = \sum_\emptyset b^J = 0$, so that τ satisfies (113.1).

Let $J \in \mathfrak{P}^\times(I)$ and $j \in J$ be given, and set $k := \omega(j)$. We find $(J\backslash\{j\})\cap\omega^<(\{k\}) = (J \cap \omega^<(\{k\}))\backslash\{j\}$, whence

$$b^J{}_k = \sum_{J\cap\omega^<(\{k\})} a = (\sum_{(J\cap\omega^<(\{k\}))\backslash\{j\}} a) + a_j = (\sum_{(J\backslash\{j\})\cap\omega^<(\{k\})} a) + a_j =$$

$$= b^{J\backslash\{j\}}{}_k + a_j.$$

On the other hand, for every $k' \in K\backslash\{k\}$ we have $(J\backslash\{j\})\cap\omega^<(\{k'\}) = J\cap\omega^<(\{k'\})$, since $j \notin \omega^<(\{k'\})$, and therefore

$$b^J|_{K\backslash\{k\}} = b^{J\backslash\{j\}}|_{K\backslash\{k\}}.$$

Combining these equalities and applying Proposition 113C we find

$$\tau_J = \sum_K b^J = \sum_{K\setminus\{k\}} b^J + b^J{}_k = \sum_{K\setminus\{k\}} b^{J\setminus\{j\}} + b^{J\setminus\{j\}}{}_k + a_j = \sum_K b^{J\setminus\{j\}} + a_j = \tau_{J\setminus\{j\}} + a_j.$$

Since $J \in \mathfrak{P}^\times(I)$ and $j \in J$ were arbitrary, τ also satisfies (113.2). By Theorem 113B we have $\sum_K b^J = \tau_J = \sum_J a$ for all $J \in \mathfrak{P}(I)$, so that the first equality in (114.2) also holds.

3. We now drop the assumption that Supp$a$ is finite and assume that the finite subset J of I is given. Then $\mathrm{Supp}(a|_J)$ is finite, and we can apply the preceding part of the proof to $a|_J$ instead of a. Combining the results of this application with Proposition 113C, we find that the assertion holds for the given J. ■

114B. Remark. The assertions of Propositions 112B and 113C are special cases of the assertion of Theorem 114A, with I, K, a, ω replaced, respectively, by K, I, $a|_K$, $1_{K\subset I}$. ■

114C. Corollary. *Let M be a commutative monoid, written additively, and let the family $a \in M^I$ and the subset J of I be given. Assume that either J or* Supp$a$ *is finite. For every partition $\mathcal{P}$ of J we have*

$$\sum_J a = \sum_{A\in\mathcal{P}} \sum_A a.$$

In particular, for every subset K of J we have

$$\sum_J a = \sum_K a + \sum_{J\setminus K} a.$$

Proof. By Propositions 112B and 113C we may, replacing a by $a|_J$ if necessary, assume that $J = I$. We recall that $\Omega_{\mathcal{P}}{}^<(\{A\}) = A$ for every $A \in \mathcal{P}$, where $\Omega_{\mathcal{P}} : I \to \mathcal{P}$ is the partition mapping. The assertion then follows from Theorem 114A with $K := \mathcal{P}$ and $\omega := \Omega_{\mathcal{P}}$. ■

114D. Corollary. *Let M be a commutative monoid, written additively. Let the finite family of finite sets $(J_i \mid i \in I)$ be given. Set $U := \dot{\bigcup_{i\in I}} J_i$. Let the family of families $a \in \bigtimes_{i\in I} M^{J_i}$ be given. Then*

$$\sum_{(i,j)\in U} (a_i)_j = \sum_{i\in I} \sum_{j\in J_i} (a_i)_j.$$

114E. Corollary. *Let M be a commutative monoid, written additively, and let the family $a \in M^I$ and the invertible mapping $\omega : I \to K$ be given. For every finite subset J of I, and for every subset J of I if* Supp$a$ *is finite, we have*

$$\sum_J a = \sum_{\omega_>(J)} a \circ \omega^{\leftarrow}.$$

Proof. For each $k \in \omega_>(J)$ we have $J \cap \omega^<(\{k\}) = J \cap \{\omega^{\leftarrow}(k)\} = \{\omega^{\leftarrow}(k)\}$. By (112.9) we have $\sum_{J\cap\omega^<(\{k\})} a = \sum_{\{\omega^{\leftarrow}(k)\}} a = a_{\omega^{\leftarrow}(k)} = (a \circ \omega^{\leftarrow})_k$. The assertion follows from Theorem 114A. ■

114F. Theorem. *Let M be a commutative monoid, written additively. Let the matrix $a \in M^{I'\times I''}$ and the subsets J' of I' and J'' of I'' be given, and assume that either J' and J'' are both finite or* Supp$a$ *is finite. Then*

$$\sum_{j'\in J'} \sum_{j''\in J''} a_{j',j''} = \sum_{J'\times J''} a = \sum_{j''\in J''} \sum_{j'\in J'} a_{j',j''}.$$

Proof. Denote the mapping $((i',i'') \mapsto i') : I' \times I'' \to I'$ by π'. For each $j' \in J'$ we have $\pi'^<(\{j'\}) = \{j'\} \times I''$, so that $(J' \times J'') \cap \pi'^<(\{j'\}) = \{j'\} \times J''$. Applying Theorem 114A (as we may, since $J' \times J''$ is finite if J', J'' are both finite, by Corollary 103G) we find

$$\sum_{J'\times J''} a = \sum_{j'\in J'} \sum_{\{j'\}\times J''} a. \tag{114.3}$$

For each fixed $j' \in J'$, $\mathrm{Supp}(a|_{\{j'\}\times J''}) \subset (J' \times J'') \cap \mathrm{Supp}a$ is finite, and each of the mappings $((j',j'') \to j'') : \{j'\} \times J'' \to J''$ and $(j'' \mapsto (j',j'')) : J'' \to \{j'\} \times J''$ is the inverse of the other. By Proposition 112B or 113C, as the case may require, and Corollary 114E, we find

$$\sum_{\{j'\}\times J''} a = \sum_{\{j'\}\times J''} (a|_{\{j'\}\times J''}) = \sum_{j''\in J''} (a|_{\{j'\}\times J''})_{j',j''} = \sum_{j''\in J''} a_{j',j''} \quad \text{for all } j' \in J'. \tag{114.4}$$

Combining (114.3) and (114.4) we obtain the first equality in the assertion. The second equality follows in the same way, with the roles of I' and I'', etc., interchanged. ■

114G. Corollary. *Let M be a commutative monoid, written additively. Let the sets I' and I'' and the subsets K and L of $I' \times I''$ be given, and assume that $K \subset L$. Let the family $a \in M^L$ be given. Assume that either K or* Supp$a$ *is finite. Then*

$$\sum_{i'\in I'} \sum_{i''\in K_{i'}} a_{i',i''} = \sum_K a = \sum_{i''\in I''} \sum_{i'\in K^{i''}} a_{i',i''},$$

where $K_{i'} := \{i'' \in I'' \mid (i',i'') \in K\}$ for all $i' \in I'$, and $K^{i''} := \{i' \in I' \mid (i',i'') \in K\}$ for all $i'' \in I''$.

Proof. Define the matrix $b \in M^{I'\times I''}$ by

$$b_{i',i''} := \begin{cases} a_{i',i''} & \text{if } (i',i'') \in K \\ 0 & \text{if } (i',i'') \in (I' \times I'')\backslash K, \end{cases}$$

and apply Theorem 114F to $J' := I'$, $J'' := I''$, and b instead of a. ■

115. Natural multiples

If $p, n \in \mathbb{N}$, Lemma 103C and Corollary 101K show that $\sum_{k\in n^\sqsubset} p = np$ and $\prod_{k\in n^\sqsubset} p = p^n$. We introduce a generalization. Let M be a commutative monoid, written additively. For every $x \in M$ and $n \in \mathbb{N}$ we define the n**th multiple of** x to be

$$nx := \sum_{k\in n^\sqsubset} x. \tag{115.1}$$

In particular, (115.1) and (112.7), (112.9), (113.4) yield

$$0x = 0 \qquad 1x = x \qquad n0 = 0. \tag{115.2}$$

If M is written multiplicatively instead, we correspondingly define the n**th power of** x to be $x^n := \prod_{k\in n^\sqsubset} x$.

115A. Lemma. *Let M be a commutative monoid, written additively. For every finite set I and every $x \in M$ we have $\sum_{i\in I} x = (\#I)x$.*

Proof. By Theorem 101B, I and $(\#I)^\sqsubset$ are equinumerous. By Corollary 114E and (115.1) we have $\sum_{i\in I} x = \sum_{k\in(\#I)^\sqsubset} x = (\#I)x$. ∎

115B. Proposition. *Let M be a commutative monoid, written additively.*

(a): *Let the finite family $p \in \mathbb{N}^L$ and $x \in M$ be given. Then*

$$(\sum_L p)x = \sum_{l\in L} p_l x.$$

(b): *Let $m, n \in \mathbb{N}$ and $x \in M$ be given. Then $(mn)x = m(nx)$.*

(c): *Let $n \in \mathbb{N}$ and the family $b \in M^I$ be given. For every finite subset J of I, and for every subset J of I if* Supp$b$ *is finite, we have*

$$n\sum_J b = \sum_{j\in J} nb_j.$$

(d): *Let the finite family $p \in \mathbb{N}^L$ and the family $b \in M^I$ be given. For every finite subset J of I, and for every subset J of I if* Supp$b$ *is finite, we have*

$$(\sum_L p)(\sum_J b) = \sum_{(l,j)\in L\times J} p_l b_j.$$

Proof. Proof of (a). Set $U := \dot{\bigcup_{l\in l}} p_l{}^\sqsubset$. By Theorem 103D and Corollary 101K, U is finite and

$$\#U = \sum_{l\in L} \#(p_l{}^\sqsubset) = \sum_{l\in L} p_l. \tag{115.3}$$

We define the mapping $\omega : U \to L$ by the rule

$$\omega((l,k)) := l \qquad \text{for all } (l,k) \in U,$$

and find $\omega^<(\{l\}) = \{l\} \times p_l^\sqsubset$ for each $l \in L$, and therefore, by Corollaries 103G and 101K,

$$\#\omega^<(\{l\}) = (\#\{l\})(\#(p_l^\sqsubset)) = 1 \cdot p_l = p_l \qquad \text{for all } l \in L. \tag{115.4}$$

We also define the family $a \in M^U$ by the rule

$$a_{l,k} := x \qquad \text{for all } (l,k) \in U. \tag{115.5}$$

We apply consecutively (115.3), Lemma 115A, (115.5), Theorem 114A, (115.5), Lemma 115A, and (115.4) to find

$$(\sum_L p)x = (\#U)x = \sum_{(l,k)\in U} x = \sum_U a = \sum_{l\in L} \sum_{\omega^<(\{l\})} a = \sum_{l\in L} \sum_{k\in\omega^<(\{l\})} x =$$

$$= \sum_{l\in L} (\#\omega^<(\{l\}))x = \sum_{l\in L} p_l x.$$

Proof of (b). From (a) we have

$$(mn)x = (\sum_{k\in m^\sqsubset} n)x = \sum_{k\in m^\sqsubset} nx = m(nx).$$

Proof of (c). We define the matrix $a \in M^{n^\sqsubset \times I}$ by the rule

$$a_{k,i} := b_i \qquad \text{for all } (k,i) \in n^\sqsubset \times I.$$

We then apply (115.1) and Theorem 114F,

$$n\sum_J b = \sum_{k\in n^\sqsubset} \sum_{j\in J} b_j = \sum_{k\in n^\sqsubset} \sum_{j\in J} a_{k,j} = \sum_{j\in J} \sum_{k\in n^\sqsubset} a_{k,j} = \sum_{j\in J} \sum_{k\in n^\sqsubset} b_j = \sum_{j\in J} nb_j.$$

Proof of (d). We combine (a) and (c) with Theorem 114F to find

$$(\sum_L p)(\sum_J b) = \sum_{l\in L} (p_l \sum_J b) = \sum_{l\in L} \sum_{j\in J} p_l b_j = \sum_{(l,j)\in L\times J} p_l b_j. \blacksquare$$

116. The Inclusion-Exclusion Principle

In Corollary 103E we proved that the union of a finite family of finite sets is finite. In Sections 112 and 114 we have developed the machinery that allows us to provide a precise determination of the cardinal of such a union, as a generalization of Corollary 103B. We therefore interrupt the general account of finite sums (the remainder of the chapter being largely concerned with additional "bookkeeping" issues) to provide this precise determination, a result known as the Inclusion-Exclusion Principle.

It happens that both the formulation of the Inclusion-Exclusion Principle and its proof are facilitated by the use of *(positive and negative) integers* instead of restricting all terms to be members of $\mathbb{N}$. The reader's indulgence is requested with regard to this foray into formally as-yet-undefined terrain. The concerned reader may, in fact, readily convert all statements and arguments in the following account into statements and arguments about natural numbers; the exercise is straightforward, but rather annoying. To satisfy that reader's curiosity, a "natural-number" formulation of the Inclusion-Exclusion Principle is recorded in Remark 116C.

We propose to give three separate proofs of this important result, as an illustration of the diversity of approaches in this combinatorial context.

116A. Theorem (Inclusion-Exclusion Principle). *For every finite family $(T_i \mid i \in I)$ of finite sets,*

$$\#\Big(\bigcup_{i\in I} T_i\Big) = \sum_{J\in\mathfrak{P}^\times(I)} (-1)^{\#J-1} \#\Big(\bigcap_{i\in J} T_i\Big). \tag{116.1}$$

Proof I. We prove by special induction (Inductive Proof Schemes 102E) that $P(I)$ holds for all finite sets I, where

$P(I) \;:\Leftrightarrow$ For every family $(T_i \mid i \in I)$ of finite sets, (116.1) holds.

If $I = \emptyset$, we have $\bigcup_{i\in I} T_i = \emptyset$ and $\mathfrak{P}^\times(I) = \emptyset$, so that both sides of (116.1) are 0, and $P(\emptyset)$ holds.

Let the non-empty finite set I and $k \in I$ be given, set $K := I\backslash\{k\}$, and assume that $P(K)$ holds. To prove $P(I)$, thus completing the induction step, we let the family $(T_i \mid i \in I)$ of finite sets be given and claim that it satisfies (116.1).

By Corollary 103B we have

$$\begin{aligned}\#\Big(\bigcup_{i\in I} T_i\Big) &= \#\Big(T_k \cup \bigcup_{i\in K} T_i\Big) = \#\Big(\bigcup_{i\in K} T_i\Big) + \#T_k - \#\Big(T_k \cap \bigcup_{i\in K} T_i\Big) = \\ &= \#\Big(\bigcup_{i\in K} T_i\Big) + \#T_k - \#\Big(\bigcup_{i\in K} (T_k \cap T_i)\Big).\end{aligned}$$

Applying the induction hypothesis $P(K)$ to the families $(T_i \mid i \in K)$ and $(T_k \cap T_i \mid i \in$

K), we then find

$$\begin{aligned}\#(\bigcup_{i\in I} T_i) = &\sum_{J\in\mathfrak{P}^\times(K)} (-1)^{\#J-1}\#(\bigcap_{i\in J} T_i) + \#T_k \;- \\ &- \sum_{J\in\mathfrak{P}^\times(K)} (-1)^{\#J-1}\,\#(\bigcap_{i\in J}(T_k\cap T_i)).\end{aligned} \tag{116.2}$$

The mappings $J\mapsto J\cup\{k\} : \mathfrak{P}(I)\backslash\mathfrak{P}(K)\to\mathfrak{P}(K)$ and $J\mapsto J\backslash\{k\} : \mathfrak{P}(I)\backslash\mathfrak{P}(K)\to\mathfrak{P}(K)$ are inverse to each other: they are both adjustments of the involutory mapping $J\mapsto J\triangle\{k\} : \mathfrak{P}(I)\to\mathfrak{P}(I)$; and they match $\emptyset$ in $\mathfrak{P}(K)$ with $\{k\}$ in $\mathfrak{P}(I)\backslash\mathfrak{P}(K)$. Therefore, using Corollary 114E,

$$\begin{aligned}\#T_k - &\sum_{J\in\mathfrak{P}^\times(K)} (-1)^{\#J-1}\#(\bigcap_{i\in J}(T_k\cap T_i)) = \\ &= \#T_k + \sum_{J\in\mathfrak{P}^\times(K)} (-1)^{\#J}\#(\bigcap_{i\in J\cup\{k\}} T_i) = \\ &= (-1)^{\#\{k\}-1}\#(\bigcap_{i\in\{k\}} T_i) + \sum_{J\in(\mathfrak{P}(I)\backslash\mathfrak{P}(K))\backslash\{k\}} (-1)^{\#J-1}\#(\bigcap_{i\in J} T_i) = \\ &= \sum_{J\in\mathfrak{P}(I)\backslash\mathfrak{P}(K)} (-1)^{\#J-1}\#(\bigcap_{i\in J} T_k).\end{aligned} \tag{116.3}$$

Combining (116.2) and (116.3), we find

$$\begin{aligned}\#(\bigcup_{i\in I} T_i) &= \sum_{J\in\mathfrak{P}^\times(K)} (-1)^{\#J-1}\#(\bigcap_{i\in J} T_i) + \sum_{J\in\mathfrak{P}(I)\backslash\mathfrak{P}(K)} (-1)^{\#J-1}\#(\bigcap_{i\in J} T_i) = \\ &= \sum_{J\in\mathfrak{P}^\times(I)} (-1)^{\#J-1}\#(\bigcap_{i\in J} T_i),\end{aligned}$$

so that (116.1) holds, as claimed. This completes Proof I.

For Proofs II and III of Theorem 116A we shall need the following auxiliary result.

116B. Lemma. *For every non-empty finite set L we have* $\displaystyle\sum_{J\in\mathfrak{P}(L)} (-1)^{\#J} = 0.$

Proof. Choose $p\in L$. The involutory mapping $J\mapsto J\triangle\{p\} : \mathfrak{P}(L)\to\mathfrak{P}(L)$ induces a bijection from the collection of subsets of L with even cardinal to the collection of subsets of L with odd cardinal. These collections are therefore equinumerous, and the conclusion follows at once. ∎

Proof II of Theorem 116A. We proceed by special induction (Induction Proof Schemes 102E), proving that $Q(S)$ holds for all finite sets S, where

$$\begin{aligned}Q(S) \;:\Leftrightarrow\;& \text{For every finite family } (T_i \mid i\in I) \text{ of finite sets such that} \\ & \bigcup_{i\in I} T_i = S,\ (116.1) \text{ holds.}\end{aligned}$$

If $S = \emptyset$ and $(T_i \mid i \in I)$ satisfies $\bigcup_{i\in I} T_i = S$, then $T_i = \emptyset$ for all $i \in I$, and each summand in the right-hand side of (116.1) is 0, as is $\#S$, the left-hand side. Thus $Q(\emptyset)$ holds.

Let the non-empty finite set S and $s \in S$ be given, set $R := S\backslash\{s\}$ and assume that $Q(R)$ holds. To prove $Q(S)$, thus completing the induction step, we let the finite family $(T_i \mid i \in I)$ of finite sets be given, assume that

$$\bigcup_{i\in I} T_i = S, \tag{116.4}$$

and claim that (116.1) holds.

We define the family $(U_i \mid i \in I)$ by $U_i := T_i\backslash\{s\} = T_i \cap R$ for all $i \in I$. Then $\bigcup_{i\in I} U_i = \bigcup_{i\in I}(T_i \cap R) = R \cap \bigcup_{i\in I} T_i = R \cap S = R$. By the induction hypothesis $Q(R)$ applied to $(U_i \mid i \in I)$, we have

$$\#R = \#(\bigcup_{i\in I} U_i) = \sum_{J\in\mathfrak{P}^\times(I)} (-1)^{\#J-1} \#(\bigcap_{i\in J} U_i). \tag{116.5}$$

Set $L := \{i \in I \mid s \in T_i\}$. By (116.4), $L \neq \emptyset$. For each $J \in \mathfrak{P}^\times(I)$,

$$\bigcap_{i\in J} T_i = \begin{cases} \{s\} \cup \bigcap_{i\in J} U_i & \text{if } J \subset L \\ \\ \bigcap_{i\in J} U_i & \text{otherwise.} \end{cases}$$

Therefore, using Lemma 116B,

$$\begin{aligned} &\sum_{J\in\mathfrak{P}^\times(I)} (-1)^{\#J-1} \#(\bigcap_{i\in J} T_i) - \sum_{J\in\mathfrak{P}^\times(I)} (-1)^{\#J-1} \#(\bigcap_{i\in J} U_i) = \\ &= \sum_{J\in\mathfrak{P}^\times(I)} (-1)^{\#J-1} (\#(\bigcap_{i\in J} T_i) - \#(\bigcap_{i\in J} U_i)) = \sum_{J\in\mathfrak{P}^\times(L)} (-1)^{\#J-1} = \\ &= 1 - \sum_{J\in\mathfrak{P}(L)} (-1)^{\#J} = 1 - 0 = 1. \end{aligned} \tag{116.6}$$

Combining (116.4), (116.5), and (116.6), we find

$$\#(\bigcup_{i\in I} T_i) = \#S = \#R + 1 = \sum_{J\in\mathfrak{P}^\times(I)} (-1)^{\#J-1} \#(\bigcap_{i\in J} U_i) + 1 = \sum_{J\in\mathfrak{P}^\times(I)} (-1)^{\#J-1} \#(\bigcap_{i\in J} T_i),$$

so that (116.1) holds, as claimed.

Proof III of Theorem 116A. Set $S := \bigcup_{i\in I} T_i$. Then all the intersections in the right-hand side of (116.1) are subsets of S. Define $L_x := \{i \in I \mid x \in T_i\}$ for each $x \in S$, and note that $L_x \neq \emptyset$ for all $x \in S$. Further define $V_J := \bigcap_{i\in J} T_i$ for all $J \in \mathfrak{P}^\times(I)$, and

$$P := \{(x, J) \in S \times \mathfrak{P}^\times(I) \mid x \in V_J\}.$$

Note that

$$\forall x \in S,\ \forall J \in \mathfrak{P}^\times(I), \quad x \in V_j \ \Leftrightarrow\ (x, J) \in P \ \Leftrightarrow\ J \subset L_x. \tag{116.7}$$

Now, using (116.7) and Lemma 116B (applied to the non-empty set L_x for each $x \in S$), as well as Corollary 103E and Corollary 114G,

$$\sum_{J\in\mathfrak{P}^\times(I)} (-1)^{\#J-1} \#\Big(\bigcap_{i\in J} T_i\Big) = \sum_{J\in\mathfrak{P}^\times(I)} (-1)^{\#J-1} \sum_{x\in V_J} 1 = \sum_{J\in\mathfrak{P}^\times(I)} \sum_{x\in V_J} (-1)^{\#J-1} =$$

$$= \sum_{(x,J)\in P} (-1)^{\#J-1} = \sum_{x\in S} \sum_{J\in\mathfrak{P}^\times(L_x)} (-1)^{\#J-1} = \sum_{x\in S} \Big(1 - \sum_{J\in\mathfrak{P}(L_x)} (-1)^{\#J}\Big) =$$

$$= \sum_{x\in S} (1 - 0) = \#S = \#\Big(\bigcup_{i\in I} T_i\Big). \ \blacksquare$$

116C. Remark. The Inclusion-Exclusion Principle may be reformulated thus: *For every finite family $(T_i \mid i \in I)$ of finite sets,*

$$\#\Big(\bigcup_{i\in I} T_i\Big) = \sum_{J\in\mathfrak{P}_{\mathrm{o}}(I)} \#\Big(\bigcap_{i\in J} T_i\Big) - \sum_{J\in\mathfrak{P}_{\mathrm{e}}^\times(I)} \#\Big(\bigcap_{i\in J} T_i\Big),$$

where $\mathfrak{P}_{\mathrm{o}}(I)$ and $\mathfrak{P}_{\mathrm{e}}^\times(I)$ are, respectively, the collection of subsets of I with odd cardinal and the collection of non-empty subsets of I with even cardinal. $\blacksquare$

117. Sums in monoids of families

Let M be a commutative monoid, written additively. Let the set I be given. We set $0_I := (0 \mid i \in I) \in M^I$ and define the mapping $((a,b) \mapsto a +_I b) : M^I \times M^I \to M^I$ by **termwise addition**, i.e., by the rule

$$(a +_I b)_i := a_i + b_i \qquad \text{for all } i \in I \text{ and } a, b \in M^I.$$

117A. Proposition. *Let M be a commutative monoid, written additively, and let the set I be given. Then the specification of 0_I as zero and the mapping $((a,b) \mapsto a +_I b) : M^I \times M^I \to M^I$ as addition endows M^I with the structure of a commutative monoid, written additively. For all $a, b \in M^I$ we have* $\mathrm{Supp}(a +_I b) \subset \mathrm{Supp}a \cup \mathrm{Supp}b$.

117B. Remark. All references to M^I as a commutative monoid will be to the structure described in Proposition 117A. We shall write 0 and + instead of 0_I and $+_I$ without risk of confusion. ■

117C. Lemma. *Let M be a commutative monoid, written additively, and let the sets I' and I'' be given. Denote the mappings $((i', i'') \mapsto i') : I' \times I'' \to I'$ and $((i', i'') \mapsto i'') : I' \times I'' \to I''$ by π' and π'', respectively. Let the matrix $a \in M^{I' \times I''}$ and the families $a' \in (M^{I''})^{I'}$ and $a'' \in (M^{I'})^{I''}$ be related by*

$$(a'_{i'})_{i''} = a_{i',i''} = (a''_{i''})_{i'} \qquad \text{for all } (i', i'') \in I' \times I'' \tag{117.1}$$

(i.e., $a'_{i'}$ is the i'th row of a and $a''_{i''}$ is the i''th column of a). Then

$$\mathrm{Supp}a = \bigcup_{i' \in \mathrm{Supp}a'} \{i'\} \times \mathrm{Supp}a'_{i'} = \bigcup_{i'' \in \mathrm{Supp}a''} \mathrm{Supp}a''_{i''} \times \{i''\}$$

$$\mathrm{Supp}a' = \pi'_>(\mathrm{Supp}a) = \bigcup_{i'' \in \mathrm{Supp}a''} \mathrm{Supp}a''_{i''}$$

$$\mathrm{Supp}a'' = \pi''_>(\mathrm{Supp}a) = \bigcup_{i' \in \mathrm{Supp}a'} \mathrm{Supp}a'_{i'}.$$

Moreover, the following statements are equivalent:

(i): Supp$a$ *is finite.*

(ii): Suppa' *is finite and* Supp$a'_{i'}$ *is finite for every* $i' \in \mathrm{Supp}a'$.

(iii): Suppa'' *is finite and* Supp$a''_{i''}$ *is finite for every* $i'' \in \mathrm{Supp}a''$.

Proof. This follows by direct verification and use of Sections 101 and 103. Note that $\mathrm{Supp}a'_{i'} \neq \emptyset$ if and only if $i' \in \mathrm{Supp}a'$. ■

117D. Proposition. *Let M be a commutative monoid, written additively, and let the sets I' and I'' be given. Let the family $a' \in (M^{I''})^{I'}$ and the subset J' of I' be given, and assume that either J' or* Suppa' *is finite. In the later case, the support of*

the family $((a'_{i'})_{i''} \mid i' \in I')$ *is finite for each* $i'' \in I''$. *In either case,*

$$(\sum_{J'} a')_{i''} = \sum_{j' \in J'} (a'_{j'})_{i''} \qquad \text{for every } i'' \in I''. \tag{117.2}$$

(The first sum in $M^{I''}$, *the second in* M.*)*

Proof. Define the matrix $a \in M^{I' \times I''}$ and the family $a'' \in (M^{I'})^{I''}$ by (117.1). Then the family $((a'_{i'})_{i''} \mid i' \in I')$ is precisely $a''_{i''}$ for every $i'' \in I''$. By Lemma 117C, $\mathrm{Supp} a''_{i''} \subset \mathrm{Supp} a'$, so the former set is finite, as asserted, if the latter one is.

To prove (117.2) we suppose first that $\mathrm{Supp} a'$ is finite and that $i'' \in I''$ is given, but allow the subset J' of I' to vary. Define the family $\tau := ((\sum_{J'} a')_{i''} \mid J' \in \mathfrak{P}(I')) \in M^{\mathfrak{P}(I')}$. If $J' \in \mathfrak{P}^\times(I')$ and $j' \in J'$ are given, we have (using Remark 117B and (117.1))

$$\tau_{J'} = (\sum_{J'} a')_{i''} = (\sum_{J' \setminus \{j'\}} a' + a'_{j'})_{i''} = (\sum_{J' \setminus \{j'\}} a')_{i''} + (a'_{j'})_{i''} = \tau_{J' \setminus \{j'\}} + (a''_{i''})_{j'}.$$

Moreover, if $J' \in \mathfrak{P}(I' \setminus \mathrm{Supp} a')$ we have $\tau_{J'} = (\sum_{J'} a')_{i''} = 0_{i''} = 0$, by (113.4). Since $\mathrm{Supp} a'$ is a finite set, τ satisfies the condition (iii) of Theorem 113B with I', $a''_{i''}$, $\mathrm{Supp} a'$ instead of I, a, K. By that theorem and (113.4) we conclude that $\tau = (\sum_{J'} a''_{i''} \mid J' \in \mathfrak{P}(I'))$, and therefore, using (117.1) again,

$$(\sum_{J'} a')_{i''} = \tau_{J'} = \sum_{J'} a''_{i''} = \sum_{j' \in J'} (a''_{i''})_{j'} = \sum_{j' \in J'} (a'_{j'})_{i''}.$$

Since $i'' \in I''$ was arbitrary, we have proved (117.2) in this case.

Suppose now that J' is a finite subset of I', but that $\mathrm{Supp} a'$ is not necessarily finite. By Proposition 112B and the preceding conclusion applied to $a'|_{J'}$ (whose support is of course finite) instead of a' we obtain the desired conclusion in this case as well. ■

117E. Theorem. *Let* M *be a commutative monoid, written additively, and let the sets* I' *and* I'' *be given. Let the family* $a' \in (M^{I''})^{I'}$ *and the subsets* J' *of* I' *and* J'' *of* I'' *be given, and assume that either* J' *and* J'' *are both finite, or that* $\mathrm{Supp} a'$ *is finite and* $\mathrm{Supp} a'_{i'}$ *is finite for each* $i' \in I'$. *In the latter case,* $\mathrm{Supp}(\sum_{J'} a')$ *is finite. In either case,*

$$\sum_{J''} \sum_{J'} a' = \sum_{j' \in J'} \sum_{J''} a'_{j'}.$$

Proof. Define $a \in M^{I' \times I''}$ and $a'' \in (M^{I'})^{I''}$ by (117.1). By Proposition 117D, we have (117.2). Hence $i'' \in \mathrm{Supp}(\sum_{J'} a')$ implies $\mathrm{Supp} a''_{i''} \neq \emptyset$, which in turn implies

$i'' \in \mathrm{Supp}a''$. By Lemma 117C we then have

$$\mathrm{Supp}(\sum_{J'} a') \subset \mathrm{Supp}a'' = \bigcup_{i' \in \mathrm{Supp}a'} \mathrm{Supp}a'_{i'}.$$

If $\mathrm{Supp}a'$ is finite and $\mathrm{Supp}a'_{i'}$ is finite for each $i' \in I'$, it follows from Corollary 103E that $\mathrm{Supp}(\sum_{J'} a')$ is finite, as asserted; Lemma 117C then further shows that $\mathrm{Supp}a$ is also finite. By Proposition 117D and Theorem 114F we have in either case.

$$\begin{aligned}\sum_{J''}\sum_{J'} a' &= \sum_{j'' \in J''} (\sum_{J'} a')_{j''} = \sum_{j'' \in J''} \sum_{j' \in J'} (a'_{j'})_{j''} = \sum_{j'' \in J''} \sum_{j' \in J'} a_{j',j''} = \\ &= \sum_{j' \in J'} \sum_{j'' \in J''} a_{j',j''} = \sum_{j' \in J'} \sum_{j'' \in J''} (a'_{j'})_{j''} = \sum_{j' \in J'} \sum_{J''} a'_{j'}.\end{aligned}$$

■

Let the set I be given. We define the subset $M^{(I)} := \{a \in M^I \mid \mathrm{Supp}a \text{ is finite}\}$ of M^I and note that by Proposition 117A we have

$$a + b \in M^{(I)} \qquad \text{for all } a, b \in M^{(I)}.$$

Of course $M^{(I)} = M^I$ if and only if either I is finite or $M = \{0\}$.

The next theorem provides a different, more "algebraic" characterization of the sum of a family with finite support in M.

117F. Theorem. *Let M be a commutative monoid, written additively, and let the set I, the subset J of I, and the mapping $S : M^{(I)} \to M$ be given. The following statements are equivalent:*

(i): $S(a) = \sum_J a$ *for all* $a \in M^{(I)}$.

(ii): S *satisfies the following three conditions:*

$$S(0) = 0 \tag{117.3}$$

$$S(u + v) = S(u) + S(v) \qquad \text{for all } u, v \in M^{(I)} \tag{117.4}$$

$$S(u) = \begin{cases} u_i & \text{if } i \in J \\ 0 & \text{if } i \in I \backslash J \end{cases} \qquad \text{for all } u \in M^{(I)} \text{ with } \mathrm{Supp}u = \{i\}. \tag{117.5}$$

Proof. (i) $\Rightarrow$ (ii). Assume that S satisfies (i). If $u, v \in M^{(I)}$, define $a' \in (M^I)^{2^{\sqsubset}}$ by $a'_0 := u$ and $a'_1 := v$. By (112.10) and Theorem 117E we have

$$S(u + v) = \sum_J (u + v) = \sum_J \sum_{2^{\sqsubset}} a' = \sum_{n \in 2^{\sqsubset}} \sum_J a'_n = \sum_J u + \sum_J v = S(u) + S(v),$$

so that (117.4) holds. Since Supp0 = ∅, (113.4) implies (117.3). If $i \in I$ and $u \in M^{(I)}$ with Supp$u = \{i\}$, then (113.4), (112.7), and (112.9) yield $S(u) = \sum_J u = \sum_{J\cap\{i\}} u$, which is $\sum_{\{i\}} u = u_i$ if $i \in J$, and $\sum_{\emptyset} u = 0$ if $i \in I\backslash J$. Thus (117.5) also holds, and (ii) is proved.

(ii) ⇒ (i). We assume that S satisfies (117.3), (117.4), (117.5), and prove by special induction that

$$P(K) \;:\Leftrightarrow\; (\forall u \in M^{(I)},\;\; \text{Supp}u = K \;\Rightarrow\; S(u) = \sum_J u)$$

holds for all $K \in \mathfrak{F}(I)$; this is sufficient to establish (i). If $u \in M^{(I)}$ and Supp$u = \emptyset$, then $u = 0$, and (117.3) and (113.4) yield $S(u) = S(0) = 0 = \sum_{\emptyset} u = \sum_{J\cap\text{Supp}u} u = \sum_J u$. We have shown that $P(\emptyset)$ holds.

Let $K \in \mathfrak{F}^\times(I)$ be given, and let $k \in K$ be such that $P(K\backslash\{k\})$ holds. Let $u \in M^{(I)}$ with Supp$u = K$ be given. Define $v, w \in M^{(I)}$ by the rules

$$(117.6)\qquad v_i := \begin{cases} u_i & \text{if } i \in I\backslash\{k\} \\ 0 & \text{if } i = k \end{cases} \qquad w_i := \begin{cases} 0 & \text{if } i \in I\backslash\{k\} \\ u_k & \text{if } i = k. \end{cases}$$

Obviously, Supp$v = K\backslash\{k\}$, Supp$w = \{k\}$, and $u = v + w$. By the induction hypothesis and the fact that $k \notin$ Supp$v$ we obtain, using (113.4), that $S(v) = \sum_J v = \sum_{J\backslash\{k\}} v$. Since (117.6) implies $v|_{I\backslash\{k\}} = u|_{I\backslash\{k\}}$, Proposition 113C shows that

$$(117.7)\qquad S(v) = \sum_{J\backslash\{k\}} v = \sum_{J\backslash\{k\}} (v|_{I\backslash\{k\}}) = \sum_{J\backslash\{k\}} (u|_{I\backslash\{k\}}) = \sum_{J\backslash\{k\}} u.$$

Now (117.4), (117.7), (117.5), (117.6) yield

$$S(u) = S(v+w) = S(v) + S(w) = \begin{cases} \sum_{J\backslash\{k\}} u + w_k = \sum_{J\backslash\{k\}} u + u_k = \sum_J u & \text{if } k \in J \\ \sum_J u + 0 = \sum_J u & \text{if } k \notin J. \end{cases}$$

Thus $P(K)$ holds. This completes the induction step. ■

▼ **117G. Remark**. In Theorem 117F, Condition (117.3) may be omitted without affecting the equivalence of (i) and (ii), provided $J \neq I$. Indeed, in this case we may choose $u \in M^{(I)}$ such that $S(u) = 0$, by (117.5) if $M \neq \{0\}$, and trivially otherwise. Then $S(0) = S(0) + 0 = S(0) + S(u) = S(0 + u) = S(u) = 0$.

The proviso $J \neq I$ cannot be omitted in general. There is a commutative monoid $M := \{0, m\}$, with $m \neq 0$ and $m + m = m$. Define $S := m_{M^{(I)} \to M}$; then (117.4) and
▲ (117.5) are satisfied for $J = I$, but (117.3) is not, and $S(0) = m \neq 0 = \sum_I 0$. ■

118. Sums without zero

We shall occasionally encounter structures that involve a binary operation, "addition", that is associative and commutative, such as addition in $\mathbb{N}^\times$, but that have no "zero", or at least none that is prescribed. In such structures it is still possible — indeed interesting — to define sums of finite families, *provided they are not empty.* Instead of developing the theory of these sums independently, as we might, we reduce it to the theory of finite sums in a commutative monoid, as described in the preceding sections. We preserve the additive notation, but the same remarks made before about other terminology and notation remain valid.

We define a **commutative semigroup (written additively)** to be a set M endowed with structure by the prescription of a mapping $((x,y)\mapsto x+y): M\times M\to M$, subject to the conditions (CM1) (associative law) and (CM2) (commutative law) in Section 111.

The device that will enable us to reduce the work with sums in a commutative semigroup to the previous work in commutative monoids is the following construction. Throughout this section we shall assume that it has been carried out.

118A. Construction. Let M be a commutative semigroup, written additively. Choose an object o not contained in M, and set $\bar M := M\cup\{o\}$. We define the mapping $((x,y)\mapsto x\bar+y):\bar M\times\bar M\to\bar M$ by the rule

$$x\bar+y := \begin{cases} x+y & \text{if } x,y\in M \\ x & \text{if } x\in M,\ y=o \\ y & \text{if } x=o,\ y\in M \\ o & \text{if } x=y=o. \end{cases} \tag{118.1}$$

It is easily verified that this mapping satisfies the associative and commutative laws. Thus $\bar M$ is a commutative monoid (written additively) when endowed with structure by the prescription of o as zero and $(x,y)\mapsto x\bar+y$ as addition. Moreover, the commutative monoid $\bar M$ has its zero isolated. We employ the symbol $\bar\sum$ rather than $\sum$ for sums in the commutative monoid $\bar M$.

(This construction is applicable even when there is a member z of M such that $x+z=x$ for all $x\in M$.) ■

118B. Theorem. *Let M be a commutative semigroup, written additively, and let the family $a\in M^I$ be given. Then there is exactly one family $\sigma\in M^{\mathfrak{F}^\times(I)}$ such that*

$$\sigma_{\{i\}} = a_i \qquad \text{for all } i\in I \tag{118.2}$$

$$\sigma_J = \sigma_{J\setminus\{j\}} + a_j \qquad \text{for all } J\in\mathfrak{F}^\times(I)\setminus\mathfrak{F}_1(I) \text{ and } j\in J. \tag{118.3}$$

Proof. 1. Suppose that $\sigma\in M^{\mathfrak{F}^\times(I)}$ satisfies (118.2) and (118.3). Define $\bar\sigma\in\bar M^{\mathfrak{F}(I)}$

by the rule

$$\bar{\sigma}_J := \begin{cases} \sigma_J & \text{if } J \in \mathfrak{P}^\times(I) \\ o & \text{if } J = \emptyset. \end{cases} \tag{118.4}$$

It follows immediately from (118.4), (118.2), (118.3), and (118.1) that

$$\bar{\sigma}_J = \bar{\sigma}_{J\setminus\{j\}} \bar{+} a_j \qquad \text{for all } J \in \mathfrak{F}^\times(I).$$

By Theorem 112A and the definition of *sum* in $\bar{M}$ we conclude that $\bar{\sigma}_J = \bar{\sum_J} a$ for all $J \in \mathfrak{F}(I)$. In particular it follows from (118.4) that

$$\sigma_J = \bar{\sum_J} a \qquad \text{for all } J \in \mathfrak{F}^\times(I). \tag{118.5}$$

There is thus at most one family $\sigma \in M^{\mathfrak{F}^\times(I)}$ that satisfies (118.2) and (118.3).

2. Conversely, since the commutative monoid $\bar{M}$ has its zero isolated, Proposition 112D shows that $\bar{\sum_J} a \in \bar{M}\setminus\{o\} = M$ for all $J \in \mathfrak{F}^\times(I)$. We may therefore *define* $\sigma \in M^{\mathfrak{F}^\times(I)}$ by the formula (118.5). It then follows from (112.8) and (112.9), applied to sums in $\bar{M}$, that σ thus defined satisfies (118.2) and (118.3). ■

Theorem 118B justifies the following definition. For every family $a \in M^I$ we set

$$\sum_J a := \sigma_J \qquad \text{for all } J \in \mathfrak{F}^\times(I),$$

where $\sigma \in M^{\mathfrak{F}^\times(I)}$ is the unique family that satisfies (118.2) and (118.3), so that

$$\sum_{\{i\}} a = a_i \qquad \text{for all } i \in I$$

$$\sum_J a = \sum_{J\setminus\{j\}} a + a_j \qquad \text{for all } J \in \mathfrak{F}^\times(I)\setminus\mathfrak{F}_1(I) \text{ and } j \in J.$$

From (118.5) in the proof of Theorem 118B we obtain

$$\sum_J a = \bar{\sum_J} a \qquad \text{for all } J \in \mathfrak{F}^\times(I). \tag{118.6}$$

Theorem 118B and (118.6) now make all the results of Sections 112 and 114 available to us, in so far as they deal with sums over non-empty finite sets. For the sake of reference,we formulate, in a remark, the conclusions obtained in this manner.

118C. Remark. The following propositions remain valid when the word "monoid" is replaced by "semigroup" and the requirement that J, J', J'' (as the case may require) be non-empty and finite is added: Proposition 112B; Remark 112C,(a); Theorems 114A and 114F; Corollaries 114C, 114D, 114E, and 114G. ■

We also wish to obtain the corresponding analogue of Theorem 117E. This requires a bit of preparation.

Let M be a commutative semigroup, written additively, and let the set I be given. We define the mapping $((a,b) \mapsto a +_I b) : M^I \times M^I \to M^I$ by **termwise addition**, i.e., by the rule

$$(a +_I b)_i := a_i + b_i \qquad \text{for all } i \in I \text{ and } a, b \in M^I.$$

118D. Proposition. *Let M be a commutative semigroup, written additively, and let the set I be given. Then the specification of the mapping $((a,b) \mapsto a +_I b) : M^I \times M^I \to M^I$ as addition endows M^I with the structure of a commutative semigroup, written additively.*

118E. Remark. All references to M^I as a commutative semigroup will be to the structure described in Proposition 118D. We shall write $+$ instead of $+_I$.

Now M^I is a subset of the commutative monoid $\bar{M}^I$, and $a \bar{+} b = a + b$ for all $a, b \in M^I$. *The symbols $\bar{+}$ and $\bar{\sum}$ refer either to $\bar{M}$ or to $\bar{M}^I$, as the context may require, but not to $M^I \cup \{o\}$, in this formula and in what follows.* ■

118F. Lemma. *Let M be a commutative semigroup, written additively, and let the sets I' and I'' be given. Let the family $a' \in (M^{I''})^{I'}$ be given. Then $\sum_{J'} a' = \bar{\sum_{J'}} a'$ for all $J' \in \mathfrak{F}^\times(I')$.*

Proof. By Proposition 117D (applied to the commutative monoids $\bar{M}$ and $\bar{M}^I$) and by (118.6) we have

$$(\bar{\sum_{J'}} a')_{i''} = \bar{\sum_{j' \in J'}} (a'_{j'})_{i''} = \sum_{j' \in J'} (a'_{j'})_{i''} \in M \qquad \text{for all } J' \in \mathfrak{F}^\times(I') \text{ and } i'' \in I''.$$

We conclude that $\bar{\sum_{J'}} a' \in M^{I''}$ for all $J' \in \mathfrak{F}^\times(I')$.

The family $\sigma' \in (M^{I''})^{\mathfrak{F}^\times(I')}$ defined by $\sigma'_{J'} := \bar{\sum_{J'}} a'$ for all $J' \in \mathfrak{F}^\times(I')$ satisfies

$$\sigma'_{\{i'\}} = \bar{\sum_{\{i'\}}} a' = a'_{i'} \qquad \text{for all } i' \in I'$$

$$\sigma'_{J'} = \bar{\sum_{J'}} a' = \bar{\sum_{J' \setminus \{j'\}}} a' \bar{+} a'_{j'} = \sigma'_{J' \setminus \{j'\}} \bar{+} a'_{j'} = \sigma'_{J' \setminus \{j'\}} + a'_{j'}$$
$$\text{for all } J' \in \mathfrak{F}^\times(I') \setminus \mathfrak{F}_1(I') \text{ and } j' \in J'.$$

By Theorem 118B and the definition of sums in $M^{I''}$, we conclude that

$$\sum_{J'} a' = \sigma'_{J'} = \bar{\sum_{J'}} a' \qquad \text{for all } J' \in \mathfrak{F}^\times(I'). \quad ■$$

118G. Theorem. *Let M be a commutative semigroup, written additively, and let the sets I' and I'' be given. Let the family $a' \in (M^{I''})^{I'}$ and the finite non-empty subsets J' of I' and J'' and I'' be given. Then*

$$\sum_{J''} \sum_{J'} a' = \sum_{j' \in J'} \sum_{J''} a'_{j'}.$$

Proof. We use consecutively: (118.6) for the family $\sum_{J'} a' \in M^{I''}$ and the set J''; Lemma 118F; Theorem 117E for the commutative monoid $\bar{M}$; (118.6) for each of the families $a'_{i'}$ with $i' \in I'$, and the set J''; and (118.6) for the family $(\sum_{J''} a'_{i'} \mid i' \in I') \in M^{I'}$ and the set J'. We find

$$\sum_{J''} \sum_{J'} a' = \bar{\sum_{J''}} \bar{\sum_{J'}} a' = \bar{\sum_{J''}} \bar{\sum_{J'}} a' = \bar{\sum_{j' \in J'}} \bar{\sum_{J''}} a'_{j'} = \bar{\sum_{j' \in J'}} \sum_{J''} a'_{j'} = \sum_{j \in J'} \sum_{J''} a'_{j'}. \quad \blacksquare$$

Chapter 12

COUNTABLE SETS

121. Countable sets

A set is said to be **countable** if there is an injection from it to $\mathbb{N}$. All finite sets are countable; a countable infinite set is said to be **countably infinite**. A set that is not countable is said to be **uncountable**. A family is said to be **countable**, **countably infinite**, **uncountable**, according as its index set is countable, countably infinite, uncountable, respectively.

121A. Proposition. *A non-empty set S is countable if and only if there is a surjection from $\mathbb{N}$ to S.*

Proof. If S is countable, we may choose an injection from S to $\mathbb{N}$. Since $S \neq \emptyset$, this injection has a left-inverse, and this is a surjection from $\mathbb{N}$ to S. Assume, conversely, that there is a surjection from $\mathbb{N}$ to S, and choose one, say $g : \mathbb{N} \to S$. We can construct a right-inverse $h : S \to \mathbb{N}$ of g by the rule

$$h(s) := \min g^{<}(\{s\}) \qquad \text{for all } s \in S.$$

Then h is injective, and therefore S is countable. ■

121B. Proposition. *Let the sets S and T and the mapping $f\colon S \to T$ be given. If T is countable and f is injective, then S is countable. If S is countable and f is surjective, then T is countable.*

121C. Proposition. *Every nest of finite sets is countable.*

Proof. Let the nest of finite sets $\mathcal{N}$ be given. By Corollary 101F the mapping $A \mapsto \#A : \mathcal{N} \to \mathbb{N}$ is injective. ■

121D. Theorem. *Let S be a given subset of $\mathbb{N}$. There is exactly one strictly isotone mapping $f\colon \mathbb{N} \to \mathbb{N}$ with $\mathrm{Rng} f = S$ if S is infinite, and none if S is finite. In the former case, $f|^S$ is an order-isomorphism from $\mathbb{N}$ to the ordered subset S of $\mathbb{N}$.*

Proof. Suppose $f\colon \mathbb{N} \to \mathbb{N}$ is a strictly isotone mapping with $\mathrm{Rng} f = S$. Then $f|^S$ is an order-isomorphism from $\mathbb{N}$ to the ordered subset S of $\mathbb{N}$, since $\mathbb{N}$ is totally ordered (Remark 62A,(a) and Proposition 62D). A first consequence is that $\mathbb{N}$ and S are equinumerous; since $\mathbb{N}$ is infinite (Corollary 101K), S must also be infinite. A

second consequence is that

$$f(0) = f|^S(\min \mathbb{N}) = \min S$$

$$f(\mathrm{seq}n) = f|^S(\min\{m \in \mathbb{N} \mid m > n\}) \quad = \min\{s \in S \mid s > f(n)\} = \\ = \min(S\backslash(\mathrm{seq}f(n))^\sqsubset) \qquad \text{for all } n \in \mathbb{N}.$$

Since S is infinite, it has no upper bound in $\mathbb{N}$. It follows that f must be the unique mapping from $\mathbb{N}$ to $\mathbb{N}$ defined recursively by the rules

$$f(0) \quad := \min S$$

$$f(\mathrm{seq}n) \quad := \min(S\backslash(\mathrm{seq}f(n))^\sqsubset) \qquad \text{for all } n \in \mathbb{N}.$$

We now consider an infinite subset S of $\mathbb{N}$ and the mapping $f\colon \mathbb{N} \to \mathbb{N}$ defined recursively by the preceding rules. It is clear that $f(n) < f(\mathrm{seq}n)$ for all $n \in \mathbb{N}$; therefore f is strictly isotone, and also satisfies $n \leq f(n)$ for all $n \in \mathbb{N}$ (Propositions 92G and 92I).

We obviously have $\mathrm{Rng}f \subset S$. To prove the reverse inclusion, let $s \in S$ be given. The set $\{n \in \mathbb{N} \mid f(n) \leq s\}$ is not empty, since it contains 0; and it has an upper bound, since $f(n) \leq s$ implies $n \leq f(n) \leq s$. We may therefore set $p := \max\{n \in \mathbb{N} \mid f(n) \leq s\}$ (Proposition 92F). Then $f(p) \leq s$; if we had $f(p) < s$, it would follow that $s \in S\backslash(\mathrm{seq}f(p))^\sqsubset$, and hence $f(\mathrm{seq}p) \leq s$, contradicting the maximality of p. Therefore $s = f(p) \in \mathrm{Rng}f$. Since $s \in S$ was arbitrary, we conclude that $\mathrm{Rng}f \supset S$. ■

121E. Remark. If S is an infinite (ordered) subset of $\mathbb{N}$, the unique order-isomorphism $g : S \to \mathbb{N}$ is given by the rule

$$g(s) := \#(S \cap s^\sqsubset) \qquad \text{for all } s \in S. \blacksquare$$

121F. Corollary. *A set is countably infinite if and only if it is equinumerous to* $\mathbb{N}$.

Proof. $\mathbb{N}$ is infinite (Corollary 101K) and countable by definition; the "if" part follows immediately. To prove the "only if" part, let S be a countably infinite set and choose an injection $\phi : S \to \mathbb{N}$. Then S is equinumerous to $\phi_>(S)$, an infinite subset of $\mathbb{N}$. By Theorem 121D, $\phi_>(S)$ is equinumerous to $\mathbb{N}$, and consequently S is equinumerous to $\mathbb{N}$ as well. ■

121G. Corollary. *The rule* $f \mapsto \mathrm{Rng}f$ *defines a bijection from the set of all strictly isotone mappings from* $\mathbb{N}$ *to* $\mathbb{N}$*, to the collection of all infinite subsets of* $\mathbb{N}$.

121H. Remark. We recall that a **sequence** is defined as a family with index set $\mathbb{N}$ or $\mathbb{N}^\times$; for our present purpose we shall restrict attention to the index set $\mathbb{N}$. If s, t are sequences, t is called a **subsequence of** s if $t = s \circ \sigma$ for some strictly isotone mapping $\sigma : \mathbb{N} \to \mathbb{N}$. Theorem 121D and Corollary 121G are useful in mediating between subsequences of a given sequence s and restrictions of s to infinite subsets of $\mathbb{N}$, thus relating a subsequence $s \circ \sigma$ to the restriction $s|_{\mathrm{Rng}\sigma}$ for every strictly isotone $\sigma : \mathbb{N} \to \mathbb{N}$. *Warning*: Unless s is injective, the same subsequence may correspond

to restrictions of s to distinct infinite subsets of $\mathbb{N}$. This concern need not, however, constitute an impediment, e.g., when the issue is the *existence* of a subsequence with some desired property, as illustrated by the following useful result. ■

121I. Proposition. *Let the ordered set D and the isotone sequence $s \in D^{\mathbb{N}}$ be given. The following statements are equivalent:*

(i): Rng$s$ *is finite.*

(ii): Rng$s$ *has a maximum.*

(iii): s *is eventually constant; i.e., $s|_{m+\mathbb{N}}$ is constant for some $m \in \mathbb{N}$.*

(iv): s *has a constant subsequence.*

(v): s *has no strictly isotone subsequence.*

Proof. Since s is isotone, Rng$s$ is a totally ordered subset (a chain) of D (and is of course not empty).

(i) ⇒ (ii). This follows from Corollary 105B.

(ii) ⇒ (iii). Set $a := \max \text{Rng}s$. Choose $m \in \mathbb{N}$ such that $s_m = a$. Since s is isotone, we have $a = s_m \prec s_n \prec a$ for all $n \in m + \mathbb{N}$; thus $s_>(m + \mathbb{N}) = \{a\}$.

(iii) ⇒ (i). Choose $m \in \mathbb{N}$ such that $s|_{m+\mathbb{N}}$ is constant, and set $a :\in s_>(m + \mathbb{N})$. Then

$$\text{Rng}s = s_>(m^\sqsubset \cup (m + \mathbb{N})) = s_>(m^\sqsubset) \cup s_>(m + \mathbb{N}) = s_>(m^\sqsubset) \cup \{a\};$$

thus Rng$s$ is finite.

(iii) ⇒ (iv). Choose $m \in \mathbb{N}$ such that $s|_{m+\mathbb{N}}$ is constant, and define $\sigma : \mathbb{N} \to \mathbb{N}$ by $\sigma(k) := m + k$ for all $k \in \mathbb{N}$. Then σ is strictly isotone and the subsequence $s \circ \sigma$ of s is constant.

(iv) ⇒ (v). Suppose, by contradiction, that s has a strictly isotone subsequence in addition to a constant subsequence. By Theorem 121D and Corollary 121G we may choose infinite (ordered) subsets A and B of $\mathbb{N}$ such that $s|_A$ is constant and $s|_B$ is strictly isotone. Choose $a \in A$. Since neither A nor B has an upper bound (Corollary 105D), we may choose $b, c \in B$ and $d \in A$ such that $a < b < c < d$. Since s is isotone, this implies $s_a \prec s_b \precneqq s_c \prec s_d$, and this contradicts $s_a = s_d$.

(v) ⇒ (ii). Assume, by contraposition, that Rng$s$ has not a maximum; hence Rng$s$, which is totally ordered, has no maximal members. It follows that $\text{Ub}_{\text{Rng}s}(\{s_n\}) \setminus \{s_n\} \neq \emptyset$ for all $n \in \mathbb{N}$. We may therefore define the mapping $\phi : \mathbb{N} \to \mathbb{N}$ by

$$\phi(n) := \min s^<(\text{Ub}_{\text{Rng}s}(\{s_n\}) \setminus \{s_n\}) \qquad \text{for all } n \in \mathbb{N}.$$

Then $s_n \precneqq s_{\phi(n)}$ for all $n \in \mathbb{N}$; since s is isotone, this implies $n < \phi(n)$ for all $n \in \mathbb{N}$. We define the mapping $\sigma : \mathbb{N} \to \mathbb{N}$ by $\sigma(k) := \phi^{\circ k}(0)$ for all $k \in \mathbb{N}$. By Proposition 92G, σ is strictly isotone, and $s_{\sigma(k)} \precneqq s_{\phi(\sigma(k))} = s_{\sigma(\text{seq}k)}$ for all $k \in \mathbb{N}$. Again by Proposition 92G, the subsequence $s \circ \sigma$ of s is strictly isotone. ■

There are several results that assert that some sets constructed from countable sets are themselves countable. We derive them all from the following theorem, which has intrinsic interest in that it justifies the binary digital system of numeration: it states that every natural number has a unique binary representation, which is

determined by the set of places (counting from the right) that carry the digit 1. A rather tedious technical modification of the proof would be required to give the theoretical justification for the digital system of numeration to any given base, such as the familiar seq seq seq seq seq seq seq seq seq seq 0.

121J. Theorem. (Binary Numeration Theorem). *The mapping* $(A \mapsto \sum_{m\in A} 2^m) : \mathfrak{F}(\mathbb{N}) \to \mathbb{N}$ *is bijective. Consequently,* $\mathfrak{F}(\mathbb{N})$ *is countably infinite.*

Proof. 1. We first record the following result, which is easily proved by induction

$$\sum_{m\in n^\sqsubset} 2^m = 2^n - 1 \qquad \text{for all } n \in \mathbb{N}. \tag{121.1}$$

The induction step uses the formula $(\mathrm{seq}n)^\sqsubset \backslash \{n\} = n^\sqsubset$ and runs as follows:

$$\sum_{m\in(\mathrm{seq}n)^\sqsubset} 2^m = \sum_{m\in n^\sqsubset} 2^m + 2^n = 2^n - 1 + 2^n = 2^n + 2^n - 1 = 2^n \cdot 2 - 1 = 2^{\mathrm{seq}n} - 1.$$

We next consider two given finite subsets S and T of $\mathbb{N}$ such that $S \cap T = \emptyset$ and $S \cup T \neq \emptyset$. Set $p := \max(S \cup T)$. Then either $p \in S$ and $T \subset p^\sqsubset$, or else $p \in T$ and $S \subset p^\sqsubset$. In the former case we have, using Corollary 114C and (121.1),

$$\sum_{m\in T} 2^m \leq \sum_{m\in T} 2^m + \sum_{m\in p^\sqsubset\backslash T} 2^m = \sum_{m\in p^\sqsubset} 2^m = 2^p - 1 < 2^p \leq \sum_{m\in S\backslash\{p\}} 2^m + 2^p = \sum_{m\in S} 2^m;$$

in the latter case we find, interchanging S and T in this argument, that $\sum_{m\in S} 2^m < \sum_{m\in T} 2^m$. We have thus proved the following assertion:

$$\forall S, T \in \mathfrak{F}(\mathbb{N}), \quad (S \cap T = \emptyset \text{ and } S \cup T \neq \emptyset) \;\Rightarrow\; \sum_{m\in S} 2^m \neq \sum_{m\in T} 2^m. \tag{121.2}$$

2. Let $A, B \in \mathfrak{F}(\mathbb{N})$ be given, and suppose that $\sum_{m\in A} 2^m = \sum_{\in B} 2^m$. Set $S := A\backslash B$ and $T := B\backslash A$. Then $S \cap T = \emptyset$ and, by Corollary 114C,

$$\sum_{m\in S} 2^m + \sum_{m\in A\cap B} 2^m = \sum_{m\in A} 2^m = \sum_{m\in B} 2^m = \sum_{m\in T} 2^m + \sum_{m\in A\cap B} 2^m.$$

By the cancellation law, we have $\sum_{m\in S} 2^m = \sum_{m\in T} 2^m$. By (121.2), this implies $S \cup T = \emptyset$, hence $S = T = \emptyset$, and thus $A = B$. Since $A, B \in \mathfrak{F}(\mathbb{N})$ were arbitrary, we have proved that the mapping $(A \mapsto \sum_{m\in A} 2^m) : \mathfrak{F}(\mathbb{N}) \to \mathbb{N}$ is injective.

3. To prove that this mapping is surjective, we prove by induction that $P(n)$ holds for all $n \in \mathbb{N}$, where

$$P(n) \;:\Leftrightarrow\; (\exists A \in \mathfrak{F}(\mathbb{N}),\ \sum_{m\in A} 2^m = n).$$

Since $\sum_{m\in\emptyset} 2^m = 0$, we see that $P(0)$ holds. Let $n \in \mathbb{N}$ be given, and assume that $P(n)$ holds. We may therefore choose $A \in \mathfrak{F}(\mathbb{N})$ such that $\sum_{m\in A} 2^m = n$. Set $p := \min(\mathbb{N}\backslash A)$. Then $p^\sqsubset \subset A$ and $p \notin A\backslash p^\sqsubset$. We find, using Corollary 114C and (121.1),

$$\mathrm{seq}n = n+1 = \sum_{m\in A} 2^m + 1 \;= \sum_{m\in A\backslash p^\sqsubset} 2^m + \sum_{m\in p^\sqsubset} 2^m + 1 = \sum_{m\in A\backslash p^\sqsubset} 2^m + 2^p - 1 + 1 =$$

$$= \sum_{m\in A\backslash p^\sqsubset} 2^m + 2^p = \sum_{m\in (A\backslash p^\sqsubset)\cup\{p\}} 2^m.$$

Hence $P(\mathrm{seq}n)$ also holds. This completes the induction step. ■

121K. Remark. The proof of Theorem 121J shows that the unique finite subset A_n of $\mathbb{N}$ that satisfies $\sum_{m\in A_n} 2^m = n$ for each $n \in \mathbb{N}$ can be obtained recursively by the following *digit-carrying rules:*

$$A_0 := \emptyset$$

$$A_{\mathrm{seq}n} := (A_n\backslash p^\sqsubset) \cup \{p\}, \text{ where } p := \min(\mathbb{N}\backslash A_n), \text{ for all } n \in \mathbb{N}. \blacksquare$$

121L. Corollary. $\mathbb{N}\times\mathbb{N}$ *is countably infinite.*

Proof. The mapping $(n \mapsto (n,n)) : \mathbb{N} \to \mathbb{N}\times\mathbb{N}$ is injective; hence $\mathbb{N}\times\mathbb{N}$ is infinite (Proposition 101E). Each of the mappings

$$((m,n) \mapsto \{m, m+n+1\}) : \mathbb{N}\times\mathbb{N} \to \mathfrak{F}_2(\mathbb{N})$$

$$(D \mapsto (\min D, \max D - \min D - 1)) : \mathfrak{F}_2(\mathbb{N}) \to \mathbb{N}\times\mathbb{N}$$

is the inverse of the other. Since $\mathfrak{F}_2(\mathbb{N}) \subset \mathfrak{F}(\mathbb{N})$, Theorem 121J and Proposition 121B show that $\mathbb{N}\times\mathbb{N}$ is countable. ■

121M. Corollary. *If S is a countable set, then $\mathfrak{F}(S)$ is again countable. If S and T are countable sets, then $S\times T$ is again countable.*

•121N. Corollary. *Let the countable family of countable sets $(T_i \mid i \in I)$ be given. The sets $\dot{\bigcup_{i\in I}} T_i$ and $\bigcup_{i\in I} T_i$ are countable.*

Proof. Choose an injection $\phi : I \to \mathbb{N}$, and •choose an injection $\psi_i : T_i \to \mathbb{N}$ for every $i \in I$. Then the mapping $((i,x) \mapsto (\phi(i), \psi_i(x))) : \dot{\bigcup_{i\in I}} T_i \to \mathbb{N}\times\mathbb{N}$ is injective. By Corollary 121L and Proposition 121B it follows that $\dot{\bigcup_{i\in I}} T_i$ is countable. The mapping $((i,x) \mapsto x) : \dot{\bigcup_{i\in I}} T_i \to \bigcup_{i\in I} T_i$ is surjective. The countability of $\bigcup_{i\in I} T_i$ now follows from Proposition 121B. ■

•121O. Corollary. *The union of every countable collection of countable sets is countable.*

•**121P. Corollary**. *A set is countable •if and only if it is the union of a nest of finite sets.*

Proof. A nest of finite sets is countable, by Proposition 121C; its union is therefore countable by •Corollary 121O. Conversely, let a countable set S be given. If S is finite, then S is the union of the nest of finite sets $\{S\}$. If S is countably infinite, we may choose, by Corollary 121F, a bijection $\phi : \mathbb{N} \to S$; then S is the union of the nest of finite sets $\{\phi_>(n^\sqsubset) \mid n \in \mathbb{N}\}$. ■

▼ **121Q. Remarks**. (a): The proof of •Corollary 121N (and consequently of •Corollaries 121O and 121P requires the choice of the family of mappings $(\psi_i \mid i \in I)$ from the product $\bigtimes_{i\in I} \mathrm{Inj}(T_i, \mathbb{N})$, where $\mathrm{Inj}(T_i, \mathbb{N}) := \{f \in \mathrm{Map}(T_i, \mathbb{N}) \mid f \text{ is injective}\} \neq \emptyset$ by assumption for each $i \in I$. The possibility of this simultaneous choice depends on the •Axiom of Choice, or rather since I is countable, on a weaker version, the •"Axiom of Countable Choice": *If $(A_n \mid n \in \mathbb{N})$ is a sequence of non-empty sets, then $\bigtimes_{n\in\mathbb{N}} A_n \neq \emptyset$.*

(b): If I is finite in Corollary 121N, then the Principle of Finite Choice (Theorem 103L) makes recourse to any such axiom unnecessary. The same is consequently true ▲ for Corollary 121O, if the collection of countable sets is finite. ■

121R. Example*. There are obvious surjections from the set $\mathbb{N}^\times \times \mathbb{N}^\times$ to the sets of strictly positive and of strictly negative rational numbers (the former surjection is $(m, n) \mapsto m/n$). By Corollaries 121M and 121O (using Remark 121Q,(b)) and Proposition 121B it follows that the set of all rational numbers is countable. ■

121S. Theorem. *Let the finite family of countable sets $(T_i \mid i \in I)$ be given. Then $\bigtimes_{i\in I} T_i$ is a countable set.*

Proof. The proof is precisely like that of Theorem 103H, using Corollary 121L instead of Corollary 103G. ■

121T. Corollary. *If S is a finite set and T a countable set, then the equinumerous sets T^S and $\mathrm{Map}(S, T)$ are countable.*

We finally show that the countable sets are "smaller" than all infinite sets. We begin with a reformulation of part of Proposition 101N.

121U. Lemma. *Let the set S be given. The following statements are equivalent:*

(i): *S is infinite.*

(ii): *$\mathfrak{F}_n(S) \neq \emptyset$ for all $n \in \mathbb{N}$.*

(iii): *the sequence $(\mathfrak{F}_n(S) \mid n \in \mathbb{N})$ in $\mathfrak{P}(\mathfrak{P}(S))$ is injective.*

Proof. The sequence $(\mathfrak{F}_n(S) \mid n \in \mathbb{N})$ is obviously disjoint. The asserted equivalence then follows from Proposition 101N. ■

•**121V. Theorem**. *Let the infinite set S be given. There exists an injective sequence $a \in S^{\mathbb{N}}$.*

Proof. It is instructive to give two proofs of this theorem. The first of these is more intuitive, but uses recursive choice on the index set $\mathbb{N}$ (in its general form), and depends on the full force of the •Axiom of Choice. The second proof uses no

recursion, and depends only on the •Axiom of Countable Choice.

Proof I. We •choose recursively a sequence $a \in S^{\mathbb{N}}$ satisfying

$$a_n \in S\backslash\mathrm{Rng}(a|_{n^\sqsubset}) \qquad \text{for all } n \in \mathbb{N}. \tag{121.3}$$

For all $m, n \in \mathbb{N}$ with $m < n$ we have $m \in n^\sqsubset$ and therefore $a_m \in \mathrm{Rng}(a|_{n^\sqsubset})$. Comparison with (121.3) shows that $a_m \neq a_n$. Thus the sequence a is injective.

To justify the recursive choice of a we apply •Theorem 82H with $I := \mathbb{N}$, $A_n := S$ for all $n \in \mathbb{N}$, and $\Phi_n : S^{n^\sqsubset} \to \mathfrak{P}(S)$ defined by the rule

$$\Phi_n(u) := S\backslash\mathrm{Rng}u \qquad \text{for all } n \in \mathbb{N} \text{ and } u \in S^{n^\sqsubset}.$$

Since $u \in S^{n^\sqsubset}$ is a finite family, its range is finite, and therefore $\Phi_n(u) \neq \emptyset$ for all $n \in \mathbb{N}$ and $u \in S^{n^\sqsubset}$. This makes the mentioned theorem applicable.

Proof II. By Lemma 121U,((i)⇒(ii)), we may •choose a sequence $(A_n \mid n \in \mathbb{N})$ in $\mathfrak{P}(S)$ such that $\#A_n = 2^n$ for all $n \in \mathbb{N}$. We now consider the sequence $(B_n \mid n \in \mathbb{N})$ in $\mathfrak{P}(S)$ defined by the rule

$$B_n := A_n \backslash \bigcup_{m \in n^\sqsubset} A_m \qquad \text{for all } n \in \mathbb{N}.$$

If $n, n' \in \mathbb{N}$ satisfy $n < n'$, then $B_n \subset A_n \subset \bigcup_{m \in n'^\sqsubset} A_m$ and $B_{n'} \subset A_{n'}\backslash B_n$. Therefore the sequence $(B_n \mid n \in \mathbb{N})$ is disjoint. Moreover, using Corollary 103E and (121.1), we find

$$\#\bigcup_{m \in n^\sqsubset} A_m \leq \sum_{m \in n^\sqsubset} \#A_m = \sum_{m \in n^\sqsubset} 2^m = 2^n - 1 < 2^n = \#A_n \qquad \text{for all } n \in \mathbb{N}.$$

Therefore $B_n \neq \emptyset$ for every $n \in \mathbb{N}$, by Corollary 101F. We may therefore •choose the sequence $a \in \bigtimes_{n \in \mathbb{N}} B_n \subset S^{\mathbb{N}}$, and this sequence is injective because the sequence $(B_n \mid n \in \mathbb{N})$ was disjoint. ■

•121W. Corollary. *If T is a countable set and S is an infinite set, then there exists an injection from T to S.*

122. Some uncountable sets

In contrast to Theorem 121J and its corollaries, in which certain sets were shown to be countable, we establish in this section that certain infinite sets cannot be countable.

122A. Theorem. *The power-set of every infinite set is uncountable.*

Proof. Suppose that S were an infinite set with countable power-set $\mathfrak{P}(S)$. Since the mapping $s \mapsto \{s\} : S \to \mathfrak{P}(S)$ is injective, S would be countable and $\mathfrak{P}(S)$ infinite (Propositions 121B and 101E). The sets S and $\mathfrak{P}(S)$, both being countably infinite and hence equinumerous to $\mathbb{N}$ (Corollary 121F), would have to be equinumerous to each other. But Proposition 32E shows that there is no surjective mapping from S to $\mathfrak{P}(S)$, let alone an invertible one. ■

•122B. Corollary. *Let the family of countable non-empty sets $(T_i \mid i \in I)$ be given. Then $\bigtimes_{i\in I} T_i$ is countable if and only if $\{i \in I \mid T_i \text{ is not a singleton}\}$ is finite.*

Proof. Set $J := \{i \in I \mid T_i \text{ is not a singleton}\}$. It is clear that $\bigtimes_{i\in I} T_i$ and $\bigtimes_{i\in J} T_i$ are equinumerous. If J is finite, then Theorem 121S shows that $\bigtimes_{i\in J} T_i$ is countable. Suppose, conversely, that $\bigtimes_{i\in J} T_i$ is countable. We •choose an injection $\lambda_i : 2^{\sqsubset} \to T_i$ for each $i \in J$, as we may, since T_i is not a singleton. Then the mapping $(\alpha \mapsto (\lambda_i(\alpha_i) \mid i \in J)) : (2^{\sqsubset})^J \to \bigtimes_{i\in J} T_i$ is injective. But $(2^{\sqsubset})^J$ is equinumerous to $\mathfrak{P}(J)$ (see proof of Corollary 103J). By Proposition 121B, $\mathfrak{P}(J)$ is countable; by Theorem 122A we conclude that J is finite. ■

122C. Proposition. *Let the countably infinite set S be given. Then $\mathfrak{P}(S)$ and $\mathfrak{P}(S)\backslash\mathfrak{F}(S)$ are equinumerous.*

Proof. Choose a bijection $\phi : S \to \mathbb{N}$. Then $\phi_> : \mathfrak{P}(S) \to \mathfrak{P}(\mathbb{N})$ is bijective and induces a bijection from $\mathfrak{P}(S)\backslash\mathfrak{F}(S)$ to $\mathfrak{P}(\mathbb{N})\backslash\mathfrak{F}(\mathbb{N})$. It is therefore sufficient to prove that $\mathfrak{P}(\mathbb{N})$ and $\mathfrak{P}(\mathbb{N})\backslash\mathfrak{F}(\mathbb{N})$ are equinumerous.

We define the mapping $f\colon \mathfrak{P}(\mathbb{N}) \to \mathfrak{P}(\mathbb{N})\backslash\mathfrak{F}(\mathbb{N})$ by the rule

$$f(A) := \begin{cases} 2A & \text{if } A \in \mathfrak{P}(\mathbb{N})\backslash\mathfrak{F}(\mathbb{N}) \\ \\ 2(\mathbb{N}\backslash A) + 1 & \text{if } A \in \mathfrak{F}(\mathbb{N}) \end{cases} \qquad A \in \mathfrak{P}(\mathbb{N}).$$

It is not difficult to see that f is injective. It then follows from the Schröder-Bernstein Theorem (Theorem 75C) that $\mathfrak{P}(\mathbb{N})$ and $\mathfrak{P}(\mathbb{N})\backslash\mathfrak{F}(\mathbb{N})$ are equinumerous. ■

The following result is a generalization of Theorem 122A (which is the special case $\mathcal{M} := \mathfrak{P}(S)$).

122D. Theorem. *Let the set S and the collection $\mathcal{M}$ of subsets of S be given. Assume that $\mathcal{M}$ is intersection-stable and satisfies*

$$\forall A \in \mathcal{M}, \quad S\backslash A \in \mathcal{M}. \tag{122.1}$$

Then $\mathcal{M}$ is either finite or uncountable.

Proof. 1. Assume that $\mathcal{M}$ is countable. Define the subcollections $\mathcal{M}_x := \{A \in \mathcal{M} \mid x \in A\}$ for every $x \in S$ (note that $S \in \mathcal{M}_x$ for every $x \in S$), and the members $K_x := \bigcap \mathcal{M}_x$ of $\mathcal{M}$ for every $x \in S$. Thus $\mathcal{D} := \{K_x \mid x \in S\}$ is a subcollection of $\mathcal{M}$.

2. We claim that $\mathcal{D}$ is disjoint. Let $x, y \in S$ be given, and assume that $K_x \neq K_y$. Then $\mathcal{M}_x \neq \mathcal{M}_y$, and hence we may choose $A \in \mathcal{M}_x \backslash \mathcal{M}_y$ or $A \in \mathcal{M}_y \backslash \mathcal{M}_x$. Consider the former case. We have $y \notin A$, so that $y \in S \backslash A$. On the other hand, $S \backslash A \in \mathcal{M}$, and we conclude that $S \backslash A \in \mathcal{M}_y$. Thus $K_x \subset A$ and $K_y \subset S \backslash A$, and therefore $K_x \cap K_y = \emptyset$. The latter case ($A \in \mathcal{M}_y \backslash \mathcal{M}_x$) yields the same conclusion (the proof shows that the two cases are indeed one). We have proved that

$$\forall x, y \in S, \quad K_x \neq K_y \;\Rightarrow\; K_x \cap K_y = \emptyset,$$

i.e., that $\mathcal{D}$ is disjoint. Since $x \in K_x$ for every $x \in S$, we have $\emptyset \notin \mathcal{D}$.

3. Since $\mathcal{M}$ is intersection-stable and satisfies (122.1), it follows that the union of every subcollection of $\mathcal{D}$ is a member of $\mathcal{M}$. Since $\mathcal{D}$ is *disjoint* and $\emptyset \notin \mathcal{D}$, the mapping

$$\mathcal{C} \mapsto \bigcup \mathcal{C} : \mathfrak{P}(\mathcal{D}) \to \mathcal{M} \tag{122.2}$$

is *injective*. It follows that $\mathfrak{P}(\mathcal{D})$ is countable; by Theorem 122A, $\mathcal{D}$ is *finite*.

4. Let $A \in \mathcal{M}$ be given. For every $x \in A$ we have $A \in \mathcal{M}_x$, and hence $x \in K_x \subset A$. We conclude that $A = \bigcup \{K_x \mid x \in A\}$. It follows that the mapping (122.2) is *surjective*. Since $\mathcal{D}$ is finite, so is $\mathfrak{P}(\mathcal{D})$ (Corollary 103J), and hence $\mathcal{M}$ is also finite. ∎

122E. Remark. Let the set S be given. A collection $\mathcal{M}$ of subsets of S is called a **σ-algebra on** S if $\mathcal{M}$ satisfies (122.1) and the union (and hence also the intersection with respect to S) of every *countable* subcollection of $\mathcal{M}$ is a member of $\mathcal{M}$. Theorem 122D shows that a σ-algebra on S cannot be countably infinite for any set S. ∎

The following results show that certain ordered sets cannot be countably infinite.

122F. Proposition. *A densely and completely ordered set is countable (if and) only if it is a singleton.*

Proof. Let the countable set D, densely and completely ordered by $\prec$, be given. Then $D \neq \emptyset$, and we may choose a surjection $\phi : \mathbb{N} \to D$ (Proposition 121A). We define the mapping f: $\mathrm{Gr}(\precneqq) \to D$ by the rule

$$f((x,y)) := \phi(\min \phi^{<}(]\!]x, y[\![)) \qquad \text{for all } (x,y) \in \mathrm{Gr}(\precneqq),$$

as we may, since the order is dense. We note that

$$x \precneqq f((x,y)) \precneqq y \;\text{ and }\; x \precneqq f((x,y)) \precneqq f((f((x,y)),y)) \precneqq y \quad \text{for all } (x,y) \in \mathrm{Gr}(\precneqq). \tag{122.3}$$

We suppose that D is not a singleton, so that $\min D \precneqq \max D$. By (122.3) we may define the sequence $((u_n, v_n) \mid n \in \mathbb{N})$ in $\mathrm{Gr}(\precneqq)$ recursively by the rule

$$(u_0, v_0) := (\min D, \max D)$$
$$(u_{n+1}, v_{n+1}) := (f((u_n, v_n)), f((f((u_n, v_n)), v_n))) \qquad \text{for all } n \in \mathbb{N}.$$

By (122.3) we have

$$u_n \precneqq u_{n+1} \precneqq v_{n+1} \precneqq v_n \qquad \text{for all } n \in \mathbb{N}. \tag{122.4}$$

It follows by Proposition 92G that the sequence $(u_n \mid n \in \mathbb{N})$ is strictly isotone and the sequence $(v_n \mid n \in \mathbb{N})$ is strictly antitone. Consequently,

$$u_m \prec u_{\max\{m,n\}} \precneqq v_{\max\{m,n\}} \prec v_n \qquad \text{for all } m, n \in \mathbb{N}. \tag{122.5}$$

We observe that for all $n \in \mathbb{N}$ we have $\phi^<(]\!]u_n, v_n[\![) \supset \phi^<(]\!]u_{n+1}, v_{n+1}[\![)$, but $\phi(\min \phi^<(]\!]u_n, v_n[\![)) = f((u_n, v_n)) = u_{n+1} \notin]\!]u_{n+1}, v_{n+1}[\![$. Therefore $\min \phi^<(]\!]u_n, v_n[\![) < \min\phi^<(]\!]u_{n+1}, v_{n+1}[\![)$ for all $n \in \mathbb{N}$, and hence, by Propositions 92G and 92I,

$$n \leq \min \phi^<(]\!]u_n, v_n[\![) \qquad \text{for all } n \in \mathbb{N}. \tag{122.6}$$

Since D is completely ordered, we may set $s := \sup\{u_n \mid n \in \mathbb{N}\}$. By (122.5), s is a lower bound of $\{v_n \mid n \in \mathbb{N}\}$. By (122.4) it follows that

$$u_n \precneqq u_{n+1} \prec s \prec v_{n+1} \precneqq v_n \qquad \text{for all } n \in \mathbb{N},$$

so that $s \in]\!]u_n, v_n[\![$, and consequently $\emptyset \neq \phi^<(\{s\}) \subset \phi^<(]\!]u_n, v_n[\![)$, for all $n \in \mathbb{N}$. Therefore

$$\min \phi^<(\{s\}) \geq \min \phi^<(]\!]u_n, v_n[\![) \qquad \text{for all } n \in \mathbb{N},$$

which contradicts (122.6). The supposition that D was not a singleton is therefore untenable. ■

122G. Corollary. *A densely and pre-completely ordered set is countable (if and) only if the order is the relation of equality in the set. A totally, densely, and pre-completely ordered set is countable (if and) only if it is either empty or a singleton.*

Proof. Let the countable set D, densely and pre-completely ordered by $\prec$, be given. Let $x, y \in D$ be given, and assume that $x \prec y$. Then the ordered subset $[\![x, y]\!]$ of D is countable and is densely and completely ordered (Proposition 72E,(b)). By Proposition 122F, $[\![x, y]\!]$ is a singleton, so that $x = y$. The relation $\prec$ is therefore narrower than $=_D$; being reflexive, it is also broader that $=_D$. The two relations are therefore equal. ■

122H. Example*. The set $\mathbb{R}$ is totally and pre-completely ordered by $\leq$, and this order is dense (see Propositions 143B and 151B). By Corollary 122G, the set $\mathbb{R}$ is uncountable. The argument used to prove Proposition 122F and hence the uncountability of $\mathbb{R}$ was first used in 1874 by Georg Ferdinand Ludwig Philipp Cantor (1845-1913). ■

123. Another characterization of finiteness

▼ The Pigeonhole Principle (Corollary 101H) shows that every injection from a finite set to itself is invertible. In contrast to this, the injection seq : $\mathbb{N} \to \mathbb{N}$ is not invertible. Using the •Axiom of Countable Choice, it is possible to generalize this contrast and obtain a characterization of finiteness (•Theorem 123B). This characteristic condition has sometimes been used (notably by Julius Wilhelm Richard Dedekind (1831-1916)) to *define* the concept of finite set. To obtain this characterization we use the following lemma, which does not require the cited •Axiom.

123A. Lemma. *Let the set S be given. The following statements are equivalent:*

(i): *There exists an injective sequence in S.*

(ii): *There exists an injection from S to S that is not invertible.*

Proof. (i) *implies* (ii). Choose an injective sequence $a \in S^{\mathbb{N}}$. Then a, considered as a surjection $a : \mathbb{N} \to \mathrm{Rng}a$, is invertible. We define the mapping $f\colon S \to S$ by the rule

$$f(s) := \begin{cases} a_{\mathrm{seq}n} & \text{if } s = a_n \qquad \text{for all } n \in \mathbb{N} \\ \\ s & \text{if } s \in S\backslash\mathrm{Rng}a. \end{cases}$$

We observe that $f_{>}(\mathrm{Rng}a) \subset \mathrm{Rng}a$ and $f_{>}(S\backslash\mathrm{Rng}a) = S\backslash\mathrm{Rng}a$. Since a, seq, and $a^{\leftarrow}$ are all injective, it follows that f is injective. But

$$\mathrm{Rng}f \subset a_{>}(\mathrm{Rng\ seq}) \cup (S\backslash\mathrm{Rng}a) = (\mathrm{Rng}a\backslash\{a_0\}) \cup (S\backslash\mathrm{Rng}a) = S\backslash\{a_0\},$$

since a is injective and $\mathrm{Rng\ seq} = \mathbb{N}\backslash\{0\}$. Therefore f is not surjective, hence not invertible.

(ii) *implies* (i). Choose an injection $f\colon S \to S$ that is not invertible, hence not surjective. Choose $u \in S\backslash\mathrm{Rng}f$, and set $a := (f^{\circ n}(u) \mid n \in \mathbb{N}) \in S^{\mathbb{N}}$. We claim that the sequence a is injective. Let $m, n \in \mathbb{N}$ be given, and assume that $m < n$. Then $f^{\circ(n-m)}(u) \in \mathrm{Rng}f$, and hence $u \neq f^{\circ(n-m)}(u)$. But $f^{\circ m}$ is injective (Proposition 94E); therefore $f^{\circ m}(u) \neq f^{\circ m}(f^{\circ(n-m)}(u)) = f^{\circ n}(u)$ (Proposition 96B). ■

•123B. Theorem. *A set S is finite if and only if every injection from S to S is invertible.*

Proof. The "only if" part follows from Corollary 101H; the "if" part follows, by contraposition, from •Theorem 121V and Lemma 123A,((i) ⇒ (ii)). ■

•123C. Corollary. *A set S is finite if and only if every surjection from S to S is invertible.*

Proof. The "only if" part follows from Corollary 101H. To prove the "if" part, let the set S be given and assume that every surjection from S to S is invertible. Now every injection from S to S has a left-inverse, and this is surjective; hence this left-inverse is invertible, and therefore so is the injection itself. From •Theorem 123B it follows that S is finite. ■

123D. Remark. Replacing, in the proof of •Theorem 123B, the use of •Theorem 121V by that of Lemma 121U,((ii) ⇒ (iii)), and combining this with Corollaries

101H and 103J, one obtains the following conclusion, which no longer depends on the •Axiom of Countable Choice: *A set S is finite if and only if every injection from* ▲ $\mathfrak{P}(\mathfrak{P}(S))$ *to* $\mathfrak{P}(\mathfrak{P}(S))$ *is invertible.* ■

Chapter 13

SOME ALGEBRAIC STRUCTURES

131. Commutative monoids and groups

In this chapter we shall discuss certain structures consisting of "algebraic operations" defined on sets. We have already introduced some structures of this kind in Chapter 11: commutative monoids (Section 111) and commutative semigroups (Section 118). We shall not engage at this point in a thorough study of such *algebraic structures*, but shall develop only so much of the terminology, notation, and "bookkeeping" properties as is needed as background for the account of the Real-Number System in the next three chapters; a second purpose, achieved at no additional expense, is to provide information required for use in linear algebra. We shall require, from Section 111, the definition of *commutative monoid*, as well as the terminology, notation, and results in Sections 111-115 and 117.

We record some notational conventions used in connection with commutative monoids. Let the commutative monoid M, written additively, be given. By virtue of the associative law we may write expressions such as $x + y + z$ for all $x, y, z \in M$, omitting parentheses without danger of ambiguity. Similarly, we may write mnx for all $x \in M$ and $m, n \in \mathbb{N}$, by virtue of Proposition 115B,(b).

For all subsets $A,\ B$ of M and all subsets S of $\mathbb{N}$ we write

$$A + B := \{x + y \mid (x, y) \in A \times B\}$$

$$SA := \{nx \mid (n, x) \in S \times A\}$$

$$A^{\times} := A \backslash \{0\}.$$

In the first and second formulas, braces are often omitted when exactly one of the sets is recorded as being a singleton: thus, with $a, b \in M$ and $n \in \mathbb{N}$ we write $a + B := \{a\} + B,\ A + b := A + \{b\},\ nA := \{n\}A,\ Sa := S\{a\}$.

In multiplicative notation, we use xyz unambiguously without parentheses, and

set

$$AB := \{xy \mid (x,y) \in A \times B\}$$

and $aB := \{a\}B$ and $Ab := A\{b\}$; but no multiplicative analogues of SA and $A^\times$ are in use. (In particular, the notation A^S is reserved for other uses: see Section 44.)

Returning to additive notation, we generalize the notation $A + B$ as follows. Let the family $(A_i \mid i \in I)$ of subsets of M be given. We define

$$\sum_{i\in I} A_i := \{\sum_{i\in I} a_i \mid a \in M^{(I)} \cap \bigtimes_{i\in I} A_i\} \subset M$$

(we recall that $M^{(I)} := \{a \in M^I \mid \mathrm{Supp}a \text{ is finite}\}$). If, in particular, I is finite, we have

$$\sum_{i\in I} A_i = \{\sum_{i\in I} a_i \mid a \in \bigtimes_{i\in I} A_i\}.$$

Warning: If A is a subset of M, we most often have $2A \neq A + A$; more generally, we usually have $nA \neq \sum_{k\in n^\sqsubset} A$ for a given $n \in \mathbb{N}$.

We define a **commutative group (written additively)** to be a commutative monoid G, written additively, endowed with additional structure by the prescription of a mapping $(x \mapsto -x) : G \to G$, subject to the following condition:

(CG) : $\quad \forall x \in G, \quad x + (-x) = 0$ (*law of opposites*).

Of course (CM2) and (CG) imply that $(-x) + x = 0$ for all $x \in G$.

We again adopt the "additive" notation in this definition, because it is the one that occurs most frequently in practice. In this notation, the mapping $x \mapsto -x$ is called **opposition**, and $-x$ is called the **opposite of** x, and is read "**minus** x". (The barbarism "negative of x" or, worse, "negative x", for $-x$ is to be avoided: there usually is nothing negative about $-x$.) When multiplicative notation is used, one uses the term **reciprocal** instead of *opposite*; the reciprocal of a member x of the group is never denoted by $-x$, but the notation varies.

131A. Examples. (a): Let the S be given. Then $\mathfrak{P}(S)$ becomes a commutative group with $\emptyset$ as zero, symmetric difference as addition, and the identity mapping $1_{\mathfrak{P}(S)}$ as opposition (Proposition 16E).

(b)*: The sets $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$, each with the number 0 as zero, and with the usual addition and opposition of real numbers, suitably adjusted, are commutative groups written additively. The sets $\mathbb{R}^\times$, $\mathbb{P}^\times$, $\mathbb{Q}^\times$, $\{1, -1\}$, each with the number 1 as unity and multiplication of real numbers and reciprocation of non-zero real numbers, suitably adjusted, are commutative groups written multiplicatively.

(c): A doubleton $\{\mathrm{even}, \mathrm{odd}\}$, with even as zero, addition defined by the rules even + even := odd + odd := even, even + odd := odd + even := odd, and opposition defined to be the identity mapping, is a commutative group written additively. (Valid but disingenuous disclaimer: "Any resemblance is purely coincidental".) ■

Let the commutative group G, written additively, be given. We define the mapping $((x,y) \mapsto x - y) : G \times G \to G$, called **subtraction**, by the rule

$$x - y := x + (-y) \qquad \text{for all } x, y \in G.$$

We call $x - y$ the **difference of** x **and** y (read "x **minus** y"). In particular (CG) implies

$$x - x = 0 \qquad \text{for all } x \in G.$$

If multiplicative notation is used, *subtraction* is replaced by **division**, *difference* by **quotient**, "*minus*" by "**over**" or "*upon*" or "*divided by*"; the notation for the quotient of x and y is usually x/y or $\frac{x}{y}$.

We return to additive notation. On the understanding that addition and subtraction are performed sequentially "from left to right", we omit parentheses in such expressions as $x + y - z := (x + y) - z$, $-x - y - z := ((-x) - y) - z$, etc.

For subsets $A,\ B$ of a commutative group written additively, we set

$$-A \ := \{-x \mid x \in A\}$$

$$A - B \ := A + (-B) = \{x - y \mid (x, y) \in A \times B\};$$

and if $a,\ b$ are members of the group, we again abbreviate thus: $a - B := \{a\} - B$, $A - b := A - \{b\}$. No multiplicative analogues of these notations are in use.

Our next result expresses one of the most fundamental properties of commutative groups.

131B. Proposition. *Let the commutative group G, written additively, be given. For all $a, b \in G$, the equation*

$$?x \in G, \quad a + x = b$$

has exactly one solution, namely $b - a$.

Proof. For every $x \in G$, if $a + x = b$, then $x = x + 0 = x + a + (-a) = (a + x) - a = b - a$. On the other hand, $a + (b - a) = b + (-a) + a = b + 0 = b$. Therefore $a + x = b$ if and only if $x = b - a$. ■

131C. Corollary. *Let the commutative group G, written additively, be given. Then*

$$\forall x, y, z \in G, \quad x + y = x + z \ \Rightarrow \ y = z \qquad \textit{(cancellation law).}$$

131D. Corollary. *Let the commutative group G, written additively, be given. For all $x, y \in G$ and all $m, n \in \mathbb{N}$ we have*

$$-0 = 0 \tag{131.1}$$

$$-(x + y) = (-x) + (-y) \tag{131.2}$$

$$-(x - y) = y - x \tag{131.3}$$

$$-(-x) = x \tag{131.4}$$

$$n(-x) = -nx \tag{131.5}$$

$$n(x-y) = nx - ny \tag{131.6}$$

$$\text{if } m \geq n, \text{ then } (m-n)x = mx - nx. \tag{131.7}$$

Proof. We shall only prove (131.5), (131.6), (131.7). By Proposition 115B(c),(a), we have

$$n(x-y) + ny = n(x-y+y) = n(x+0) = nx = nx + 0 = nx - ny + ny$$

$$(m-n)x + nx = (m-n+n)x = mx = mx - nx + nx,$$

and the cancellation law (Corollary 131C) yields (131.6) and (131.7). From (115.2) we have $n0 = 0$. From (131.6) with $y := 0$ we obtain $n(-x) = n(0-x) = n0 - nx = 0 - nx = -nx$, so that (131.5) holds. ∎

131E. Corollary. *Let the commutative group G, written additively, be given. Let the family $a \in G^I$ be given, and set*

$$-a := (-a_i \mid i \in I) \in G^I.$$

Then $\mathrm{Supp}(-a) = \mathrm{Supp}a$. *For every finite subset J of I, and for every subset J of I if* $\mathrm{Supp}a$ *is finite, we have*

$$\sum_J (-a) = -\sum_J a. \tag{131.8}$$

Proof. From (131.1) and (131.4) we obtain $\mathrm{Supp}(-a) = \mathrm{Supp}a$. From (131.1) and (131.2) we have

$$-\sum_{\emptyset} a = -0 = 0$$

$$-\sum_J a = -(\sum_{J\setminus\{j\}} a + a_j) = -\sum_{J\setminus\{j\}} a + (-a_j) = -\sum_{J\setminus\{j\}} a + (-a)_j$$

for all $J \in \mathfrak{F}^{\times}(I)$ and $j \in J$.

It follows from Theorem 112A (cf. (112.7), (112.8)) that (131.8) holds for all $J \in \mathfrak{F}(I)$. If $\mathrm{Supp}a = \mathrm{Supp}(-a)$ is finite, we apply (131.8) to the finite set $J \cap \mathrm{Supp}a$ instead of J and find, using (113.4), that

$$\sum_J (-a) = \sum_{J\cap \mathrm{Supp}(-a)} (-a) = \sum_{J\cap \mathrm{Supp}a} (-a) = -\sum_{J\cap \mathrm{Supp}a} a = -\sum_J a$$

holds for all $J \in \mathfrak{P}(I)$. ∎

132. Commutative rings

We define a **commutative ring** to be a commutative group A, written *additively*, endowed, in addition (no pun intended!), with the structure of a commutative monoid, written *multiplicatively*, subject to the following condition:

$$\text{(CR)}: \qquad \forall x, y, z \in A, \quad (x+y)z = (xz) + (yz) \qquad (\textit{distributive law}).$$

Of course the distributive law and the commutative law for multiplication imply that $z(x+y) = (zx) + (zy)$ for all $x, y, z \in A$.

We adopt the usual convention concerning parentheses: in the absence of parentheses indicating the contrary, multiplication is to be executed with priority over addition, subtraction and opposition. Thus, the right-hand side of (CR) may be written $xz + yz$.

A commutative ring with 0 as its only member is said to be **trivial**.

132A. Examples. (a): Let the set S be given. The $\mathfrak{P}(S)$ becomes a commutative ring with $\emptyset$ as zero, symmetric difference as addition, the identity mapping as opposition, S as unity, and intersection as multiplication (Proposition 16E). This commutative ring is trivial if and only if $S = \emptyset$.

(b)*: The set $\mathbb{Z}$, with the usual zero, addition, opposition, and multiplication, and with 1 as unity, is a commutative ring. ■

We record some simple consequences of the definition.

132B. Proposition. *Let the commutative ring A be given. For all $x, y, z \in A$ we have*

$$0x = 0 \qquad \textit{where } 0 \textit{ is the zero of } A \tag{132.1}$$

$$(-x)y = x(-y) = -xy \quad \textit{and} \quad (-x)(-y) = xy \tag{132.2}$$

$$(x-y)z = xz - yz. \tag{132.3}$$

Proof. For every $x \in A$ the distributive law yields

$$0x + 0x = (0+0)x = 0x = 0x + 0,$$

and the cancellation law for addition shows that $0x = 0$. This establishes (132.1).

Let $x, y, z \in A$ be given. The distributive law and (132.1) yield

$$xy + (-x)y = (x + (-x))y = 0y = 0 = xy + (-xy),$$

$$(x-y)z + yz = (x - y + y)z = (x+0)z = xz = xz + 0 = xz - yz + yz,$$

and the cancellation law shows that $(-x)y = -xy$ and $(x-y)z = xz - yz$. This establishes part of (132.2) and also (132.3). The rest of (132.2) follows by the commutative law of multiplication. ■

In (132.1), $0x$ was the product of the members 0 and x of A; (132.1) shows that there is no notational clash with $0x$, the 0th natural multiple of x, which is also $0 \in A$, by (115.2).

132C. Corollary. *Let the commutative ring A, with unity e, be given. Then A is trivial if (and only if) $e = 0$.*

Proof. If $e = 0$, then (132.1) shows that $x = ex = 0x = 0$ for all $x \in A$. ■

We require some generalized versions of the distributive law.

132D. Proposition. *Let the commutative ring A, with unity e, be given.*

(a): *Let the family $a \in A^I$ and $y \in A$ be given. Then* $\mathrm{Supp}(a_i y \mid i \in I) \subset \mathrm{Supp} a$. *For every finite subset J of I, and for every subset J of I if* Supp$a$ *is finite, we have*

$$(\sum_J a)y = \sum_{j\in J} a_j y. \tag{132.4}$$

(b): *Let the families $a \in A^I$ and $b \in A^K$ be given. Then* $\mathrm{Supp}(a_i b_k \mid (i,k) \in I \times K) \subset \mathrm{Supp} a \times \mathrm{Supp} b$. *For all finite subsets J of I and L of K, and for all subsets J of I and L of K if* Supp$a$ *and* Supp$b$ *are finite, we have*

$$(\sum_J a)\ (\sum_L b) = \sum_{(j,l)\in J\times L} a_j b_l.$$

(c): $(mx)(ny) = (mn)(xy)$ *for all* $x, y \in A$ and $m, n \in \mathbb{N}$; *in particular*, $(me)y = my$ *for all* $y \in A$ *and* $m \in \mathbb{N}$.

Proof. Proof of (a). The inclusion $\mathrm{Supp}(a_i y \mid i \in I) \subset \mathrm{Supp} a$ follows from (132.1). From (132.1) and the distributive law we have

$$(\sum_\emptyset a)y = 0y = 0 = \sum_{j\in\emptyset} a_j y$$

$$(\sum_J a)y = (\sum_{J\setminus\{j\}} a + a_j)y = (\sum_{J\setminus\{j\}} a)y + a_j y \qquad \text{for all } J \in \mathfrak{F}^\times(I) \text{ and } j \in J.$$

It follows from Theorem 112A (cf. (112.7), (112.8)), that (132.4) holds for all $J \in \mathfrak{F}(I)$. If Supp$a$ is finite, the proof is completed for all $J \in \mathfrak{P}(I)$ as in Corollary 131E.

Proof of (b). By (a), the commutative law for multiplication, and Theorem 114F, we have

$$(\sum_J a)\ (\sum_L b) = \sum_{j\in J}(a_j \sum_L b) = \sum_{j\in J}\sum_{l\in L} a_j b_l = \sum_{(j,l)\in J\times L} a_j b_l.$$

Proof of (c). Let $x, y \in A$ and $m, n \in \mathbb{N}$ be given. By (b) and Lemma 115A and Corollaries 101K and 103G,

$$(mx)(ny) = (\sum_{i\in m^\sqsubset} x)\ (\sum_{k\in n^\sqsubset} y) = \sum_{(i,k)\in m^\sqsubset\times n^\sqsubset} xy = (\#(m^\sqsubset \times n^\sqsubset))(xy) = (mn)(xy).$$ ■

On account of Proposition 132D,(c), we may drop parentheses in expressions such as nxy for $n \in \mathbb{N}$ and $x, y \in A$.

132E. Theorem. *Let the commutative ring A, the family of sets $(K_i \mid i \in I)$, and the family $u \in \bigtimes_{i\in I} A^{(K_i)}$ be given. For all finite subsets J of I we have*

$$\prod_{i\in J}\sum_{K_i} u_i = \sum_{h\in P}\prod_{i\in J}(u_i)_{h_i}, \quad \text{where } P := \bigtimes_{i\in J} K_i. \tag{132.5}$$

Proof. We denote the assertion (132.5) by $Q(J)$, and show by special induction that $Q(J)$ holds for all $J \in \mathfrak{F}(I)$. Suppose that $J := \emptyset$. Then $P = \{\emptyset\}$, and

$$\sum_{h\in P}\prod_{i\in J}(u_i)_{h_i} = \sum_{h\in\{\emptyset\}}\prod_{i\in\emptyset}(u_i)_{h_i} = \sum_{h\in\{\emptyset\}} e = e = \prod_{i\in\{\emptyset\}}\sum_{K_i} u_i = \prod_{i\in J}\sum_{K_i} u_i,$$

where e is the unity of A. Therefore $Q(\emptyset)$ holds.

Let now $J \in \mathfrak{F}^\times(I)$ and $j \in J$ be given, and assume that $Q(J\backslash\{j\})$ holds. Set $J' := J\backslash\{j\}, P' := \bigtimes_{i\in J'} K_i$. By Proposition 44A, the mapping

$$h \mapsto (h|_{J'}, h_j) : P \to P' \times K_j \tag{132.6}$$

is bijective. In the following computation we use consecutively: the characterizing property of the product; the induction hypothesis; the generalized distributive law of Proposition 132D,(b); the bijectivity of (132.6), together with Corollary 114E (applied to multiplication) for re-indexing; and again the characterizing property of the product. We find

$$\begin{aligned}\prod_{i\in J}\sum_{K_i} u_i &= (\prod_{i\in J'}\sum_{K_i} u_i)\,(\sum_{K_j} u_j) = (\sum_{l\in P'}\prod_{i\in J}(u_i)_{l_i})\,(\sum_{k\in K_j}(u_j)_k) = \\ &= \sum_{(l,k)\in P'\times K_j}(\prod_{i\in J'}(u_i)_{l_i})(u_j)_k = \sum_{h\in P}(\prod_{i\in J'}(u_i)_{h_i})(u_j)_{h_j} = \\ &= \sum_{h\in P}\prod_{i\in J}(u_i)_{h_i}.\end{aligned}$$

This shows that $Q(J)$ holds, and completes the induction step. ∎

To illustrate the application of Theorem 132E, we present a derivation of the Binomial Theorem. In the next corollary, addition of families of members of A is defined termwise, as in Section 117.

132F. Corollary. *Let the commutative ring A and the finite families $a, b \in A^I$ be given. Then*

$$\prod_I (a+b) = \sum_{K\in\mathfrak{P}(I)}(\prod_K a)(\prod_{I\backslash K} b).$$

Proof. Define $u \in (A^{2^\sqsubset})^I$ by $(u_i)_0 := b_i$ and $(u_i)_1 := a_i$ for all $i \in I$. From Theorem 132E, we then obtain

$$\prod_I (a+b) = \prod_{i\in I}((u_i)_0 + (u_i)_1) = \prod_{i\in I}\sum_{2^\sqsubset} u_i = \sum_{h\in(2^\sqsubset)^I}\prod_{i\in I}(u_i)_{h_i}. \tag{132.7}$$

The mapping

$$K \mapsto \chi_{K\subset I} : \mathfrak{P}(I) \to (2^\sqsubset)^I$$

is bijective (cf. proof of Corollary 103J). By Corollary 114E (for re-indexing) and Corollary 114C we have

$$\begin{aligned}\sum_{h\in(2^\sqsubset)^I}\prod_{i\in I}(u_i)_{h_i} &= \sum_{K\in\mathfrak{P}(I)}\prod_{i\in I}(u_i)_{\chi_{K\subset I}(i)} = \\ &= \sum_{K\in\mathfrak{P}(I)}(\prod_{i\in K}(u_i)_1)(\prod_{i\in I\setminus K}(u_i)_0) = \\ &= \sum_{K\in\mathfrak{P}(I)}(\prod_K a)(\prod_{I\setminus K} b).\end{aligned} \tag{132.8}$$

Combination of (132.7) and (132.8) yields the assertion to be proved. ■

132G. Theorem (Binomial Theorem). *Let the commutative ring A be given. Let $x, y \in A$ and $n \in \mathbb{N}$ be given. Then*

$$(x+y)^n = \sum_{k\in(n+1)^\sqsubset}\binom{n}{k}x^k y^{n-k}. \tag{132.9}$$

Proof. By Corollary 132F we have, using Lemma 115A (multiplicative version), Corollary 101K, and Theorem 103A,

$$\begin{aligned}(x+y)^n &= \prod_{k\in n^\sqsubset}(x+y) = \sum_{K\in\mathfrak{P}(n^\sqsubset)}(\prod_{i\in K}x)(\prod_{j\in n^\sqsubset\setminus K}y) = \\ &= \sum_{K\in\mathfrak{P}(n^\sqsubset)} x^{\#K}y^{\#(n^\sqsubset\setminus K)} = \sum_{K\in\mathfrak{P}(n^\sqsubset)} x^{\#K}y^{n-\#K}.\end{aligned} \tag{132.10}$$

By Proposition 101N and Corollary 101K, the disjoint family $(\mathfrak{F}_k(n^\sqsubset)\,|\,k\in(n+1)^\sqsubset)$ has no empty term, and its union is $\mathfrak{P}(n^\sqsubset)$. By Corollary 114C and Lemma 115A we have

$$\begin{aligned}\sum_{K\in\mathfrak{P}(n^\sqsubset)} x^{\#K}y^{n-\#K} &= \sum_{k\in(n+1)^\sqsubset}\sum_{K\in\mathfrak{F}_k(n^\sqsubset)} x^k y^{n-k} = \\ &= \sum_{k\in(n+1)^\sqsubset}(\#\mathfrak{F}_k(n^\sqsubset))x^k y^{n-k} = \sum_{k\in(n+1)^\sqsubset}\binom{n}{k}x^k y^{n-k}.\end{aligned} \tag{132.11}$$

Combination of (132.10) and (132.11) yields (132.9). ■

133. Fields

A non-trivial commutative ring F (with unity e) is called a **field** if it satisfies the following condition:

(F) : $\forall x \in F^\times,\ \exists y \in F, \quad xy = e.$

We note that in a field $e \in F^\times$, by Corollary 132C.

133A. Examples. (a): let the set S be given. Then the commutative ring $\mathfrak{P}(S)$ described in Example 132A,(a) is a field if and only if S is a singleton; in that case $\mathfrak{P}(S)$ is a doubleton.

(b)*: The sets $\mathbb{R}$ and $\mathbb{Q}$, with the usual zero, addition, opposition, and multiplication, and with 1 as unity, are fields; $\mathbb{Z}$ is not.

(c): The commutative group {even, odd} described in Example 131A,(c), becomes a field with odd as unity, and multiplication defined by even·even := even·odd := odd·even := even, and odd·odd := odd. ■

132B. Proposition. *Let the field F be given. For each $a \in F^\times$, $b \in F$, the equation*

$$?x \in F, \quad ax = b \tag{133.1}$$

has exactly one solution. That solution is 0 *if and only if* $b = 0$.

Proof. By (F) we may choose $c \in F$ such that $ac = e$, wher e is the unity of F. For every $x \in F$, if $ax = b$, then $xex = (ac)x = cax = c(ax) = cb$. On the other hand, $a(cb) = (ac)beb = b$. Therefore $ax = b$ if and only if $x = cb$. This prove that (133.1) has exactly one solution. The final assertion follow from (132.1). ■

Given $a \in F^\times$ and $b \in F$, the only solution of (133.1) is denoted by b/a or $\frac{b}{a}$ (read "b **over** a" or "b *upon* a" or "b *divided by* a"). The mapping $((x, y) \mapsto x/y) : F \times F^\times \to F$ is called **division**. For given $x \in F^\times$, e/x is called the **reciprocal** of x (or *multiplicative inverse of* x).

133C. Corollary. *Let the field F be given. Then*

$$\forall x \in F^\times,\ \forall y, z \in F, \quad xy = xz \Rightarrow y = z \text{ (cancellation law for multiplication).}$$

We observe that $0 \cdot 0 = 0 = 0e$; hence the condition $x \neq 0$ cannot be omitted in Corollary 133C.

133D. Corollary. *Let the field F, with unity e, be given. Then $F^\times F^\times \subset F^\times$; and the set $F^\times$, with e as unity, $((x, y) \mapsto xy) : F^\times \times F^\times \to F^\times$ as multiplication, and e/x as the reciprocal of x for all $x \in F^\times$, is a commutative group, written multiplicatively.*

The set $F^\times$, thus endowed with the structure of a commutative group, written multiplicatively, is called the **multiplicative group of** the field F; in contrast, the commutative group F, written additively, is called the **additive group of** the field F.

133E. Corollary. *Let the field F, with unity e, be given. For all $x, y \in F$ and $u, v, w \in F^\times$ and $n \in \mathbb{N}$ we have*

$$x/e = x$$
$$-(x/u) = (-x)/u = x/(-u)$$
$$(x/u) + (y/v) = (vx + uy)/(uv)$$
$$(x/u) - (y/v) = (vx - uy)/(uv)$$
$$(x/u)(y/v) = (xy)/(uv)$$
$$e/(v/u) = u/v$$
$$(x/u)/(v/w) = (xw)/(uv)$$
$$n(x/u) = (nx)/u.$$

Proof. As a sample, we shall only prove the third assertion. We have $uv \in F^\times$, and

$$\begin{aligned} uv((x/u) + (y/v)) &= uv(x/u) + uv(y/v) = v(u(x/u)) + u(v(y/v)) = \\ &= vx + uy = uv((vx + uy)/(uv)). \end{aligned}$$

The cancellation law for multiplication then yields $(x/u) + (y/v) = (vx + uy)/(uv)$. ■

133F. Remarks. (a): The rules of operation in commutative monoids and groups, commutative rings, and fields that we have obtained in Sections 131, 132, 133 (and in the relevant sections of Chapter 11) have all a familiar form. In the sequel we shall use them freely, and shall hardly ever cite chapter and verse for them. The reader should have no difficulty in locating the appropriate reference when in doubt.

(b): One "familiar" rule that is conspicuously missing, even in the current section on fields, is $\mathbb{N}^\times F^\times \subset F^\times$ for a field F, i.e.,

$$\forall x \in F,\ \forall n \in \mathbb{N}, \quad nx = 0 \ \Rightarrow (n = 0 \text{ or } x = 0).$$

The reason for the absence of this rule is that it is not valid in general: in Examples 133A,(c) we have 2odd = odd + odd = even, which is the zero of this field. ■

Chapter 14

THE REAL NUMBERS: COMPLETE ORDERED FIELDS

141. Introduction

The purpose of this chapter and the next is to introduce the systems $\mathbb{R}$, $\mathbb{Q}$, and $\mathbb{Z}$ of real numbers, of rational numbers, and of integers, respectively, without engaging in a thorough or prolonged discussion of their properties.

Of the various methods in current use for the introduction of the Real-Number System, we adopt one that stresses the operational structure of the system, rather than one that purports to address the question of what real numbers "really are". We therefore introduce the Real-Number System $\mathbb{R}$ as an instance of a *complete ordered field*, as defined in Section 143. We encounter here the same questions that we addressed when we introduced the Natural-Number System as an instance of a counting system in Chapter 9. One is the question regarding the "*essential uniqueness*" of a complete ordered field: we shall show that any two complete ordered fields are naturally isomorphic as regards their complete-ordered-field structure (Section 144) — although they may in other respects be quite different. Once this is shown, it matters little which particular complete ordered field we accept, by choice, by construction, or by authority, as being *the* Real-Number System.

The question regarding the *existence* of a complete ordered field then remains. We judged the corresponding question regarding the existence of a counting system to be a foundational issue, beyond the scope of this work. We take very much the same view of our current existence question, but we shall establish that the two problems are equivalent, by (a) showing that every complete ordered field "includes" a counting system, and (b) showing how to construct a complete ordered field from a counting system. We regard this analysis of the existence question as tangential to our chief concerns, and therefore relegate the execution of the building process (b) to an appendix (Chapter 16). Both here and in the proof of the natural-isomorphism result mentioned before we use a quite faithful adaptation of the theory of ratios attributed to Eudoxos (Εύδοξος, 4th century B.C.E.), which gave the first effective account of what we call positive irrational numbers (*irrational*, when applied to numbers, means

"without ratio", not "without reason").

Common mathematical discourse regards $\mathbb{N}$ as a subset of $\mathbb{R}$, and indeed accepts the inclusions $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. We subscribe to this view, which we shall discuss further in Section 151. This view rules out, however, a traditional version of the building process (b), one that starts from the Natural-Number System $\mathbb{N}$ itself and proceeds by successive construction of $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ ($\mathbb{Z}$ is sometimes skipped) in such a way that none of these sets is a subset of any other. This version, even when adjusted by suitable modifications to circumvent this obstacle, is not, in our view, so straightforward as the one we have adopted.

142. Ordered fields

We define an **ordered field** to be a field F endowed with additional structure by the prescription of a subset P of F, called the **positive half of** F, subject to the following conditions:

(OF1) : $P + P \subset P$
(OF2) : $PP \subset P$
(OF3) : $P \cap (-P) = \{0\}$
(OF4) : $P \cup (-P) = F$.

Conditions (OF1) and (OF2) are often expressed by saying that P *is stable with respect to addition and multiplication*.

In this section and in the next, whenever F denotes an ordered field, the unity of F and the positive half of F will always be denoted by e and P, respectively.

142A. Lemma. *Let the ordered field F be given. Then*

$$F\backslash P = -P^\times \qquad F\backslash P^\times = -P \tag{141.1}$$

$$P^\times \cup \{0\} \cup (-P^\times) = F \tag{142.2}$$

$$P + P^\times \subset P^\times \tag{142.3}$$

$$P^\times P^\times \subset P^\times \tag{142.4}$$

$$\forall x, y \in P^\times, \quad x/y \in P^\times. \tag{142.5}$$

Proof. (142.1) and (142.2) follow immediately from (OF3) and (OF4). Let $x, y \in P$ be given, so that $x + y \in P$ by (OF1). If $x + y = 0$, then $y = -x \in -P$, and by (OF3) we find $y = 0$. Thus $y \in P^\times$ implies $x + y \in P^\times$. Since $x, y \in P$ were arbitrary, this establishes (142.3). By (OF2), $P^\times P^\times \subset PP \subset P$; but by Corollary 133D, $P^\times P^\times \subset F^\times F^\times \subset F^\times$. Hence $P^\times P^\times \subset P \cap F^\times = P^\times$. This establishes (142.4). Let $x \in F$ and $y \in P^\times$ be given. If $x/y \in F\backslash P^\times = -P$, then $x = y(x/y) \in P^\times(-P) = -P^\times P \subset -P = F\backslash P^\times$; hence $x \in P^\times$ implies $x/y \in P^\times$. This establishes (142.5). ■

Let the ordered field F be given. We define the relation $<$ in F by the rule

$$\forall x, y \in F, \quad x < y \ :\Leftrightarrow\ y - x \in P^\times.$$

The term *ordered field* is justified by the following result.

142B. Proposition. *Let the ordered field F be given. Then $<$ is a total strict order in F.*

Proof. For all $x \in F$ we have $x - x = 0 \notin P^\times$, so that $<$ is irreflexive. Let $x, y \in F$ be given. By (142.2), either $y - x \in P^\times$ or $y - x = 0$ or $x - y = -(y - x) \in P^\times$.

Hence $<$ is total. Let $x, y, z \in F$ be given, and assume that $x < y$ and $y < z$. Then $z - x = (z - y) + (y - x) \in P^\times + P^\times \subset P^\times$, by (142.3). Therefore $<$ is transitive. ∎

We denote the (lax) order corresponding to the strict-order $<$ in F by $\leq$ (the notation $\leqq$ is also in common use). We shall always regard the set F as ordered by $\leq$. We find

$$\forall x, y \in F, \quad x \leq y \;\Leftrightarrow\; y - x \in P,$$
$$P = \mathrm{Ub}(\{0\}).$$

A member of F is said to be **positive, negative, strictly positive, strictly negative** if it is in P, $-P$, $P^\times$, $-P^\times$, respectively. (A conflicting convention that is not uncommon, but that we do not use, employs the respective terms *non-negative, non-positive, positive, negative* instead; it is mannerly to declare one's choice between these conventions when confusion might otherwise arise.)

142C. Proposition. *Let the ordered field F be given. Then the following monotonicity laws hold:*

$$\forall x, y, z \in F, \quad x < y \;\Leftrightarrow\; x + z < y + z \;\Leftrightarrow\; x - z < y - z \tag{142.6}$$

$$\forall x, y \in F, \quad x < y \;\Leftrightarrow\; -y < -x \tag{142.7}$$

$$\forall x, y \in F,\ \forall z \in P^\times, \quad x < y \;\Leftrightarrow\; xz < yz \;\Leftrightarrow\; x/z < y/z \tag{142.8}$$

$$\forall x, y \in F,\ \forall u, v \in P^\times, \quad vx < uy \;\Leftrightarrow\; x/u < y/v. \tag{142.9}$$

Proof. For all $x, y, z \in F$ we have $(y + z) - (x + z) = (y - z) - (x - z) = y - x$ and $(-x) - (-y) = -x + y = y - x$; this establishes (142.6) and (142.7). Let $x, y \in F$ and $z \in P^\times$ be given. If $x < y$, then $yz - xz = (y - x)z \in P^\times P^\times \subset P^\times$, by (142.4), and hence $xz < yz$; if, on the other hand, $x \geq y$, then $xz - yz = (x - y)z \in PP \subset P$, and hence $xz \geq yz$. This shows that $x < y \;\Leftrightarrow\; xz < yz$. Applying this equivalence to x/z and y/z instead of x and y, we find that $x/z < y/z \;\Leftrightarrow\; x < y$. This establishes (142.8). Let $x, y \in F$ and $u, v \in P^\times$ be given. Then $uv \in P^\times$ by (142.4), and by (142.8) $vx < uy$ if and only if $x/u = (vx)/(vu) < (uy)/(uv) = y/v$. This establishes (142.9). ∎

142D. Proposition. *Let the ordered field F be given.*

(a): *for every $x \in F$ we have $xx \in P$.*

(b): *$e \in P^\times$; for all $x \in P^\times$ we have $e/x \in P^\times$.*

(c): *The mapping $(x \mapsto xx) : P \to P$ is strictly isotone, hence injective.*

Proof. Proof of (a). Let $x \in F$ be given. By (OF4) and (OF2) we have $xx \in (PP) \cup ((-P)(-P)) = (PP) \cup (PP) = PP \subset P$.

Proof of (b). By (a), $e = ee \in P$, but $e \neq 0$, hence $e \in P^\times$. The remaining assertion follows from (142.5).

Proof of (c). Let $x, y \in P$ be given, and assume that $x < y$. Then $y - x \in P^\times$, and $y + x = (y - x) + x + x \in P^\times + P + P \subset P^\times$, by (142.3). Therefore $yy - xx =$

$(y-x)(y+x) \in P^\times P^\times \subset P^\times$, by (142.4), so that $xx < yy$. Thus $(x \mapsto xx) : P \to P$ is strictly isotone. Since P is totally ordered, this mapping is injective (Remark 62A,(a)). ■

142E. Theorem. *Let the ordered field F be given. Set*

$$N := \bigcap \{A \in \mathfrak{P}(F) \mid 0 \in A,\ A + e \subset A\}, \tag{142.10}$$

the smallest member of the intersection-stable collection $\mathcal{A} := \{A \in \mathfrak{P}(F) \mid 0 \in A, A + e \subset A\}$. Then $0 \in N$, $N + e \subset N \subset P$; and the set N, with 0 as zero and the mapping $(x \mapsto x + e) : N \to N$ as successor-mapping, is a counting system.

Proof. The collection $\mathcal{A}$ is obviously intersection-stable, and hence $N \in \mathcal{A}$, so that $0 \in N$ and $N + e \subset N$. By Proposition 142D,(b) we have $e \in P^\times$ and hence, by (OF1), $P + e \subset P + P \subset P$; by (OF3) we have $0 \in P$. Thus $P \in \mathcal{A}$, and hence $N \subset P$.

Define $\sigma\colon N \to N$ by the rule $\sigma(x) := x + e$ for all $x \in N$. Then $\mathrm{Rng}\sigma = N + e \subset P + P^\times \subset P^\times$, by (142.3). Hence $0 \notin \mathrm{Rng}\sigma$, and (Count I) holds for the proposed counting system; σ is injective, by the cancellation law for addition in F, so that (Count II) holds. If S is a subset of N such that $0 \in S$ and $\sigma_>(S) \subset S$, then $S + e \subset S$, so that $S \in \mathcal{A}$. Hence $N \subset S \subset N$, and equality holds, thus (Count III) also holds for N, with 0 and σ. ■

The next result shows that the mapping $(n \mapsto ne) : \mathbb{N} \to F$ permits an identification of $\mathbb{N}$ with the counting system described in Theorem 142E. (We do not actually make this identification, in which n would serve as a symbol for ne for each $n \in \mathbb{N}$, at this time; but see Section 151.)

142F. Proposition. *Let the ordered field F be given, and let the subset N of F be defined by* (142.10). *Then $\mathbb{N}e = N$ (so that $\mathbb{N}e \subset P$), and the mapping $(n \mapsto ne) : \mathbb{N} \to \mathbb{N}e$ is an order-isomorphism. Moreover,*

$$0e = 0 \qquad 1e = e \tag{142.11}$$

$$\forall m, n \in \mathbb{N}, \quad (m+n)e = me + ne \quad (mn)e = (me)(ne) \quad m^n e = (me)^n \tag{142.12}$$

$$\forall m, n \in \mathbb{N}, \quad m \geq n \ \Rightarrow\ (m-n)e = me - ne. \tag{142.13}$$

Proof. We note that (142.11) follows from (115.2); the first part of (142.12) from Proposition 115B,(a); the second part of (142.12) from Proposition 132D,(c); and (142.13) from (131.7). The third part of (142.12) follows from the second by induction: the core of the induction step reads

$$m^{\mathrm{seq}n}e = (m^n m)e = (m^n e)(me) = (me)^n(me) = (me)^{\mathrm{seq}n}.$$

We now prove the first part of the statement. Theorem 142E describes the counting system N, and asserts that $N \subset P$. By Theorem 95A, there exists exactly one mapping $\phi : \mathbb{N} \to N$ such that $\phi(0) = 0$ and $\phi(\mathrm{seq}n) = \phi(n) + e$ for all $n \in \mathbb{N}$, and this

mapping is bijective. We now prove that $\phi(n) = ne$ for all $n \in \mathbb{N}$; this will show that $\mathbb{N}e = N$. By (142.11), $\phi(0) = 0 = 0e$. Let $n \in \mathbb{N}$ be such that $\phi(n) = ne$. By (142.11) and (142.12) we then have $\phi(\text{seq}n) = \phi(n) + e = ne + 1e = (n+1)e = (\text{seq}n)e$. This completes the induction step.

It remains to prove that ϕ is an order-isomorphism. Let $m, n \in \mathbb{N}$ be given. If $m \geq n$, then (142.13) yields $me - ne = (m-n)e \in \mathbb{N}e \subset P$, and hence $me \geq ne$. Thus ϕ is isotone. Since $\mathbb{N}$ is totally ordered and ϕ is bijective, ϕ is indeed an order-isomorphism (Proposition 62D). ■

142G. Corollary. *Let the ordered field F be given.*

(a): *The following cancellation laws hold:*

$$\forall n \in \mathbb{N}^\times,\ \forall x, y \in F, \quad nx = ny \ \Rightarrow\ x = y \tag{142.14}$$

$$\forall x \in F^\times,\ \forall m, n \in \mathbb{N}, \quad mx = nx \ \Rightarrow\ m = n. \tag{142.15}$$

(b): *$\mathbb{N}^\times P^\times \subset P^\times$, and the following monotonicity laws hold:*

$$\forall n \in \mathbb{N}^\times,\ \forall x, y \in F, \quad x < y \ \Leftrightarrow\ nx < ny \tag{142.16}$$

$$\forall x \in P^\times,\ \forall m, n \in \mathbb{N},\ m < n \ \Leftrightarrow\ mx < nx. \tag{142.17}$$

Proof. By Proposition 142F, the mapping $(n \mapsto ne) : \mathbb{N} \to P$ is injective. Hence $\mathbb{N}^\times e = (\mathbb{N}e)\backslash\{0e\} \subset P^\times$. Then $\mathbb{N}^\times P^\times = (\mathbb{N}^\times e)P^\times \subset P^\times P^\times \subset P^\times$ by (142.4). Moreover, (142.14), (142.15) then follow from the cancellation law for multiplication, and (142.16), (142.17) from (142.8) and the fact that $(n \mapsto ne) : \mathbb{N} \to \mathbb{N}e$ is an order-isomorphism. For instance, if $x \in P^\times$ and $m, n \in \mathbb{N}$ are given, we have

$$m < n \ \Leftrightarrow\ me < ne \ \Leftrightarrow\ mx = (me)x < (ne)x = nx. \ \blacksquare$$

142H. Proposition. *Let the ordered field F be given. The order $\leq$ in F is dense.*

Proof. By Proposition 142D,(b) and Corollary 142G,(b) we have $2e \in \mathbb{P}^\times$. Let $x, y \in F$ be given, and assume that $x < y$, so that $y - x \in \mathbb{P}^\times$. Set $z := (x+y)/(2e)$. Then $z - x = y - z = (y-x)/(2e) \in \mathbb{P}^\times$; hence $x < z < y$. ■

Let the ordered field F be given. We define the mapping $(x \mapsto |x|) : F \to P$, called the **absolute-value mapping**, by the rule

$$|x| := \max\{x, -x\} = \begin{cases} x & \text{if } x \in P \\ -x & \text{if } x \in -P \end{cases} \quad \text{for all } x \in F.$$

For every $x \in F$, $|x|$ is called the **absolute value of** x. We also define the mapping $\text{sgn}_F : F \to \{e, 0, -e\}$ by the rule

$$\text{sgn}x := \text{sgn}_F x := \begin{cases} e & \text{if } x \in P^\times \\ 0 & \text{if } x = 0 \\ -e & \text{if } x \in -P^\times. \end{cases}$$

For every $x \in F$, $\mathrm{sgn}x$ is called the **sign of** x (sometimes also the *signum* of x, in order to prevent prosodic confusion with the sine). We record the fundamental properties of these mappings.

142I. Proposition. *Let the ordered field F be given. For all $x, y \in F$ and $z \in F^\times$ we have*

$$|x| = 0 \;\Leftrightarrow\; x = 0$$
$$-|x| \le x \le |x|$$
$$|-x| = |x|$$
$$||x| - |y|| \le |x + y| \le |x| + |y|$$
$$|xy| = |x||y|$$
$$|x/z| = |x|/|z|$$
$$\mathrm{sgn}(xy) = (\mathrm{sgn}x)(\mathrm{sgn}y)$$
$$|x|\mathrm{sgn}x = x.$$

Proof. We shall only prove that $|x + y| \le |x| + |y|$ for all $x, y \in F$. We have $x \le |x|$, $-x \le |x|$, $y \le |y|$, $-y \le |y|$, and therefore, by (142.6), $x + y \le |x| + |y|$ and $-(x + y) = -x - y \le |x| + |y|$; hence $|x + y| = \max\{x + y, -(x + y)\} \le |x| + |y|$. ■

An ordered field F is said to be **archimedean** if $\mathrm{Ub}(\mathbb{N}e) = \emptyset$. (By the testimony of Archimedes himself, this property of the field of real numbers, in the variant form used by him, was used earlier by Eudoxos.) The essential property of archimedean ordered fields is given in the following result.

142J. Proposition. *Let the archimedean ordered field F be given. If $x, y \in P$ and $x < y$, then there exist $p, q \in \mathbb{N}^\times$ such that $qx < pe < qy$.*

This proposition can be roughly restated thus: *Between any two distinct members of P there is another of the form $(pe)/(qe)$, with $p, q \in \mathbb{N}^\times$.*

Proof. We have $y - x \in P^\times$. Since F is archimedean, we may choose $q \in \mathbb{N}^\times$ such that $qe > e/(y-x)$. By (142.8) and (131.6) we have $qy - qx = q(y-x) = (qe)(y-x) > e$, and hence $qx + e < qy$ by (142.6). Again since F is archimedean, we may define $p := \min\{n \in \mathbb{N} \mid ne > qx\}$. Since $0e = 0 \le qx$ by Corollary 142G, we have $p \in \mathbb{N}^\times$, and $pe - e = (p - 1)e \le qx$. Therefore (142.6) yields $qx < pe \le qx + e < qy$. ■

143. Complete ordered fields

An ordered field F is said to be **complete** if the ordered set F is pre-completely ordered.

If F is an ordered field, then $x < x+e$ for every $x \in F$, so that F has no maximum and cannot, therefore, be a *completely* ordered set. The term "complete", as defined here, may thus be somewhat ill-fitting, but it is not actually misleading; we therefore adhere to this traditional terminology.

143A. Lemma. *Let the ordered field F be given. The following statements are equivalent:*

(i): *F is a complete ordered field.*

(ii): *The ordered subset P of F is pre-completely ordered.*

(iii): *Every closed order-interval of F is completely ordered.*

(iv): *The closed order-interval $[\![0, e]\!]$ of F is completely ordered.*

Proof. F is totally ordered, and every closed order-interval of P is a closed order-interval of F. From Proposition 72J we infer the implications (i) $\Leftrightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (iv). It remains to prove that (iv) implies (iii). Let the closed order-interval $[\![a, b]\!]$ of F be given. Since singleton subsets of F are completely ordered, we may assume that $a < b$. The mappings $(x \mapsto a + (b-a)x) : F \to F$ and $(x \mapsto (x-a)/(b-a)) : F \to F$ are isotone, by Proposition 142C, and each is the inverse of the other. They are therefore order-isomorphisms, and the former induces an order-isomorphism from $[\![0, e]\!]$ onto $[\![a, b]\!]$. If $[\![0, e]\!]$ is completely ordered, it follows that $[\![a, b]\!]$ is completely ordered. ■

143B. Proposition. *Let the complete ordered field F be given. Every order-convex subset of F is either empty or a singleton or an uncountable set. In particular, F is uncountable.*

Proof. Let the order-convex subset A of F be given. By Proposition 72J, A is totally and pre-completely ordered by $\leq$; by Proposition 142H, this order is dense. The conclusion follows by Corollary 122G. ■

143C. Proposition. *Let the complete ordered field F be given. The mapping $(x \mapsto x^2) : P \to P$ is an order-isomorphism.*

Proof. By Proposition 142D,(c), the mapping is (strictly) isotone and injective. Since P is totally ordered, it will be enough to show that the mapping is surjective. (Proposition 62D).

Let $y \in P$ be given. Set $a := \max\{e, y\} \in P^\times$. Then $y = ey \leq ay \leq a^2$, by (142.8). For all $t \in P$ we have $at + y \in P$, $t + a \in P^\times$, by (OF1), (OF2), and (142.3). By (142.5) we may define the mapping $f\colon [\![0, a]\!] \to P$ by the rule

$$f(t) := (at + y)/(t + a) = a - (a^2 - y)/(t + a) \qquad \text{for all } t \in [\![0, a]\!].$$

We see that $\operatorname{Rng} f \subset [\![0, a]\!]$. By Proposition 142C, applied repeatedly, we infer that f is (strictly) isotone.

By Lemma 143A, $[\![0, a]\!]$ is completely ordered. By the Knaster Fixed-Point Theorem (Theorem 75A), we may choose a fixed point x of the isotone mapping $f|^{[\![0,a]\!]}$. Then $(ax + y)/(x + a) = x$, and hence $x^2 = y$.

Since $y \in P$ was arbitrary, the mapping $(x \mapsto x^2) : P \to P$ is surjective. ■

For every $y \in P$ the unique solution of the equation

$$?x \in P, \quad x^2 = y$$

is called the **square root of** y, and denoted by $\sqrt{y}$. The mapping $(y \mapsto \sqrt{y}) : P \to P$ is the inverse of $(x \mapsto x^2) : P \to P$, and is therefore an order-isomorphism. ■

143D. Proposition. *Every complete ordered field is archimedean.*

Proof. Let the complete ordered field F, with unity e, be given. Since $\mathbb{N}e+e \subset \mathbb{N}e$, we have $\mathrm{Ub}(\mathbb{N}e + e) \supset \mathrm{Ub}(\mathbb{N}e)$. But the mapping $(x \mapsto x - e) : F \to F$ is an order-isomorphism, by Proposition 142C (its inverse is $(x \mapsto x+e) : F \to F$), and therefore

$$\mathrm{Ub}(\mathbb{N}e) = \mathrm{Ub}(\mathbb{N}e + e - e) = \mathrm{Ub}(\mathbb{N}e + e) - e \supset \mathrm{Ub}(\mathbb{N}e) - e. \tag{143.1}$$

Since $e \in P^\times$ (Proposition 142D,(b)), it follows from (143.1) that $\mathrm{Ub}(\mathbb{N}e)$ has no minimum; hence $\mathbb{N}e$ has no supremum. Since F is pre-completely ordered and $\mathbb{N}e \neq \emptyset$, it follows that $\mathrm{Ub}(\mathbb{N}e) = \emptyset$. ■

143E. Proposition. *Let the complete ordered field and the mapping $\rho : F \to F$ be given. Then ρ satisfies*

$$\rho(x + y) = \rho(x) + \rho(y) \quad \textit{for all } x, y \in F \tag{143.2}$$

$$\rho_>(P) \subset P \tag{143.3}$$

(if and) only if $\rho(e) \in P$ and

$$\rho(x) = x\rho(e) \quad \textit{for all } x \in F. \tag{143.4}$$

Proof. 1. By (143.2) we have $\rho(y)+\rho(x-y) = \rho(y+(x-y)) = \rho(x)$ for all $x, y \in F$. It follows that

$$\rho(x - y) = \rho(x) - \rho(y) \quad \text{for all } x, y \in F. \tag{143.5}$$

In particular,

$$\rho(0) = \rho(0 - 0) = \rho(0) - \rho(0) = 0 \tag{143.6}$$

$$\rho(-x) = \rho(0 - x) = \rho(0) - \rho(x) = 0 - \rho(x) = -\rho(x) \quad \text{for all } x \in F. \tag{143.7}$$

Let $x \in F$ be given. We have $\rho(0x) = \rho(0) = 0 = 0x$, by (143.6). Let $n \in \mathbb{N}$ be such that $\rho(nx) = n\rho(x)$. Then (143.2), (142.11), (142.12) imply

$$\rho((n + 1)x) = \rho(nx + 1x) = \rho(nx) + \rho(x) = n\rho(x) + 1\rho(x) = (n + 1)\rho(x).$$

We have proved by induction that

$$\rho(nx) = n\rho(x) \quad \text{for all } n \in \mathbb{N} \text{ and } x \in F. \tag{143.8}$$

From (143.3) and (143.5) it follows that ρ is isotone. Since $e \in P$ (Proposition 142D,(b)), (143.3) yields $\rho(e) \in P$.

2. Let $x \in P$ be given. Suppose first that $\rho(e) = 0$. Since F is archimedean (Proposition 143D), we may choose $n \in \mathbb{N}$ such that $0 \leq x \leq ne$. By (143.6), (143.8), and the isotonicity of ρ we find

$$0 = \rho(0) \leq \rho(x) \leq \rho(ne) = n\rho(e) = n0 = 0.$$

We have proved that $\rho(x) = 0 = x\rho(e)$.

Suppose next that $\rho(e) \neq 0$. Let $p, q \in \mathbb{N}^\times$ be given. Since ρ is isotone, it follows from (143.8) that $pe \leq qx$ implies $(pe)\rho(e) = p\rho(e) = \rho(pe) \leq \rho(qx) = q\rho(x)$, and hence $pe \leq (q\rho(x))/\rho(e) = q(\rho(x)/\rho(e))$; and, similarly, $pe \geq qx$ implies $pe \geq q(\rho(x)/\rho(e))$. We have therefore neither $qx < pe < q(\rho(x)/\rho(e))$ nor $qx > pe > q(\rho(x)/\rho(e))$. Since $p, q \in \mathbb{N}^\times$ were arbitrary and F is archimedean, it follows from Proposition 142J that $x = \rho(x)/\rho(e)$; thus $\rho(x) = x\rho(e)$ in this case too.

We have shown that $\rho(x) = x\rho(e)$ for all $x \in P$. It follows from (143.7) and (OF4) that $\rho(x) = x\rho(e)$ for all $x \in F$. Thus (143.4) holds. ■

143F. Corollary. *Let the complete ordered field F and the mapping $\rho : F \to F$ be given. Then ρ satisfies* (143.2) *and*

$$\rho(x^2) = \rho(x)^2 \quad \textit{for all } x \in P \tag{143.9}$$

(if and) only if $\rho = 0_{F \to F}$ or $\rho = 1_F$.

Proof. By Propositions 143C and 142D,(a), (143.9) implies (143.3). In view of Proposition 143E, it is now sufficient to show that $\rho(e) = 0$ or $\rho(e) = e$, where e is the unity of F. Since $e \in P$, (143.9) yields

$$(\rho(e) - e)\rho(e) = \rho(e)^2 - e\rho(e) = \rho(e^2) - \rho(e) = \rho(e) - \rho(e) = 0,$$

and therefore $\rho(e) = 0$ or $\rho(e) = e$. ■

144. Essential uniqueness of complete ordered fields

The aim of this section is to prove that any two complete ordered fields are naturally isomorphic, and thus have essentially the same structure, in the very strong sense described in the following main result.

144A. Theorem. *Let the complete ordered fields F', F'' be given, with respective zeros $0'$, $0''$, unities e', e'', and positive halves P', P''. There exists exactly one mapping $\phi' : F' \to F''$ satisfying*

$$\phi'(x' + y') = \phi'(x') + \phi'(y') \quad \textit{for all } x', y' \in F' \tag{144.1}$$

$$\phi'(x'^2) = \phi'(x')^2 \quad \textit{for all } x' \in P' \tag{144.2}$$

$$\phi' \neq 0''_{F' \to F''}. \tag{144.3}$$

There exists exactly one mapping $\phi'' : F'' \to F'$ satisfying

$$\phi''(x'' + y'') = \phi''(x'') + \phi''(y'') \quad \textit{for all } x'', y'' \in F'' \tag{144.4}$$

$$\phi''(x''^2) = \phi''(x'')^2 \quad \textit{for all } x'' \in P'' \tag{144.5}$$

$$\phi'' \neq 0'_{F'' \to F'} \tag{144.6}$$

Each of the mappings ϕ', ϕ'' is the inverse of the other. Each is an order-isomorphism, and they satisfy

$$\begin{aligned} &\phi'(-x') = -\phi'(x') \quad \textit{for all } x' \in F' \\ &\phi''(-x'') = -\phi''(x'') \quad \textit{for all } x'' \in F'' \end{aligned} \tag{144.7}$$

$$\phi'(0') = 0'' \qquad \phi''(0'') = 0' \tag{144.8}$$

$$\phi'(e') = e'' \qquad \phi''(e'') = e' \tag{144.9}$$

$$\begin{aligned} &\phi'(x'y') = \phi'(x')\phi'(y') \quad \textit{for all } x', y' \in F' \\ &\phi''(x''y'') = \phi''(x'')\phi''(y'') \quad \textit{for all } x'', y'' = F'' \end{aligned} \tag{144.10}$$

$$\phi'_>(P') = P'' \qquad \phi''_>(P'') = P'. \tag{144.11}$$

For the proof of this theorem we shall rely on three lemmas. In each of them, F and F_1 are given complete ordered fields, with respective zeros 0 and 0_1, unities e and e_1, and positive halves P and P_1.

144B. Lemma. *There exists a mapping $\omega : P \to P_1$ such that*

$$\forall x \in P,\ \forall m, n \in \mathbb{N}^\times, \begin{cases} me \le nx \ \Rightarrow\ me_1 \le n\omega(x) \\ \\ me \ge nx \ \Rightarrow\ me_1 \ge n\omega(x). \end{cases} \tag{144.12}$$

Proof. Let $x \in P$ be given. Since F is archimedean (Proposition 143D), there exist $m', n' \in \mathbb{N}^\times$ such that $m'e \ge n'x$ (even with $n' := 1$). For all such m', n', and for all $m, n \in \mathbb{N}^\times$ with $me \le nx$, we have, by (142.16), $n'me \le n'nx = nn'x \le nm'e$; it follows by (142.17) that $n'm \le nm'$, and consequently $(n'e_1)(me_1) = n'me_1 \le nm'e_1 = (ne_1)(m'e_1)$; and hence, by (142.9), $(me_1)/(ne_1) \le (m'e_1)/(n'e_1)$. We conclude that the subset $\{(me_1)/(ne_1) \mid m, n \in \mathbb{N}^\times,\ me \le nx\}$ of P_1 has an upper bound. Since P_1 is pre-completely ordered (Lemma 143A) and has the minimum 0_1, this subset has a supremum.

We may therefore define $\omega : P \to P_1$ by the rule

$$\omega(x) := \sup\{(me_1)/(ne_1) \mid m, n \in \mathbb{N}^\times,\ me \le nx\} \quad \text{for all } x \in P.$$

Let $x \in P$ and $m', n' \in \mathbb{N}^\times$ be given. If $m'e \le n'x$, then $(m'e_1)/(n'e_1) \le \omega(x)$, and hence, by (142.8), $m'e_1 \le (n'e_1)\omega(x) = n'\omega(x)$. If, on the other hand, $m'e \ge n'x$, we showed in the preceding paragraph that $(me_1)/(ne_1) \le (m'e_1)/(n'e_1)$ for *all* $m, n \in \mathbb{N}^\times$ with $me \le nx$; and therefore $\omega(x) \le (m'e_1)/(n'e_1)$, and consequently $m'e_1 \ge (n'e_1)\omega(x) = n'\omega(x)$. Thus (144.12) is established. ∎

144C. Lemma. *Let $\omega : P \to P_1$ be a mapping that satisfies* (144.12). *Then $\omega(e) = e_1$ and*

$$\omega(x + y) = \omega(x) + \omega(y) \quad \textit{for all } x, y \in P \tag{144.13}$$

$$\omega(xy) = \omega(x)\omega(y) \quad \textit{for all } x, y \in P. \tag{144.14}$$

Proof. 1. We have $1e = 1e$. It follows from (144.12) that $e_1 = 1e_1 = 1\omega(e) = \omega(e)$.

Let $x, y \in P$ and $m, n \in \mathbb{N}^\times$ be given. Assume first that $me_1 < n\omega(x + y)$. It follows from (144.12) that $me < n(x + y)$. If $me < nx$, we set $a := m,\ b := 0,\ c := 1$. If $nx \le me < ny$, we set $a := 0,\ b := m,\ c := 1$. If $me \ge nx$ and $me \ge ny$, we have, by (142.6), $me - ny < nx \le me$. Since F is archimedean, Proposition 142J permits us to choose $p, q \in \mathbb{N}^\times$ such that $q(me - ny) < pe < qnx \le qme$ (the last inequality by (142.16)); therefore $p \le qm$, by (142.17), and we set $a := p,\ b := qm - p,\ c := q$. In all three cases, $c \in \mathbb{N}^\times$, and $ae \le cnx,\ be\ \le cny$, and $a + b = cm$. Then (144.12) yields $ae_1 \le cn\omega(x)$, $be_1 \le cn\omega(y)$ (even when $a = 0$ or $b = 0$), and therefore, by (142.6),

$$cme_1 = ae_1 + be_1 \le cn\omega(x) + cn\omega(y) = cn(\omega(x) + \omega(y));$$

consequently, by (142.16), $me_1 \le n(\omega(x)+\omega(y))$.

Assume next that $me_1 > n\omega(x+y)$. It follows from (144.12) that $me > n(x+y)$. Then $0 \le ny < me$ and $nx < me - ny \le me$. By Proposition 142J we may choose $p, q \in \mathbb{N}^\times$ such that $qnx < pe < q(me-ny) \le qme$. Then $p \le qm$ and $qnx < pe$ and $qny < (qm-p)e$. We infer from (144.12) that $pe_1 \ge qn\omega(x)$ and $(qm-p)e_1 \ge qn\omega(y)$, and therefore

$$qme_1 = pe_1 + (qm-p)e_1 \ge qn\omega(x) + qn\omega(y) = qn(\omega(x+\omega(y));$$

consequently, by (142.16), $me_1 \ge n(\omega(x)+\omega(y))$.

We infer from the preceding two paragraphs that we have neither $n(\omega(x)+\omega(y)) < me_1 < n\omega(x+y)$ nor $n(\omega(x)+\omega(y)) > me_1 > n\omega(x+y)$. Since $m, n \in \mathbb{N}^\times$ were arbitrary and F_1 is archimedean, it follows from Proposition 142J that $\omega(x)+\omega(y) = \omega(x+y)$. Since $x, y \in P$ were arbitrary, (144.13) is established.

2. From (144.13) we have $\omega(0) = \omega(0+0) = \omega(0)+\omega(0)$, and hence $\omega(0) = \omega(0)-\omega(0) = 0$. Therefore $\omega(0y) = \omega(0) = 0 = 0\omega(y) = \omega(0)\omega(y)$ for all $y \in P$.

Let $x \in P^\times$, $y \in P$, and $m, n \in \mathbb{N}^\times$ be given. Assume first that $me_1 < n\omega(xy)$. It follows from (144.12) that $me < nxy$, and hence $(me)/x < ny$. By Proposition 142J we may choose $p, q \in \mathbb{N}^\times$ such that $(qme)/x = q((me)/x) < pe < qny$. Then $qme < px$ and $pe < qny$; from (144.12) we infer that $qme_1 \le p\omega(x)$ and $pe_1 \le qn\omega(y)$; by (142.8) it follows that $pqme_1 \le pqn\omega(x)\omega(y)$ and hence, by (142.16), $me_1 \le n\omega(x)\omega(y)$. By exactly the same argument with the inequalities reversed it follows that $me_1 > n\omega(xy)$ implies $me_1 \ge n\omega(x)\omega(y)$. We thus have neither $n\omega(x)\omega(y) < me_1 < n\omega(xy)$ nor $n\omega(x)\omega(y) > me_1 > n\omega(xy)$. Since $m, n \in \mathbb{N}^\times$ were arbitrary and F_1 is archimedean, Proposition 142J yields that $\omega(xy) = \omega(x)\omega(y)$. Since $x \in P^\times$ and $y \in P$ were arbitrary, and since we have shown before that $\omega(0y) = \omega(0)\omega(y)$ for all $y \in P$, (144.14) is established. ∎

144D. Lemma. *Let $\omega : P \to P_1$ be a mapping that satisfies* (144.13) *and* (144.14). *Define $\psi : F \to F_1$ by the rule*

$$\psi(x) := \begin{cases} \omega(x) & \text{if } x \in P \\ -\omega(-x) & \text{if } x \in F \backslash P = -P^\times. \end{cases} \tag{144.15}$$

Then

$$\psi(-x) = -\psi(x) \quad \text{for all } x \in F \tag{144.16}$$

$$\psi(x+y) = \psi(x) + \psi(y) \quad \text{for all } x, y \in F \tag{144.17}$$

$$\psi(xy) = \psi(x)\psi(y) \quad \text{for all } x, y \in F. \tag{144.18}$$

Proof. (144.16) is an immediate consequence of (144.15) and the fact that $\omega(0) = 0$ (which follows from (144.13) as in the preceding proof). (144.18) follows from

(144.14), (144.15), (144.16): for example, if $x \in P$, $y \in F\backslash P = -P^{\times}$, we have $-(xy) = x(-y) \in PP \subset P$, and $\psi(xy) = -\omega(-(xy)) = -\omega(x(-y)) = -\omega(x)\omega(-y) = \omega(x)(-\omega(-y)) = \psi(x)\psi(y)$.

We claim that

$$\forall x, y, z \in F, \quad x + y + z = 0 \;\Rightarrow\; \psi(x) + \psi(y) + \psi(z) = 0_1. \tag{144.19}$$

In view of the symmetry of (144.19) and on account of (OF4), it is enough to prove (144.19) under the additional restriction that $x, y \in P$ or $x, y \in -P$. The latter case is reduced to the former by replacing $x,\ y,\ z$ by $-x,\ -y,\ -z$, respectively, and applying (144.16). Now if $x, y \in P$ and $x + y + z = 0$, then $-z = x + y \in P + P \subset P$, and (144.15), (144.13), (144.16) yield

$$\psi(x) + \psi(y) + \psi(z) = \omega(x) + \omega(y) - \psi(-z) = \omega(x + y) - \omega(x + y) = 0_1,$$

as claimed, thus establishing (144.19).

For all $x, y \in F$ we have $x + y + (-(x + y)) = 0$. It follows, using (144.16) and (144.19), that

$$\psi(x) + \psi(y) - \psi(x + y) = \psi(x) + \psi(y) + \psi(-(x + y)) = 0_1 \quad \text{for all } x, y \in F,$$

and thus (144.17) is also established. ■

Proof of Theorem 144A. 1. Relying on Lemmas 144B, 144C, 144D with $F := F'$ and $F_1 := F''$, and again with $F := F''$ and $F_1 := F'$, we may choose mappings $\psi'' : F' \to F''$ and $\psi'' : F'' \to F'$ that satisfy

$$\psi'(e') = e'' \qquad\qquad \psi''(e'') = e' \tag{144.20}$$

$$\psi'_{>}(P') \subset P'' \qquad\qquad \psi''_{>}(P'') \subset P' \tag{144.21}$$

$$\begin{aligned} &\psi'(-x') = -\psi'(x') \quad \text{for all } x' \in F' \\ &\psi''(-x'') = -\psi''(x'') \quad \text{for all } x'' \in F'' \end{aligned} \tag{144.22}$$

$$\begin{aligned} &\psi'(x' + y') = \psi'(x') + \psi'(y') \quad \text{for all } x', y' \in F' \\ &\psi''(x'' + y'') = \psi''(x'') + \psi''(y'') \quad \text{for all } x'', y'' \in F'' \end{aligned} \tag{144.23}$$

$$\begin{aligned} &\psi'(x'y') = \psi'(x')\psi'(y') \quad \text{for all } x', y' \in F' \\ &\psi''(x''y'') = \psi''(x'')\psi''(y'') \quad \text{for all } x'', y'' \in F''. \end{aligned} \tag{144.24}$$

We observe that $\phi' := \psi'$ and $\phi'' := \psi''$ satisfy (144.1), (144.2), (144.3), (144.4), (144.5) (144.6), (144.7), (144.8) (from (144.7)), (144.9), and (144.10).

From (144.23) we have

$$(\psi'' \circ \psi')(x' + y') = \psi''(\psi'(x' + y')) = \psi''(\psi'(x') + \psi'(y')) = \psi''(\psi'(x')) + \psi''(\psi'(y')) =$$

$$= (\psi'' \circ \psi')(x') + (\psi'' \circ \psi')(y') \quad \text{for all } x', y' \in F';$$

and from (144.24) a similar computation yields

$$(\psi'' \circ \psi')(x'^2) = ((\psi'' \circ \psi')(x'))^2 \quad \text{for all } x' \in P'.$$

From (144.20) we have $(\psi'' \circ \psi')(e') = \psi''(e'') = e' \neq 0'$. Applying Corollary 143F to $F := F'$ and $\rho := \psi'' \circ \psi'$, we conclude that $\psi'' \circ \psi' = 1_{F'}$.

Repeating the argument with F' and F'', ψ' and ψ'' interchanged, we conclude that $\psi' \circ \psi'' = 1_{F''}$. Thus ψ' and ψ'' are invertible, and each is the inverse of the other. Therefore equality must hold in (144.21).

By (144.21), (144.22), (144.23), ψ' and ψ'' are isotone, and therefore they are order-isomorphisms. Thus $\phi' := \psi'$ and $\phi'' := \psi''$ are order-isomorphisms, each is the inverse of the other, and they satisfy (144.11).

2. It remains to show that $\phi' := \psi'$ and $\phi'' := \psi''$ are the *only* mappings satisfying (144.1), (144.2), (144.3) and (144.4), (144.5), (144.6).

Let $\phi'' : F'' \to F'$ be a mapping that satisfies (144.4), (144.5), (144.6). From (144.4) and (144.23), and from (144.5) and (144.24), we obtain, by computations similar to the preceding ones,

$$(\phi'' \circ \psi')(x' + y') = (\phi'' \circ \psi')(x') + (\phi'' \circ \psi')(y') \quad \text{for all } x', y' \in F'$$

$$(\phi'' \circ \psi')(x'^2) = ((\phi'' \circ \psi')(x'))^2 \quad \text{for all } x' \in P'.$$

By (144.6) we have $(\phi'' \circ \psi') \circ \psi'' = \phi'' \circ (\psi' \circ \psi'') = \phi'' \circ 1_{F''} = \phi'' \neq 0'_{F'' \to F'}$, and therefore $\phi'' \circ \psi' \neq 0'_{F' \to F'}$. We may therefore apply Corollary 143F with $F := F'$ and $\rho := \phi'' \circ \psi'$ and find that $\phi'' \circ \psi' = 1_{F'}$. Thus ϕ'' is a left-inverse of the invertible mapping ψ', and therefore ϕ'' must be ψ'', the inverse of ψ'.

If $\phi' : F' \to F''$ satisfies (144.1), (144.2), (144.3), the same argument, with F' and F'', ψ' and ψ'' interchanged, shows that $\phi' = \psi'$. ■

This page intentionally left blank

Chapter 15

THE REAL-NUMBER SYSTEM

151. The Real-Number System

The real numbers, the operations with them, and the relations among them constitute a fundamental cluster of notions in Mathematics. It is plain that some essential properties of the system of real numbers are summarized by stating that it constitutes an ordered field. Long before the concept was formalized, it was realized that more was required: a careful analysis of the use made of real numbers, such as originates with the theory of ratios of Eudoxos, demands that this ordered field be *complete*.

It might seem that with this specification we have not exhausted the demands implicit in the structure of the system: we have not yet accounted, say, for notions of distance, of nearness, of continuity, etc. Our "essential-uniqueness result", Theorem 144A shows us, however, that such additional demands cannot require a choice among complete ordered fields, since these are all naturally isomorphic; the additional notions will have to be accommodated by deriving them from the complete-ordered-field structure itself.

Concerning the question of *existence* of complete ordered fields, we refer to the introductory comments in Section 141 and the discussion in Chapter 16. We shall lose no further sleep over this question here.

With the same turn of phrase used concerning the natural numbers we say: We shall adopt, or pretend to adopt, the naive view that one specific complete ordered field is revealed to us, or singled out by us, as the **Real-Number System**, to be denoted by $\mathbb{R}$, and its members to be known as **real numbers**.

We should like to work in a mathematical world in which, justifying common usage, the natural numbers are themselves real numbers. We achieve this as follows. In an ordered field F with unity e there is the counting system N described by Theorem 142E. We observe that neither the definition of complete ordered field, nor the construction of N with its zero and successor-mapping, nor the proof of Theorem 142E makes any appeal to the Natural-Number System. We also recall that *any* counting system could serve as the Natural-Number System. We therefore take it for granted that *the counting system described by Theorem 142E for the special (complete) ordered field* $\mathbb{R}$ *is in fact the Natural-Number System* $\mathbb{N}$ *itself.* This state of affairs

may be interpreted as part of the revelation, or the choice, that gives us both systems; or it may, less satisfyingly, be regarded as a deliberate change in recognizing which counting system is *the* Natural-Number System.

With this state of affairs granted, the unity of $\mathbb{R}$ is the natural number 1, and the nth multiple of this unity is the natural number n for each $n \in \mathbb{N}$.

Of course this state of affairs would be anything but desirable if there were a clash between the operations and relations defined for the natural numbers on the one hand, and the similarly designated operations and relations defined for the real numbers, as adjusted to the subset of natural numbers, on the other. Proposition 142F shows that there is no such clash as regards order, zero, unity, addition, multiplication, power formation (exponentiation), and subtraction (so far as meaningful in $\mathbb{N}$). Proposition 132D,(c) shows that, for $n \in \mathbb{N}$ and $r \in \mathbb{R}$, there is no clash between nr, the nth multiple of the real number r, and nr, the product of the real numbers n and r.

Most notation and terminology for $\mathbb{R}$ is the same as that introduced for all (complete) ordered fields. We mention two items of specialized notation. The positive half of $\mathbb{R}$ will be denoted by $\mathbb{P}$. Order-intervals in $\mathbb{R}$ are written with ordinary (square) brackets: for given $a, b \in \mathbb{R}$ with $a \leq b$ we set

$$\begin{aligned} [a,b] &:= [\![a,b]\!] = \{t \in \mathbb{R} \mid a \leq t \leq b\} \\ [a,b[&:= [\![a,b[\!] = \{t \in \mathbb{R} \mid a \leq t < b\} \\ ]a,b] &:=]\!]a,b]\!] = \{t \in \mathbb{R} \mid a < t \leq b\} \\ ]a,b[&:=]\!]a,b[\!] = \{t \in \mathbb{R} \mid a < t < b\}. \end{aligned}$$

In much mathematical writing one finds one of a pair of parentheses used instead of the everted bracket: $[a,b), (a,b], (a,b)$ instead of $[a,b[,\ ]a,b],\ ]a,b[$. This notation is less suggestive than the one we adhere to, and is actually confusing: (a,b) already denotes a *pair*.

Every order-convex subset of $\mathbb{R}$ is called an **interval**. This terminology involves a mild clash with the term *order-interval* as introduced in Section 61, but is too entrenched (and useful) to be discarded. The clash will, moreover, be almost entirely eliminated when all intervals will be shown, in the next section, to actually be order-intervals in the *Extended-Real-Number System.*

151A. Proposition. *A subset of $\mathbb{R}$ is an interval if and only if it is an order-interval of $\mathbb{R}$, or $\mathbb{R}$ itself, or one of the sets $a + \mathbb{P},\ a + \mathbb{P}^\times,\ a - \mathbb{P},\ a - \mathbb{P}^\times$ for some $a \in \mathbb{R}$.*

Proof. This is an immediate consequence of Proposition 72J((i) $\Rightarrow$ (iv)) and the definition of the order $\leq$ in $\mathbb{R}$. ■

An interval is said to be **bounded** if it is order-bounded; it follows from Proposition 151A that this is the case precisely when it is an order-interval. An interval is said to be **genuine** if it is neither empty nor a singleton.

151B. Proposition. *Every genuine interval is an uncountable set. In particular,* $\mathbb{R}$ *is uncountable.*

Proof. This is a reformulation of Proposition 143B. (See also Example 122H.) ■

We set $\mathbb{Z} := \mathbb{N} \cup (-\mathbb{N})$. The members of $\mathbb{Z}$ are called **integers** (sometimes they are called *whole numbers*). A real number is said to be **integral** if it is in $\mathbb{Z}$. The natural numbers are thus the positive integers. We show that $\mathbb{Z}$ is stable under addition, opposition, and multiplication; this will enable us to adjust these operations to $\mathbb{Z}$.

151C. Proposition. $\mathbb{Z} + \mathbb{Z} = \mathbb{Z},\ -\mathbb{Z} = \mathbb{Z},\ \mathbb{Z}\mathbb{Z} = \mathbb{Z}$.

Proof. We have $\mathbb{N} + \mathbb{N} \subset \mathbb{N}, \mathbb{N}\mathbb{N} \subset \mathbb{N}$, and $0, 1 \in \mathbb{N} \subset \mathbb{Z}$. The assertion will therefore be established once we prove that $\mathbb{N} - \mathbb{N} \subset \mathbb{Z}$. Let $m, n \in \mathbb{N}$ be given. If $m \geq n$, we have $m - n \in \mathbb{N} \subset \mathbb{Z}$; if $m < n$, we have $m - n = -(n - m) \in -\mathbb{N} \subset \mathbb{Z}$. ■

151D. Corollary. $\mathbb{Z}$ *is a commutative ring with* 0 *as zero,* 1 *as unity, and addition, opposition, and multiplication adjusted from the corresponding operations in* $\mathbb{R}$. *This commutative ring is not a field.*

Proof. All but the last statement is obvious in view of Proposition 151C. For every $n \in \mathbb{Z}^\times$ we have $|2n| = 2|n| \geq 2 \cdot 1 = 2 > 1$, and therefore there is no $n \in \mathbb{Z}$ such that $2n = 1$. ■

It is convenient to extend to $\mathbb{Z}$ the notation introduced by (92.9) for $\mathbb{N}$:

$$m..n := \{k \in \mathbb{Z} \mid m \leq k \leq n\} \quad \text{for all } m, n \in \mathbb{Z}$$

(see comments following (92.9) to explain this choice of notation).

We are now able to extend the notion of natural multiples (Section 115) in a commutative *group* to that of *integral multiples*. Let the commutative group G, written additively, be given. We define

$$(-n)x := -nx = n(-x) \qquad \text{for every } x \in G \text{ and } n \in \mathbb{N}^\times,$$

thus completing the definition of nx for all $n \in \mathbb{Z}$ and $x \in G$. The following result is obtained by straightforward verification.

151E. Proposition. *The statements of Corollary 131D and Proposition 132,*(c) *remain valid when* $\mathbb{N}$ *is replaced by* $\mathbb{Z}$.

In a commutative group G written multiplicatively we have, correspondingly, the notion of *powers* x^n for all $n \in \mathbb{Z}$ and $x \in G$, obtained by defining

$$x^{-n} := e/x^n = (e/x)^n \qquad \text{for every } x \in G \text{ and } n \in \mathbb{N}^\times$$

(where e is the unity of G). This notation is applicable, in particular, to the multiplicative group of a field, and more in particular to the multiplicative group $\mathbb{R}^\times$ of $\mathbb{R}$. We note that, by (142.4) and Proposition 142D,(b), we have

$$r^n \in \mathbb{R}^\times \qquad \text{for all } n \in \mathbb{Z} \text{ and } r \in \mathbb{R}^\times. \tag{151.1}$$

We introduce two integer-valued mappings from $\mathbb{R}$ to $\mathbb{R}$.

151F. Proposition. *Let $r \in \mathbb{R}$ be given. The set $\mathbb{Z} \cap (r + \mathbb{P}) = \mathbb{Z} \cap \mathrm{Ub}(\{r\})$ has a minimum and the set $\mathbb{Z} \cap (r - \mathbb{P}) = \mathbb{Z} \cap \mathrm{Lb}(\{r\})$ has a maximum.*

Proof. Assume that $r \in \mathbb{P}$. The set $\mathbb{Z} \cap (r + \mathbb{P}) = \mathbb{Z} \cap \mathbb{P} \cap (r + \mathbb{P}) = \mathbb{N} \cap \mathrm{Ub}(\{r\})$ is a non-empty subset of $\mathbb{N}$, since $\mathbb{R}$ is archimedean; it therefore has a minimum. The set $\mathbb{N} \cap (r - \mathbb{P}) = \mathbb{N} \cap [0, r]$ is finite (the preceding minimum is an upper bound in $\mathbb{N}$) and it contains 0. Therefore $\mathbb{N} \cap (r - \mathbb{P})$ has a maximum, and this is also a maximum of $\mathbb{Z} \cap (r - \mathbb{P})$.

Opposition in $\mathbb{R}$ is an order-antimorphism that maps $\mathbb{Z}$ onto $\mathbb{Z}$ (Proposition 151C). If $r \in -\mathbb{P}$, therefore, $-\max(\mathbb{Z} \cap (-r - \mathbb{P})) = -\max(-(\mathbb{Z} \cap (r + \mathbb{P})))$ is a minimum of $\mathbb{Z} \cap (r + \mathbb{P})$, and $-\min(\mathbb{Z} \cap (-r + \mathbb{P})) = -\min(-(\mathbb{Z} \cap (r - \mathbb{P})))$ is a maximum of $\mathbb{Z} \cap (r - \mathbb{P})$. ∎

By virtue of Proposition 151F we may define the functions $\lceil, \lfloor \in \mathrm{Map}(\mathbb{R}, \mathbb{R})$ by the rules

$$\lceil r := \min(\mathbb{Z} \cap (r + \mathbb{P})), \quad \lfloor r := \max(\mathbb{Z} \cap (r - \mathbb{P})) \quad \text{for all } r \in \mathbb{R}. \tag{151.2}$$

151G. Proposition. *The functions $\lceil$, $\lfloor$ have the following properties:*

$$\lceil \text{ and } \lfloor \text{ are isotone and idempotent, and } \mathrm{Rng}\lceil = \mathrm{Rng}\lfloor = \mathbb{Z}. \tag{151.3}$$

$$\lceil(-r) = -\lfloor r \quad \text{for all } r \in \mathbb{R}. \tag{151.4}$$

$$(\lceil r = n \;\Leftrightarrow\; r \in]n-1, n]) \quad \text{and} \quad (\lfloor r = n \;\Leftrightarrow\; r \in [n, n+1[) \quad \text{for all } r \in \mathbb{R} \text{ and } n \in \mathbb{Z}. \tag{151.5}$$

$$r - 1 < \lfloor r \leq r \leq \lceil r < r + 1 \quad \text{for all } r \in \mathbb{R}. \tag{151.6}$$

$$\lceil(2r) \in \{2\lceil r, 2\lceil r - 1\} \quad \text{and} \quad \lfloor(2r) \in \{2\lfloor r, 2\lfloor r + 1\} \quad \text{for all } r \in \mathbb{R}. \tag{151.7}$$

Proof. We shall prove only one-half of (151.7). Let $r \in \mathbb{R}$ be given. Then (151.5) yields $\lceil r - 1 < r \leq \lceil r$. By Proposition 142C, $2\lceil r - 2 < 2r \leq 2\lceil r$. Therefore $\lceil(2r) = 2\lceil r$ or $\lceil(2r) = 2\lceil r - 1$ according as $2r > 2\lceil r - 1$ or $2r \leq 2\lceil r - 1$. ∎

On account of (151.2) or (151.5), $\lceil r$ and $\lfloor r$ are called the **ceiling of** r and the **floor of** r, respectively, for every $r \in \mathbb{R}$. The functions $\lceil$ and $\lfloor$ themselves are called the **ceiling-function** and the **floor-function**, respectively. (They are also known as the *least-integer function* and the *greatest-integer function*. The notation $[r]$ instead of $\lfloor r$ is often encountered; we do not use it.)

We set $\mathbb{Q} := \{m/n \mid (m, n) \in \mathbb{Z} \times \mathbb{N}^\times\} = \{m/n \mid (m, n) \in \mathbb{Z} \times \mathbb{Z}^\times\}$. A real number is said to be **rational** or **irrational**, and is called a **rational number** or an **irrational number**, according as it is in $\mathbb{Q}$ or not. We record the fact that $\mathbb{Q}$ is stable under addition, opposition, and multiplication, as well as division by non-zero numbers; this will enable us to adjust these operations to $\mathbb{Q}$.

151H. Proposition. $\mathbb{Q} + \mathbb{Q} = \mathbb{Q}, -\mathbb{Q},\ \mathbb{Q}\mathbb{Q} = \mathbb{Q}$. *Moreover,*

$$p/q \in \mathbb{Q} \qquad \textit{for all } (p, q) \in \mathbb{Q} \times \mathbb{Q}^\times.$$

Proof. This follows from Corollary 133E. ■

151I. Corollary. $\mathbb{Q}$ *is an archimedean ordered field with* 0 *as zero,* 1 *as unity,* $\mathbb{Q} \cap \mathbb{P}$ *as positive half, and addition, opposition, and multiplication adjusted from the corresponding operations in* $\mathbb{R}$.

We next show that between any two distinct real numbers there is always a rational number.

151J. Proposition. *Let* $s, t \in \mathbb{R}$ *be given. If* $s < t$*, then there exists* $q \in \mathbb{Q}$ *such that* $s < q < t$.

Proof. Since $-\mathbb{Q} = \mathbb{Q}$ and $0 \in \mathbb{Q}$, it is sufficient to prove the assertion when $s, t \in \mathbb{P}$. But then the conclusion follows immediately from Proposition 142J and (142.8), since $\mathbb{R}$ is archimedean. ■

151K. Remarks. (a): The set $\mathbb{Q}$ is infinite, since it includes the infinite set $\mathbb{N}$. It is shown in Example 121R that $\mathbb{Q}$ is countable.

(b): Since $\mathbb{R}$ is uncountable (Proposition 151B) and $\mathbb{Q}$ is countable, it follows that there exists an irrational number. It then follows at once from Proposition 142J that between any two distinct real numbers there is an irrational number. It is actually not difficult to produce a specific irrational number: e.g., $\sqrt{2}$ is irrational. For completeness (no pun intended) we include a proof of this fact. ■

151L. Proposition. *For all* $p, q \in \mathbb{N}^\times,\ 2q^2 \neq p^2$*. Consequently,* $\sqrt{2}$ *is irrational.*

Proof. Set $A := \{q \in \mathbb{N}^\times \mid \exists p \in \mathbb{N}^\times,\ 2q^2 = p^2\}$. We are to show that $A = \emptyset$.

Let $q \in A$ be given. We may choose $p \in \mathbb{N}^\times$ such that

$$2q^2 = p^2. \tag{151.8}$$

Since $1 < 2 < 2 \cdot 2 = 4$, we find

$$q^2 < 2q^2 = p^2 < 4q^2 < 4p^2.$$

Since the mapping $x \mapsto x^2 : \mathbb{P} \to \mathbb{P}$ is strictly isotone (Proposition 142D,(c)), we conclude that

$$q < p < 2q < 2p.$$

From this we obtain $q' := 2q - p \in \mathbb{N}^\times$ and $p' := 2(p - q) \in \mathbb{N}^\times$, and $q' < 2q - q = q$. From (151.8) we also have

$$2q'^2 = 2 \cdot 4q^2 - 2 \cdot 4qp + 2p^2 = 4p^2 - 2 \cdot 4pq + 4q^2 = p'^2.$$

We conclude that $q' \in A$, but $q' < q$. Thus A has not a minimum, and hence is empty. ■

152. The Extended-Real-Number System

We observed at the beginning of Section 143 that the complete ordered field $\mathbb{R}$ has no maximum; since opposition is an order-antimorphism from $\mathbb{R}$ to $\mathbb{R}$, $\mathbb{R}$ has no minimum either. It is sometimes convenient to regard the pre-completely ordered set $\mathbb{R}$ as an order-interval of a completely ordered set, as described in Example 72I,(b), by relying on Theorem 72G. To that end we choose, or assume to be specified, a doubleton $\{\infty, \infty'\}$ that is disjoint from $\mathbb{R}$, and set $\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty, \infty'\}$; we then define in $\bar{\mathbb{R}}$ a relation, again denoted by $\leq$, by requiring that its restriction to $\mathbb{R}$ be the order of $\mathbb{R}$ and that $\infty' \leq t$ and $t \leq \infty$ for all $t \in \bar{\mathbb{R}}$. Then $\bar{\mathbb{R}}$ is totally and completely ordered by $\leq$, and $\infty' = \min\bar{\mathbb{R}} = \inf\mathbb{R}$ and $\infty = \max\bar{\mathbb{R}} = \sup\mathbb{R}$. The corresponding strict-order is again denoted by $<$.

It is desirable to extend to $\bar{\mathbb{R}}$, or as far as possible into $\bar{\mathbb{R}}$, the algebraic operations in $\mathbb{R}$, while preserving the validity of as many as possible of their basic properties. Thus, we define the **opposition** $(t \mapsto -t) : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ by requiring that it agree with opposition on $\mathbb{R}$ and satisfy $-\infty := \infty', -\infty' := \infty$. From now on we shall always write $-\infty$ for ∞'. We define the **addition** $((s,t) \mapsto s+t) : (\bar{\mathbb{R}} \times \bar{\mathbb{R}}) \backslash \{(\infty, -\infty), (-\infty, \infty)\} \to \bar{\mathbb{R}}$ by requiring that it agree with addition on $\mathbb{R} \times \mathbb{R}$ and satisfy

$$\infty + t := t + \infty := \infty \qquad \text{for all } t \in \mathbb{R} \cup \{\infty\}$$

$$-\infty + t := t + (-\infty) := -\infty \qquad \text{for all } t \in \mathbb{R} \cup \{-\infty\};$$

the **subtraction** $((s,t) \mapsto s-t) : (\bar{\mathbb{R}} \times \bar{\mathbb{R}}) \backslash \{(\infty, \infty), (-\infty, -\infty)\} \to \bar{\mathbb{R}}$ by the rule $s - t := s + (-t)$ for all (s,t) in the domain; and the **multiplication** $((s,t) \mapsto st) : \bar{\mathbb{R}} \times \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ by requiring that it agree with multiplication on $\mathbb{R} \times \mathbb{R}$ and satisfy

$$\infty \cdot t := t \cdot \infty := \begin{cases} \infty & \text{if } t > 0 \\ 0 & \text{if } t = 0 \\ -\infty & \text{if } t < 0 \end{cases} \qquad \text{for all } t \in \bar{\mathbb{R}}$$

$$(-\infty) \cdot t := t \cdot (-\infty) := \begin{cases} -\infty & \text{if } t > 0 \\ 0 & \text{if } t = 0 \\ \infty & \text{if } t < 0 \end{cases} \qquad \text{for all } t \in \bar{\mathbb{R}}.$$

It must be noted that many results (e.g., concerning limits) that involve multiplication in $\bar{\mathbb{R}}$ are valid only if the domain of multiplication is restricted to exclude multiplication of 0 by either ∞ or $-\infty$.

It may be seen that the associative, commutative, and neutrality laws for both addition and multiplication, as well as the distributive law, are preserved; but note that

the domain of addition is *not* $\bar{\mathbb{R}}\times\bar{\mathbb{R}}$, so that we have not a commutative ring. The formulas (131.1), (131.2), (131.3), (131.4), (131.5), (132.1), (132.2), (132.3) also remain valid in $\bar{\mathbb{R}}$. The cancellation laws fail. The strict monotonicity laws in Proposition 142C and Corollary 142G,(b) fail, but their lax analogues remain valid. Notations such as $A^\times$, $A+B$, $A-B$, AB for subsets A, B of $\bar{\mathbb{R}}$ are defined as in Sections 131 and 132, so long as the definition remains meaningful.

The completely ordered set $\bar{\mathbb{R}}$, thus endowed with structure, is called the **Extended-Real-Number System**; its members are called **extended-real numbers**; ∞ is called (**plus**) **infinity**, and is regarded as strictly positive; $-\infty$ is called **minus infinity**, and is regarded as strictly negative. An extended-real number is said to be **finite** or **infinite** according as it is in $\mathbb{R}$ or not.

We set $\bar{\mathbb{P}} := \mathbb{P}\cup\{\infty\} = \{t\in\bar{\mathbb{R}} \mid t\geq 0\}$, and extend the use of plain square brackets for order-intervals in $\mathbb{R}$ to those in $\bar{\mathbb{R}}$. In particular,

$$\begin{gathered}[a,\infty[= a+\mathbb{P} \quad]a,\infty[= a+\mathbb{P}^\times \quad]-\infty,a] = a-\mathbb{P}\\ ]-\infty,a[= a-\mathbb{P}^\times \quad]-\infty,\infty[= \mathbb{R},\end{gathered} \tag{152.1}$$

which are intervals in $\mathbb{R}$ (Proposition 151A). The notations recorded in (152.1) are often used in $\mathbb{R}$ even when "∞" and "$-\infty$" are not regarded as names for actual objects, but merely convenient notational devices.

152A. Proposition. $\bar{\mathbb{P}}+\bar{\mathbb{P}}\subset\bar{\mathbb{P}}$, $\bar{\mathbb{P}}+\bar{\mathbb{P}}^\times\subset\bar{\mathbb{P}}^\times$, $\bar{\mathbb{P}}\bar{\mathbb{P}}\subset\bar{\mathbb{P}}$, $\bar{\mathbb{P}}^\times\bar{\mathbb{P}}^\times\subset\bar{\mathbb{P}}^\times$. *The set $\bar{\mathbb{P}}$, with 0 as zero, 1 as unity, and addition and multiplication adjusted from the corresponding operations in $\bar{\mathbb{R}}$, is both a commutative monoid written additively and a commutative monoid written multiplicatively, and the distributive law holds:*

$$\forall r,s,t\in\bar{\mathbb{P}},\qquad (r+s)t = rt+st.$$

As an ordered subset of $\bar{\mathbb{R}}$, $\bar{\mathbb{P}}$ is completely and totally ordered, and the following monotonicity laws hold:

$$\forall r,s,t\in\bar{\mathbb{P}},\quad r\leq s \;\Rightarrow\; r+t\leq s+t \tag{152.2}$$

$$\forall r,s,t\in\bar{\mathbb{P}},\quad r\leq s \;\Rightarrow\; rt\leq st. \tag{152.3}$$

In the commutative monoid $\bar{\mathbb{P}}$ written additively it is possible to extend the notion of *sum* from families with finite support to *all* families.

Suppose first that the family $a\in\bar{\mathbb{P}}^I$ has finite support, and let $J\in\mathfrak{P}(I)$ be given. By (152.2) we have

$$\sum_{J\cap\mathrm{Supp}a} a = \sum_J a = \sum_K a + \sum_{J\setminus K} a \geq \sum_K a \qquad \text{for every } K\in\mathfrak{F}(J),$$

and therefore

$$\sum_J a = \max\{\sum_K a \mid K\in\mathfrak{F}(J)\}.$$

Now let a family $a \in \bar{\mathbb{P}}^I$ be given; we make no assumption about its support. The preceding discussion shows that there will be no notational clash if we define

$$\sum_J a := \sup\{\sum_K a \mid K \in \mathfrak{F}(J)\} \qquad \text{for all } J \in \mathfrak{P}(I), \tag{152.4}$$

as we may, since $\bar{\mathbb{P}}$ is completely ordered; and we may call $\sum_J a$ the **sum of** the family a **over** J.

This extended notion of sum shares many of the properties of the sums of families with finite support, as described in Sections 113, 114, 115, 117; we shall not develop these here, except for the few results that follow.

152B. Remark. Let the family $a \in \bar{\mathbb{P}}^I$ and $J \in \mathfrak{P}(I)$ be given, and assume that $\sum_J a$ is finite. It is obvious that a_j is finite for every $j \in J$, and that the set $K_n := \{j \in J \mid a_j > (1/n)\sum_J a\}$ is finite, with $\#K_n < n$, for all $n \in \mathbb{N}^\times$. Therefore $J \cap \text{Supp} a = \bigcup_{n\in\mathbb{R}^\times} K_n$ (cf. Proposition 142J) is countable, by •Corollary 121N. ■

152C. Lemma. *Let the sequence $a \in \bar{\mathbb{P}}^{\mathbb{N}^\times}$ and $m \in \mathbb{N}$ be given. Then*

$$\sum_J a = \sup\{\sum_{J\cap n^\sqsupset} a \mid n \in m + \mathbb{N}\} \qquad \textit{for all } J \in \mathfrak{P}(\mathbb{N}^\times).$$

Proof. For every $K \in \mathfrak{F}(J)$ we may choose an upper bound n of $K \cup m^\sqsupset$ and find that $n \in m + \mathbb{N}$ and, by (152.2),

$$\sum_J a \geq \sum_{J\cap n^\sqsupset} a = \sum_K a + \sum_{(J\setminus K)\cap n^\sqsupset} a \geq \sum_K a.$$

The assertion then follows from (152.4). ■

152D. Lemma. *Let the family $a \in \bar{\mathbb{P}}^I$ be given. The mapping $J \mapsto \sum_J a : \mathfrak{P}(I) \to \bar{\mathbb{P}}$ is $(\subset_I, \leq)$-isotone.*

152E. Proposition. *Let the family $a \in \bar{\mathbb{P}}^I$ and the disjoint collection $\mathcal{C}$ of subsets of I be given. Then*

$$\sum_{\bigcup\mathcal{C}} a = \sum_{A\in\mathcal{C}} \sum_A a.$$

Proof. 1. Let $K \in \mathfrak{F}(\bigcup\mathcal{C})$ be given. Set $\mathcal{D} := \{A \in \mathcal{C} \mid A \cap K \neq \emptyset\}$. Then $\{A \cap K \mid A \in \mathcal{D}\}$ is a (finite) partition of K. By Corollary 144C,

$$\sum_K a = \sum_{A\in\mathcal{D}} \sum_{A\cap K} a \leq \sum_{A\in\mathcal{D}} \sum_A a \leq \sum_{A\in\mathcal{C}} \sum_A a$$

(the first inequality is valid because $\mathcal{D}$ is finite). Since $K \in \mathfrak{F}(\bigcup\mathcal{C})$ was arbitrary, we have

$$\sum_{\bigcup\mathcal{C}} a \leq \sum_{A\in\mathcal{C}} \sum_A a. \tag{152.5}$$

2. Let $\mathcal{D} \in \mathfrak{F}(\mathcal{C})$ be given. If $\sum_{\bigcup \mathcal{D}} a = \infty$, it follows that

$$\sum_{\bigcup \mathcal{D}} a \geq \sum_{A \in \mathcal{D}} \sum_{A} a. \tag{152.6}$$

In order to prove (152.6) in every case, we may therefore assume that $\sum_{\bigcup \mathcal{D}} a \in \mathbb{P}$; it follows by Lemma 152D that $\sum_{A} a \in \mathbb{P}$ for all $A \in \mathcal{D}$.

Let $\epsilon \in \mathbb{P}^\times$ be given. We may choose a family $(B_A \mid A \in \mathcal{D}) \in \bigtimes_{A \in \mathcal{D}} \mathfrak{F}(A)$ such that

$$\forall A \in \mathcal{D}, \quad \sum_{A} a < \epsilon + \sum_{B_A} a.$$

Then $K := \bigcup_{A \in \mathcal{D}} B_A \in \mathfrak{F}(\bigcup \mathcal{D})$, and by Corollary 114C and Lemma 152D (the latter also to deal with possibly empty terms),

$$\sum_{A \in \mathcal{D}} \sum_{A} a \leq \sum_{A \in \mathcal{D}} (\epsilon + \sum_{B_A} a) = \epsilon(\#\mathcal{D}) + \sum_{K} a \leq \epsilon(\#\mathcal{D}) + \sum_{\bigcup \mathcal{D}} a.$$

Since $\epsilon \in \mathbb{P}^\times$ was arbitrary, we have shown that (152.6) also holds when $\sum_{\bigcup \mathcal{D}} a \in \mathbb{P}$.

3. Since $\mathcal{D} \in \mathfrak{F}(C)$ was arbitrary in Part 2, we have, using Lemma 152D,

$$\forall \mathcal{D} \in \mathfrak{F}(\mathcal{C}), \quad \sum_{A \in \mathcal{D}} \sum_{A} a \leq \sum_{\bigcup \mathcal{D}} a \leq \sum_{\bigcup \mathcal{C}} a.$$

By the definition of the sum of the family $(\sum_{A} a \mid A \in \mathcal{C})$, we conclude that

$$\sum_{A \in \mathcal{C}} \sum_{A} a \leq \sum_{\bigcup \mathcal{C}} a.$$

Together with (152.5) this yields the desired conclusion. ∎

153. Binary digital expansion

In this section we shall establish the familiar fact that for every real number $r \in]0,1]$ there is a sequence $s \in \{0,1\}^{\mathbb{N}^\times}$ such that $r = \sum_{n\in\mathbb{N}^\times} s_n 2^{-n}$. Such a sequence may be called a *binary digital expansion of* r; for each $n \in \mathbb{N}^\times$, s_n is the *nth binary digit* (or *bit*) in the expansion. It is clear that the binary digital expansion s is known once its support $A := \mathrm{Supp} s$ is known: indeed, $s = \chi_{A\subset\mathbb{N}^\times}$, and we also have $r = \sum_{n\in A} 2^{-n}$.

It is well known that for some r there are two binary expansions; this however, happens if and only if there is one with finite support. If we insist on a *non-terminating* expansion, i.e., one with infinite support, then there is exactly one for each $r \in]0,1]$; indeed, we shall see that the rule $A \mapsto \sum_{n\in A} 2^{-n}$ defines a bijection from the collection of all infinite subsets of $\mathbb{N}^\times$ to the interval $]0,1]$.

153A. Lemma. *Let the non-empty subset A of $\mathbb{N}^\times$ be given. Then*

$$\forall m,n\in\mathbb{N},\quad m\le n \;\Rightarrow \sum_{k\in(n^\sqsupset\setminus m^\sqsupset)\cap A} 2^{-k} \le 2^{-m}-2^{-n} \tag{153.1}$$

$$\sum_{k\in A} 2^{-k} \in]0,1]. \tag{153.2}$$

Proof. By (121.1) we have

$$\sum_{j\in l^\sqsubset} 2^j = 2^l - 1 \qquad \text{for all } l\in\mathbb{N}. \tag{153.3}$$

Let $m,n\in\mathbb{N}$ be given, with $m\le n$. Then Proposition 142C and (153.3) yield

$$\sum_{k\in(n^\sqsupset\setminus m^\sqsupset)\cap A} 2^{-k} \le \sum_{k\in(n^\sqsupset\setminus m^\sqsupset)\cap A} 2^{-k} + \sum_{k\in(n^\sqsupset\setminus m^\sqsupset)\setminus A} 2^{-k} = \sum_{k\in n^\sqsupset\setminus m^\sqsupset} 2^{-k} =$$

$$= \sum_{j\in(n-m)^\sqsubset} 2^{-(n-j)} = 2^{-n}\sum_{j\in(n-m)^\sqsubset} 2^j = 2^{-n}(2^{n-m}-1) = 2^{-m}-2^{-n}.$$

This establishes (153.1).

From (153.1) with $m := 0$ we have $\sum_{k\in A\cap n^\sqsupset} 2^{-k} \le 1-2^{-n} < 1$ for all $n\in\mathbb{N}$. On the other hand, since $A\ne\emptyset$, we have $\sum_{k\in A\cap(\min A)^\sqsupset} 2^{-k} = 2^{-\min A} > 0$. It follows from Lemma 152C that (153.2) holds. ■

153B. Lemma. *Let $r\in]0,1]$ and the subset A of $\mathbb{N}^\times$ be given. Then*

$$A = \{n\in\mathbb{N}^\times \mid \lceil(2^n r)\in 2\mathbb{N}\} \tag{153.4}$$

if and only if

$$\sum_{k\in A\cap n^{\sqsupset}} 2^{n-k} = \lceil(2^n r) - 1 \quad \textit{for all } n \in \mathbb{N}. \tag{153.5}$$

If A satisfies these conditions, then A is infinite.

Proof. 1. We assume that (153.4) holds, and prove (153.5) by induction. We have $\lceil(2^0 r) = \lceil r = 1$, so that $0 \notin A$, and $\sum_{k\in A\cap 0^{\sqsupset}} 2^{0-k} = 0 = 1 - 1 = \lceil(2^0 r) - 1$. Suppose now that $n \in \mathbb{N}$ is such that $\sum_{k\in A\cap n^{\sqsupset}} 2^{n-k} = \lceil(2^n r) - 1$. We distinguish two cases.

If $n + 1 \in A$, then $\lceil(2^{n+1} r) \in 2\mathbb{N}$, and by (151.7) we have $\lceil(2^{n+1} r) = 2\lceil(2^n r)$. Therefore

$$\sum_{k\in A\cap(n+1)^{\sqsupset}} 2^{n+1-k} = \sum_{k\in A\cap n^{\sqsupset}} 2^{n+1-k} + 1 = 2\sum_{k\in A\cap n^{\sqsupset}} 2^{n-k} + 1 = 2(\lceil(2^n r) - 1) + 1 =$$

$$= \lceil(2^{n+1} r) - 1.$$

If, on the other hand, $n + 1 \notin A$, then $\lceil(2^{n+1} r) \notin 2\mathbb{N}$, and by (151.7) we have $\lceil(2^{n+1} r) = 2\lceil(2^n r) - 1$. Therefore

$$\sum_{k\in A\cap(n+1)^{\sqsupset}} 2^{n+1-k} = \sum_{k\in A\cap n^{\sqsupset}} 2^{n+1-k} = 2\sum_{k\in A\cap n^{\sqsupset}} 2^{n-k} = 2(\lceil(2^n r) - 1) = \lceil(2^{n+1} r) - 1.$$

This completes the induction step.

2. We now assume that (153.5) holds. Let $n \in \mathbb{N}$ be given. If $n + 1 \in A$, then

$$\lceil(2^{n+1} r) = \sum_{k\in A\cap(n+1)^{\sqsupset}} 2^{n+1-k} + 1 = \sum_{k\in A\cap n^{\sqsupset}} 2^{n+1-k} + 2 = 2(\sum_{k\in A\cap n^{\sqsupset}} 2^{n-k} + 1) \in 2\mathbb{N};$$

if $n + 1 \notin A$, then

$$\lceil(2^{n+1} r) = \sum_{k\in A\cap(n+1)^{\sqsupset}} 2^{n+1-k} + 1 = \sum_{k\in A\cap n^{\sqsupset}} 2^{n+1-k} + 1 = 2\sum_{k\in A\cap n^{\sqsupset}} 2^{n-k} + 1 \notin 2\mathbb{N}.$$

Since $n \in \mathbb{N}$ was arbitrary, this establishes (153.4).

3. We assume that A satisfies (153.5). Suppose that $m \in \mathbb{N}$ were an upper bound of A. Then we should have

$$\sum_{k\in A} 2^{n-k} = \lceil(2^n r) - 1 < 2^n r \le \lceil(2^n r) = \sum_{k\in A} 2^{n-k} + 1 \qquad \text{for all } n \in m + \mathbb{N}. \tag{153.6}$$

In particular, we should have $\sum_{k\in A} 2^{-k} = 2^{-m}\sum_{k\in A} 2^{m-k} < 2^{-m}2^m r = r$ (Proposition 142C and (151.1)). The sequence $(2^n \mid n \in m + \mathbb{N})$ is strictly isotone; hence its range

is infinite and has no upper bound in $\mathbb{N}$. Since $\mathbb{R}$ is archimedean, we might therefore choose $n \in m + \mathbb{N}$ such that $2^n > 1/(r - \sum_{k\in A} 2^{-k})$. Then

$$\sum_{k\in A} 2^{n-k} + 1 = 2^n \sum_{k\in A} 2^{-k} + 1 < 2^n r,$$

contradicting (153.6). Therefore our supposition that A had an upper bound in $\mathbb{N}$ is untenable, and A is infinite. ∎

Lemmas 153A and 153B justify the definition of the mappings $\mathrm{Bin} :]0,1] \to \mathfrak{P}(\mathbb{N}^\times)\backslash\mathfrak{F}(\mathbb{N}^\times)$ and $\mathrm{bin} : \mathfrak{P}(\mathbb{N}^\times)\backslash\mathfrak{F}(\mathbb{N}^\times) \to]0,1]$ by the rules

$$\mathrm{Bin}(r) := \{n \in \mathbb{N}^\times \mid \lceil (2^n r) \in 2\mathbb{N}\} \qquad \text{for all } r \in]0,1] \tag{153.7}$$

$$\mathrm{bin}(A) := \sum_{k\in A} 2^{-k} \qquad \text{for all } A \in \mathfrak{P}(\mathbb{N}^\times)\backslash\mathfrak{F}(\mathbb{N}^\times). \tag{153.8}$$

153C. Theorem. *For every $r \in]0,1]$ there is exactly one infinite subset A of $\mathbb{N}^\times$ such that $r = \sum_{k\in A} 2^{-k}$, namely $A := \mathrm{Bin}(r)$. More precisely, each of the mappings* Bin *and* bin *is the inverse of the other.*

Proof. 1. Let $r \in]0,1]$ be given. By (151.6) and Lemma 153B with $A := \mathrm{Bin}(r)$ we have

$$2^n r - 1 \le \sum_{k\in \mathrm{Bin}(r)\cap n^\sqsupset} 2^{n-k} < 2^n r \qquad \text{for all } n \in \mathbb{N}.$$

By Proposition 142C and (151.1) this implies

$$r - 2^{-n} \le \sum_{k\in \mathrm{Bin}(r)\cap n^\sqsupset} 2^{-k} < r \qquad \text{for all } n \in \mathbb{N}.$$

For given $s \in]0, r[$ we may choose $n \in \mathbb{N}$ such that $2^n \ge 1/(r-s)$ (cf. proof of Lemma 153B), and hence $r - 2^{-n} \ge s$. It follows from Lemma 152C that

$$\mathrm{bin}(\mathrm{Bin}(r)) = \sum_{k\in \mathrm{Bin}(r)} 2^{-k} = \sup\{\sum_{k\in \mathrm{Bin}(r)\cap n^\sqsupset} 2^{-k} \mid n \in \mathbb{N}\} = r.$$

Since $r \in]0,1]$ was arbitrary, we have $\mathrm{bin} \circ \mathrm{Bin} = 1_{]0,1]}$.

2. Let $A \in \mathfrak{P}(\mathbb{N}^\times)\backslash\mathfrak{F}(\mathbb{N}^\times)$ and $m \in \mathbb{N}$ be given. Since A is infinite, we may set $m' := \min(A\backslash m^\sqsupset)$. Using Lemma 153A (Formula (153.1)), we find

$$\sum_{k\in A\cap m^\sqsupset} 2^{-k} < \sum_{k\in A\cap m^\sqsupset} 2^{-k} + 2^{-m'} = \sum_{k\in A\cap m'^\sqsupset} 2^{-k} \le \sum_{k\in A\cap m'^\sqsupset} 2^{-k} + \sum_{k\in A\cap (n^\sqsupset\backslash m'^\sqsupset)} 2^{-k} =$$

$$= \sum_{k\in A\cap n^\sqsupset} 2^{-k} = \sum_{k\in A\cap m^\sqsupset} 2^{-k} + \sum_{k\in A\cap (n^\sqsupset\backslash m^\sqsupset)} 2^{-k} < \sum_{k\in A\cap m^\sqsupset} 2^{-k} + 2^{-m}$$

$$\text{for all } n \in m' + \mathbb{N}.$$

By Proposition 142C and (151.1) we may "multiply through" the preceding chain of equalities and inequalities by 2^m. It then follows from Lemma 152C that

$$\sum_{k\in A\cap m^\sqsupset} 2^{m-k} = 2^m \sum_{k\in A\cap m^\sqsupset} 2^{-k} < 2^m \sum_{k\in A} 2^{-k} \leq 2^m \sum_{k\in A\cap m^\sqsupset} 2^{-k} + 1 = \sum_{k\in A\cap m^\sqsupset} 2^{m-k} + 1.$$

Using (153.8) and (151.5), we may rewrite this as

$$\lceil (2^m \mathrm{bin}(A)) = \sum_{k\in A\cap m^\sqsupset} 2^{m-k} + 1. \tag{153.9}$$

Now (153.9) holds for all $m \in \mathbb{N}$. Applying Lemma 153B with $r := \mathrm{bin}(A)$ and comparing (153.9) with (153.5) and (153.7) with (153.4), we conclude that $A = \mathrm{Bin}(\mathrm{bin}(A))$. Since $A \in \mathfrak{P}(\mathbb{N}^\times)\backslash\mathfrak{F}(\mathbb{N}^\times)$ was arbitrary, we have $\mathrm{Bin}\circ\mathrm{bin} = 1_{\mathfrak{P}(\mathbb{N}^\times)\backslash\mathfrak{F}(\mathbb{N}^\times)}$. ∎

153D. Remarks. (a): Let $r \in]0,1]$ be given. By Theorem 153C we have

$$r = \sum_{k\in \mathrm{Bin}(r)} 2^{-k} = \sum_{k\in\mathbb{N}^\times} \chi_{\mathrm{Bin}(r)}(k) 2^{-k}.$$

As mentioned at the beginning of this section (and noting that $\mathrm{Bin}(r)$ is an infinite set), the sequence $\chi_{\mathrm{Bin}(r)} \in \{0,1\}^{\mathbb{N}^\times}$ is called the **non-terminating binary digital expansion of** r.

(b): In order to extend the construction of non-terminating binary digital expansions to all strictly positive real numbers, we may proceed by multiplying a given $r \in \mathbb{P}^\times$ by 2^{-m} for a suitable $m \in \mathbb{N}$ so that $2^{-m}r \in]0,1]$, and then applying Theorem 153C to $2^{-m}r$ instead of r. To express the result (we omit the details of the proof), we define Ξ to be the collection of all infinite subsets of $\mathbb{Z}$ that have lower bounds (and hence a minimum). Then the formula

$$\mathrm{Bin}_\infty(r) := \{n \in \mathbb{Z} \mid \lceil (2^n r) \in 2\mathbb{N}\} \qquad \text{for all } r \in \mathbb{P}^\times$$

defines a bijection $\mathrm{Bin}_\infty : \mathbb{P}^\times \to \Xi$ such that

$$r = \sum_{k\in \mathrm{Bin}_\infty(r)} 2^{-k} = \sum_{k\in\mathbb{Z}} \chi_{\mathrm{Bin}_\infty(r)}(k) 2^{-k} \qquad \text{for all } r \in \mathbb{P}^\times.$$

(c): We shall not be concerned here with *terminating* binary digital expansions, except to note that if A is a non-empty finite subset of $\mathbb{N}^\times$ then $\sum_{k\in A} 2^{-k} \in]0,1[$ and $\mathrm{Bin}(\sum_{k\in A} 2^{-k}) = A \mathbin{\triangle} (\max A + \mathbb{N})$, where $\triangle$ denotes symmetric difference of sets. ∎

It is possible to describe the order $\leq$ in $]0,1]$ by means of the non-terminating binary digital expansions, as we now show, by examining the least index at which they differ.

153E. Proposition.

$$\forall r, s \in]0,1], \quad r < s \;\Leftrightarrow\; (\mathrm{Bin}(r) \neq \mathrm{Bin}(s) \quad \textit{and} \quad \min(\mathrm{Bin}(r) \triangle \mathrm{Bin}(s)) \in \mathrm{Bin}(s)).$$

Proof. Since Bin is bijective, we may stipulate that $r \neq s$ and $\mathrm{Bin}(r) \neq \mathrm{Bin}(s)$. Then $\mathrm{Bin}(r) \triangle \mathrm{Bin}(s) \neq \emptyset$.

Assume first that $m := \min(\mathrm{Bin}(r) \triangle \mathrm{Bin}(s)) \in \mathrm{Bin}(s)$. Then $\mathrm{Bin}(r) \cap (m-1)^{\sqsupset} = \mathrm{Bin}(s) \cap (m-1)^{\sqsupset}$ and $m \notin \mathrm{Bin}(r)$. We therefore have, using Lemma 153A,

$$\sum_{k \in \mathrm{Bin}(r) \cap n^{\sqsupset}} 2^{-k} = \sum_{k \in \mathrm{Bin}(r) \cap (m-1)^{\sqsupset}} 2^{-k} + \sum_{k \in \mathrm{Bin}(r) \cap (n^{\sqsupset} \setminus m^{\sqsupset})} 2^{-k} <$$

$$< \sum_{k \in \mathrm{Bin}(s) \cap (m-1)^{\sqsupset}} 2^{-k} + 2^{-m} = \sum_{k \in \mathrm{Bin}(s) \cap m^{\sqsupset}} 2^{-k} \leq \sum_{k \in \mathrm{Bin}(s)} 2^{-k}$$

for all $n \in m + \mathbb{N}$.

By Lemma 152C and Theorem 153C, $r = \sum_{k \in \mathrm{Bin}(r)} 2^{-k} \leq \sum_{k \in \mathrm{Bin}(s)} 2^{-k} = s$. Since $r \neq s$, we have $r < s$.

If, on the other hand, $\min(\mathrm{Bin}(r) \triangle \mathrm{Bin}(s)) \in \mathrm{Bin}(r)$, the preceding argument with r and s interchanged, shows that $r > s$. ■

153F. Remark. By Theorem 153C and Proposition 122C, $]0,1]$ is equinumerous to $\mathfrak{P}(\mathbb{N}^{\times}) \setminus \mathfrak{F}(\mathbb{N}^{\times})$, hence to $\mathfrak{P}(\mathbb{N}^{\times})$, hence also to $\mathfrak{P}(\mathbb{N})$. By Theorem 122A $]0,1]$ is uncountable (the idea for this proof of the uncountability of $]0,1]$ goes back to Georg Ferdinand Ludwig Philipp Cantor (1845-1918)). It is easy to deduce that every genuine interval is uncountable. Although this was already established in Proposition 151B (also due to Cantor, as noted in Example 122H), the present observation provides a more precise determination of the "degree of uncountability" of genuine intervals: every one is equinumerous with the uncountable set $\mathfrak{P}(\mathbb{N})$. ■

Chapter 16

THE REAL NUMBERS: EXISTENCE

161. Construction of a complete ordered field

As was explained in Sections 141 and 151, we are concerned with the question regarding the existence of a complete ordered field only to the extent of ascertaining that this question is equivalent to the one regarding the existence of a counting system. Cultural Anthropology and the history of Mathematics will surely bear out the contention that it is more natural to accept consciously the existence of the Natural-Number System — no matter how much sophistication this may have required — than to extend the same degree of acceptance to the existence of the entire Real-Number System. It has therefore long been regarded as a matter of interest to show, by means of a construction, that the acceptance of the former compels the acceptance of the latter.

Instead of carrying out this program by first constructing, from a given counting system, structures imitating (technically: isomorphic to) $\mathbb{Z}$ and $\mathbb{Q}$, we prefer to use as our intermediate step a structure imitating $\overline{\mathbb{P}}$ (cf. Section 152), and to obtain this structure from a counting system by a version of the Eudoxian theory of ratios.

We define a **positivity system** to be a completely and totally ordered set P (with strict-order and lax order here denoted by $<$ and $\leq$, respectively, and with $\infty := \max P$), endowed, in addition, with the structures of a commutative monoid, written additively, and of a commutative monoid, written multiplicatively (with unity e), subject to the following conditions:

(PS1): $\forall x, y, z \in P, \quad (x + y)z = (xz) + (yz)$
(PS2): $\forall x, y \in P, \quad x + y = x \;\Rightarrow\; (x = \infty \text{ or } y = 0)$
(PS3): $\forall x, y \in P, \quad x + y = \infty \;\Rightarrow\; (x = \infty \text{ or } y = \infty)$
(PS4): $\forall x, y \in P, \quad x \leq y \;\Leftrightarrow\; (\exists z \in P, \quad x + z = y)$
(PS5): $e \notin \{0, \infty\}$
(PS6): $\forall x \in P, \quad x \notin \{0, \infty\} \;\Rightarrow\; (\exists y \in P, \quad xy = e\}$.

In dealing with a positivity system, we shall use the distributive law (PS1) without explicit mention. We shall also accept the notational conventions, especially those concerning parentheses, that are in common use for addition and multiplication. We set $P_0 := P\backslash\{\infty\}$ and $P_0^\times := {P_0}^\times = P\backslash\{0,\infty\}$.

▼ **161A. Remarks**. (a): In verifying that a given structure is a positivity system P, it is not necessary to verify that $x+0=x$ for all $x\in P$. Indeed, let $x\in P_0$ be given. Since $x\le x$, we may, by (PS4), choose $y\in P$ such that $x+y=x$. From (PS2) it follows that $y=0$. On the other hand, by (PS4), $\infty\le\infty+0\le\max P=\infty$, so that equality holds.

(b): From (PS4) we find $\min P\le 0\le 0+\min P=\min P$; it follows that $\min P=0$, and from this and (PS5) we have $0<e<\infty$. ■

161B. Lemma. *Let the positivity system P be given. Then*

$$P_0+P_0\subset P_0 \tag{161.1}$$

$$P_0P_0\subset P_0 \tag{161.2}$$

$$0P_0=\{0\} \tag{161.3}$$

$$P_0^\times P_0^\times\subset P_0^\times \tag{161.4}$$

$$\forall x\in P_0^\times,\qquad \exists y\in P_0^\times,\qquad xy=e. \tag{161.5}$$

Proof. (161.1) is an immediate consequence of (PS3).

For every $x\in P$ we have, by (PS4), $\infty\le\infty+x=x+\infty\le\infty$; hence

$$\forall x\in P,\qquad x+\infty=\infty. \tag{161.6}$$

From (PS4) and (161.6) we have

$$\infty=e\infty\le e\infty+\infty\infty=(e+\infty)\infty=\infty\infty\le\infty. \tag{161.7}$$

Let $x\in P_0^\times$ be given. By (PS6) we may choose $y\in P$ such that $xy=e$. Then (PS4) and (161.6) yield

$$\infty=e\infty=xy\infty\le xy\infty+x\infty=x(y\infty+\infty)=x\infty\le\infty.$$

Combining this with (161.7), we find

$$\forall x\in P^\times,\qquad x\infty=\infty. \tag{161.8}$$

Let $x\in P_0$ be given. Then $x+x0=xe+x0=x(e+0)=xe=x$, and therefore, by (PS2), $x0=0$. This establishes (161.3). We have $0\infty=(0+0)\infty=0\infty+0\infty$. It follows from (PS2) that

$$0\infty\in\{0,\infty\}. \tag{161.9}$$

Let $x, y \in P_0^\times$ be given. By (PS6), we may choose $u, v \in P$ such that $xu = e = yv$. Then (PS5) yields

$$(xy)(uv) = (xu)(yv) = ee = e \in P_0^\times.$$

From this and (161.3), (161.8), (161.9) it follows that $xy \in P_0^\times$. This establishes (161.4), and (161.2) follows from this and (161.3).

If $x \in P_0^\times, y \in P$, and $xy = e$, then (161.3) and (161.8) show that $y \in P_0^\times$. This, together with (PS6), establishes (161.5). ■

Let the positivity system P be given. For given $x, y \in P$ we consider the problem

$$?z \in P, \qquad x + z = y. \tag{161.10}$$

By (PS4), this problem has a solution if and only if $x \leq y$. Now suppose that $x \leq y < \infty$. If z', z'' are solutions of (161.10) with $z' \leq z''$, we may, by (PS4), choose $u \in P$ such that $z' + u = z''$. Then $y = x + z'' = x + z' + u = y + u$. By (PS2) we find that $u = 0$, and hence $z'' = z' + 0 = z'$. Since P is totally ordered, this discussion shows that, if $x \leq y < \infty$, then (161.10) has exactly one solution; by (PS4), this solution is in P_0. In that case we define

$$y - x :\in \{z \in P \mid x + z = y\} \subset P_0.$$

We note that

$$\forall x, y \in P_0, \qquad x \leq y \;\Rightarrow\; (y - x = 0 \;\Leftrightarrow\; x = y).$$

We set

$$F := \{(x, y) \in P_0 \times P_0 \mid xy = 0\}, \tag{161.11}$$

and define the mapping $p : P_0 \times P_0 \to P_0 \times P_0$ by the rule

$$p((x, y)) := (x - \min\{x, y\},\ y - \min\{x, y\}) \qquad \text{for all } (x, y) \in P_0 \times P_0. \tag{161.12}$$

In the following two lemmas P is a given positivity system, and F and p are defined by (161.11) and (161.12).

161C. Lemma. (a): *p is idempotent and* $\mathrm{Rng}p = F$.

(b): $p((x + z, y + z)) = p((x, y))$ *for all* $x, y, z \in P_0$.

Proof. Proof of (a). Let $x, y \in P_0$ be given. Then $x - \min\{x, y\} = 0$ or $y - \min\{x, y\} = 0$ according as $x \leq y$ or $x \geq y$; it follows from (161.3) that $p((x, y)) \in F$. We conclude that $\mathrm{Rng}p \subset F$.

On the other hand, let $(x, y) \in F$ be given. By (161.4) either $x = 0$ or $y = 0$, and hence $\min\{x, y\} = 0$, since $0 = \min P$ (Remark 161A,(b)). It follows that $p((x, y)) = (x, y)$. Thus $F \subset \mathrm{Rng}p$, and every member of F is a fixed point of p. It follows that $\mathrm{Rng}p = F$ and, by Proposition 26C, that p is idempotent.

Proof of (b). We first assume that $x \leq y$. Then $x + z + (y - x) = x + (y - x) + z = y + z$. By (PS4) we have $x + z \leq y + z$ and $(y + z) - (x + z) = y - x$. Hence

$p((x+z,y+z)) = (0,y-x) = p((x,y))$. Similarly, if $x \geq y$, then $p((x+z,y+z)) = (x-y,0) = p((x,y))$. ■

161D. Lemma. *For all $(x,y),(u,v) \in F$ we have $(y,x) \in F$ and $(xu+yv, xv+yu) \in F$.*

Proof. The first assertion is obvious. The second follows from (161.3) by observing that

$$(xu+yv)(xv+yu) = (xx+yy)(uv)+(uu+vv)(xy) = (xx+yy)0+(uu+vv)0 = 0+0 = 0. \blacksquare$$

161E. Theorem. *Let the positivity system P be given, and let the set F and the mapping p be defined by* (161.11) *and* (161.12). *Then F, with $(0,0)$ as zero; $(e,0)$ as unity; addition, opposition, and multiplication defined respectively by the rules*

$$(x,y)+(u,v) := p((x+u,y+v)) \qquad \text{for all } (x,y),(u,v) \in F$$

$$-(x,y) := (y,x) \qquad \text{for all } (x,y) \in F$$

$$(x,y)(u,v) := (xu+yv, xv+yu) \qquad \text{for all } (x,y),(u,v) \in F$$

(as is permitted by Lemmas 161B, 161C, and 161D); and the subset $P_0 \times \{0\}$ of F as positive half, is a complete ordered field.

Proof. The commutative laws for addition and multiplication hold as trivial consequences of the corresponding laws in P. The associative law for multiplication holds as a trivial consequence of the associative laws and the distributive law in P.

Let $(x,y),(u,v),(w,z) \in F$ be given. Set $m := \min\{x+u, y+v\}$, $n := \min\{u+w, v+z\}$. By Lemma 161C and (161.3),

$$\begin{aligned}((x,y)+(u,v))+(w,z) &= p((x+u-m+w, y+v-m+z)) = \\ &= p((x+u+w, y+v+z)) = p((x+u+w-n, y+v+z-n)) = \\ &= (x,y)+((u,v)+(w,z)),\end{aligned}$$

$$\begin{aligned}((x,y)+(u,v))(w,z) &= (x+u-m, y+v-m)(w,z) = \\ &= ((x+u-m)w+(y+v-m)z, (x+u-m)z+(y+v-m)w) = \\ &= p(((x+u-m)w+(y+v-m)z, (x+u-m)z+(y+v-m)w)) = \\ &= p((x+u)w+(y+v)z, (x+u)z+(y+v)w)) = \\ &= p((xw+yz+uw+vz, xz+yw+uz+vw)) = \\ &= (xw+yz, xz+yw)+(uw+vz, uz+vw) = \\ &= ((x,y)(w,z))+((u,v)(w,z)),\end{aligned}$$

$$(x,y)+(0,0) = p((x+0,y+0)) = p((x,y)) = (x,y),$$

$$((x,y)+(-(x,y)) = (x,y)+(y,x) = p((x+y,y+x)) = p((0,0)) = (0,0),$$

$$(e,0)(x,y) = (ex+0y, ey+0x) = (x,y).$$

This establishes the associative law for addition, the neutrality laws for addition and multiplication, the law of opposites, and the distributive law. We also have $0 \neq e$, and therefore $(0,0) \neq (e,0)$.

Let $(x,y) \in F\backslash\{(0,0)\}$ be given. By (161.3), (161.4), either $x \neq 0 = y$ or $y \neq 0 = x$. By (161.5) we may choose $z \in P_0^\times$ such that $xz = e$ or $yz = e$, respectively, and we respectively find

$$(x,y)(z,0) = (x,0)(z,0) = (xz + 00, x0 + 0z) = (e,0)$$

$$(x,y)(0,z) = (0,y)(0,z) = (00 + yz, 0z + y0) = (e,0).$$

Let $(x,y),(u,v) \in P_0 \times \{0\}$ be given. Then $y = 0 = v$. Since $0 = \min P$ (by Remark 9A,(b)),

$$(x,y) + (u,v) = (x,0) + (u,0) = p((x+u, 0+0)) = (x+u,0) \in P_0 \times \{0\},$$

$$(x,y)(u,v) = (x,0)(u,0) = (xu + 00, x0 + 0u) = (xu,0) \in P_0 \times \{0\}.$$

Moreover,

$$-(P_0 \times \{0\}) = \{-(x,y) \mid (x,y) \in P_0 \times \{0\}\} = \{(y,x) \mid (x,y) \in P_0 \times \{0\}\} = \{0\} \times P_0.$$

Therefore

$$(P_0 \times \{0\}) \cap (-(P_0 \times \{0\})) = (P_0 \times \{0\}) \cap (\{0\} \times P_0) = \{0\} \times \{0\} = \{(0,0)\}.$$

By (161.3) and (161.4)

$$\begin{aligned} F = \{(x,y) \in P_0 \times P_0 \mid x = 0 \text{ or } y = 0\} &= (P_0 \times \{0\}) \cup (\{0\} \times P_0) = \\ &= (P_0 \times \{0\}) \cup (-(P_0 \times \{0\})). \end{aligned}$$

We have proved that F with the indicated structure is an ordered field; it remains to show that this ordered field is complete. By Lemma 143A it is sufficient to show that the ordered subset $P_0 \times \{0\}$ of F is pre-completely ordered.

Let $(x,y),(u,v) \in P_0 \times \{0\}$ be given. Then $y = 0 = v$, and

$$(u,v) - (x,y) = (u,0) - (x,0) = (u,0) + (0,x) =$$

$$= p((u,x)) = \begin{cases} (u-x,0) \in P_0 \times \{0\} & \text{if } x \leq u \\ (0,x-u) \in \{0\} \times P_0 = -(P_0 \times \{0\}) & \text{if } x \geq u. \end{cases}$$

This shows that the mapping $(x \mapsto (x,0)) : P_0 \to P_0 \times \{0\}$ is an order-isomorphism. But $P_0 = [\![\min P, \max P[\![$ is pre-completely ordered, by Proposition 72E,(b). It follows ▲ that $P_0 \times \{0\}$ is pre-completely ordered. ■

162. Construction of a positivity system

▼ We now propose to show how to construct a positivity system from a given counting system.

Let the counting system N be given. Except for using the symbol N, rather than $\mathbb{N}$, we shall feel free to use in N the terminology, notation, and results pertaining to the Natural-Number System, as described in Chapter 9. We may do this on account of Theorem 95A.

We set $\Lambda := N \times N^\times$, and define in Λ the relations $\precsim$, $\prec$, $\approx$ by the rules

$$\forall (m,m'),(n,n') \in \Lambda, \quad (m,m') \precsim (n,n') \quad :\Leftrightarrow \quad mn' \leq m'n$$

$$\forall (m,m'),(n,n') \in \Lambda, \quad (m,m') \prec (n,n') \quad :\Leftrightarrow \quad mn' < m'n$$

$$\forall (m,m'),(n,n') \in \Lambda, \quad (m,m') \approx (n,n') \quad :\Leftrightarrow \quad mn' = m'n.$$

These notations are compatible, since

$$\forall \mu,\nu \in \Lambda, \quad \mu \precsim \nu \quad \Leftrightarrow \quad (\mu \prec \nu \text{ or } \mu \approx \nu)$$

$$\forall \mu,\nu \in \Lambda, \quad \mu \prec \nu \quad \Leftrightarrow \quad \text{not}(\nu \precsim \mu) \quad \Leftrightarrow \quad (\mu \precsim \nu \text{ and not } \mu \approx \nu)$$

$$\forall \mu,\nu \in \Lambda, \quad \mu \approx \nu \quad \Leftrightarrow \quad (\mu \precsim \nu \text{ and } \nu \precsim \mu).$$

The relation $\precsim$ is reflexive, transitive, and total, but not antisymmetric. It also satisfies

$$\forall n \in N, \ \forall m',n' \in N^\times, \quad (0,m') \precsim (n,n')$$

$$\forall n \in N, \forall m',n' \in N^\times, \quad (n,n') \precsim (0,m') \ \Leftrightarrow \ n = 0.$$

We define the mappings $((\mu,\nu) \mapsto \mu+\nu) : \Lambda \times \Lambda \to \Lambda, ((\mu,\nu) \mapsto \nu - \mu) : \mathrm{Gr}(\precsim) \to \Lambda$, and $((\mu,\nu) \mapsto \mu\nu) : \Lambda \times \Lambda \to \Lambda$ by the rules

$$(m,m') + (n,n') := (mn' + m'n, m'n') \quad \text{for all } (m,m'),(n,n') \in \Lambda$$

$$(n,n') - (m,m') := (nm' - n'm, m'n') \quad \text{for all } ((m,m'),(n,n')) \in \mathrm{Gr}(\precsim)$$

$$(m,m')(n,n') := (mn, m'n') \quad \text{for all } (m,m'),(n,n') \in \Lambda.$$

We shall not record explicitly all the properties of these mappings and their relationships with the previously defined relations: all those we shall need can be immediately verified. We note, in particular, that "addition" and "multiplication" satisfy associative and commutative laws, and that

(162.1)

$$\forall \mu,\nu,\xi \in \Lambda, \quad \mu \precsim \nu \ \Rightarrow \ (\mu+\xi \precsim \nu+\xi \quad \text{and} \quad (\nu+\xi) - (\mu+\xi) \approx \nu - \mu)$$

$$\forall \mu,\nu,\xi \in \Lambda, \quad (\mu+\nu)\xi \approx (\mu\xi) + (\nu\xi).$$

We shall use and omit parentheses as is usual when "addition" and "multiplication" satisfy associative laws, and "multiplication" is agreed to have priority over "addition" and "subtraction". For all subsets $A,\ B$ of Λ we shall use the abbreviations $A+B := \{\alpha+\beta \mid (\alpha,\beta)\in A\times B\}$, $B-A := \{\beta-\alpha \mid (\alpha,\beta)\in (A\times B)\cap \mathrm{Gr}(\precapprox)\}$, $AB := \{\alpha\beta \mid (\alpha,\beta)\in A\times B\}$. We define the mappings $U, L \in \mathrm{Map}(\mathfrak{P}(\Lambda),\mathfrak{P}(\Lambda))$ by the rules

$$U(A) := \{\mu\in\Lambda \mid \forall\alpha\in A, \alpha \precapprox \mu\} \qquad \text{for all } A\in\mathfrak{P}(\Lambda)$$

$$L(A) := \{\mu\in\Lambda \mid \forall\alpha\in A,\ \mu \precapprox \alpha\}$$

(the sets of all "upper bounds" and of all "lower bounds" of A with respect to the reflexive, transitive, and total, but not antisymmetric, relation $\precapprox$ in Λ). Then the pair (U,L) is the Galois correspondence from $\mathfrak{P}(\Lambda)$ to $\mathfrak{P}(\Lambda)$ (both times ordered by inclusion) determined by the relation $\precapprox$ in Λ according to Proposition 74C. We use this fact and its consequences in what follows. In particular, both U and L are antitone, and $L\circ U\circ L = L,\ U\circ L\circ U = U$; the composite $L\circ U$ is a closure mapping in $\mathfrak{P}(\Lambda)$ ordered by inclusion, and $\mathrm{Rng}L = \mathrm{Rng}(L\circ U)$ is the set of fixed points of $L\circ U$ and is intersection-stable (Theorem 74E,(a),(b), and Theorem 73,(a)).

162A. Theorem. *Let the counting system N be given. The collection* $\mathrm{Rng}L$*, ordered by inclusion, with $L(\Lambda)$ as zero, $L(\{1,1)\})$ as unity, and addition and multiplication defined by the rules*

$$X\oplus Y := L(U(X+Y)) \quad \textit{for all } X,Y\in\mathrm{Rng}L$$

$$X\odot Y := L(U(XY)) \qquad \textit{for all } X,Y\in\mathrm{Rng}L,$$

is a positivity system.

Proof. 1. $\mathrm{Rng}L = \mathrm{Rng}(L\circ U)$ is intersection-stable, hence completely ordered by inclusion (Proposition 71F). We next claim that $\mathrm{Rng}L$ is a nest, i.e., that it is totally ordered by inclusion. Let $X,Y\in\mathrm{Rng}L$ be given, and suppose that $Y\backslash X\neq\emptyset$. Choose $\eta\in Y\backslash X$. Since $\eta\notin X = L(U(X))$ and $\precapprox$ is total, we may choose $\mu\in U(X)$ such that $\mu\prec\eta$. For every $\nu\in U(Y)$ we then have $\mu\precapprox\eta\precapprox\nu$. Since $\precapprox$ is transitive, we have $\nu\in U(X)$. We conclude that $U(Y)\subset U(X)$, and therefore $X = L(U(X))\subset L(U(Y)) = Y$. This establishes our claim.

We note that $\max\mathrm{Rng}L = \Lambda$, $\min\mathrm{Rng}L = \bigcap\mathrm{Rng}L = L(\Lambda) = \{0\}\times N^\times$, $L(\{(1,1)\}) = \{(m,m')\in\Lambda \mid m\le m'\}$.

To continue the proof we require two lemmas.

162B. Lemma. *Let $A,B\in\mathfrak{P}(\Lambda)\backslash\{\emptyset\}$ be given. Then*

$$L(U(A+B)) = L(U(A))\oplus L(U(B))$$

$$L(U(AB)) = L(U(A))\odot L(U(B)).$$

Proof. $A+B\subset L(U(A))+L(U(B))$; hence

$$L(U(A+B))\subset L(U(L(U(A))+L(U(B)))) = L(U(A))\oplus L(U(B)). \tag{162.2}$$

In exactly the same way we see that

$$L(U(AB)) \subset L(U(A)) \odot L(U(B)). \tag{162.3}$$

Let $\mu \in \Lambda\backslash U(L(U(A)) + L(U(B)))$ be given. We may then choose $\xi \in L(U(A))$, $\eta \in L(U(B))$ such that $\mu \prec \xi + \eta$. We now distinguish three cases. In the first case, $\mu \in \Lambda\backslash U(A) \subset \Lambda\backslash U(A+B)$; in the second, $\mu \in \Lambda\backslash U(B) \subset \Lambda\backslash U(A+B)$ (here we need the assumption that neither A nor B is empty). We are left with the third case, in which $\mu \in U(A) \cap U(B)$. Then $\xi \precapprox \mu$, and $\mu - \xi \prec \eta$; this implies $\mu - \xi \in \Lambda\backslash U(L(U(B))) = \Lambda\backslash U(B)$, and we may choose $\beta \in B$ such that $\mu - \xi \prec \beta$. We then have $\mu \prec \xi + \beta$; but $\beta \succapprox \mu$, and therefore $\mu - \beta \prec \xi$. Then $\mu - \beta \in \Lambda\backslash U(L(U(A))) = \Lambda\backslash U(A)$, and we may choose $\alpha \in A$ such that $\mu - \beta \prec \alpha$. Then $\mu \prec \alpha + \beta \in A + B$, and so $\mu \in \Lambda\backslash U(A+B)$ in the third case too. We have shown that $\Lambda\backslash U(L(U(A)) + L(U(B))) \subset \Lambda\backslash U(A+B)$. It follows that $U(A+B) \subset U(L(U(A)) + L(U(B)))$, and therefore

$$L(U(A+B)) \supset L(U(L(U(A)) + L(U(B)))) = L(U(A)) \oplus L(U(B)). \tag{162.4}$$

Let $\mu \in \Lambda\backslash U(L(U(A))L(U(B)))$ be given. We may then choose $(x, x') \in L(U(A))$, $\eta \in L(U(B))$ such that $\mu \prec (x, x')\eta$. We cannot have $x = 0$, and we deduce that $\mu(x', x) \prec \eta$. Therefore $\mu(x', x) \in \Lambda\backslash U(L(U(B))) = \Lambda\backslash U(B)$, and we may therefore choose $(b, b') \in B$ such that $\mu(x', x) \prec (b, b')$. We cannot have $b = 0$, and we infer that $\mu(b', b) \prec (x, x')$, so that $\mu(b', b) \in \Lambda\backslash U(L(U(A))) = \Lambda\backslash U(A)$. We may therefore choose $\alpha \in A$ such that $\mu(b', b) \prec \alpha$. Then $\mu \prec \alpha(b, b') \in AB$, and therefore $\mu \in \Lambda\backslash U(AB)$. We have shown that $\Lambda\backslash U(L(U(A))L(U(B))) \subset \Lambda\backslash U(AB)$. It follows that $U(AB) \subset U(L(U(A))L(U(B)))$, and therefore

$$L(U(AB)) \supset L(U(L(U(A))L(U(B)))) = L(U(A)) \odot L(U(B)). \tag{162.5}$$

The assertion follows by combining (162.2) with (162.4) and (162.3) with (162.5).■

162C. Lemma. *Let $X \in \mathrm{Rng}L\backslash\{\Lambda\}$ and $\delta \in \Lambda\backslash L(\Lambda)$ be given. Then there exists $k \in N^\times$ such that $(k-1, 1)\delta \in X$, $(k, 1)\delta \notin X$.*

Proof. Choose $(m, m') \in \Lambda\backslash X = U(X)\backslash X$, and set $(d, d') := \delta$. Then $d \neq 0$, and

$$(m, m') \precapprox (md'd, d') = (md', 1)\delta,$$

so that $(md', 1)\delta \in U(X)\backslash X = \Lambda\backslash X$. We may therefore define $k := \min\{n \in N \mid (n, 1)\delta \in \Lambda\backslash X\}$. Since $(0, 1)\delta = (0, d') \in \bigcap \mathrm{Rng}L \subset X$, we find $k \in N^\times$. This k verifies the assertion. ■

Proof of Theorem 162A, continued. 2. Both addition and multiplication as defined in $\mathrm{Rng}L$ obviously satisfy the commutative law. The fact that both operations satisfy the associative law follows from Lemma 162B: for all $X, Y, Z \in \mathrm{Rng}L$,

$$\begin{aligned}(X \oplus Y) \oplus Z &= L(U(X+Y)) \oplus L(U(Z)) = L(U(X+Y+Z)) = \\ &= L(U(X)) \oplus L(U(Y+Z)) = X \oplus (Y \oplus Z)\end{aligned}$$

$$(X \odot Y) \odot Z = L(U(XY)) \odot L(U(Z)) = L(U(XYZ)) =$$
$$= L(U(X)) \odot L(U(YZ)) = X \odot (Y \odot Z).$$

From Lemma 162B and (162.1) (for the middle equality in the following chain) we obtain the distributive law (PS1):

$$\begin{aligned}(X \oplus Y) \odot Z &= L(U(X+Y)) \odot L(U(Z)) = L(U(X+Y)Z)) = \\ &= L(U((XZ)+(YZ))) = \\ &= L(U(XZ) \oplus L(U(YZ)) = (X \odot Z) \oplus (Y \odot Z).\end{aligned}$$

Let $X \in \mathrm{Rng}L$ be given. Then $X = \{(1,1)\}X \subset (L(\{1,1)\}))X$. On the other hand, if $(m,m') \in L(\{1,1)\}), (x,x') \in X$, then $m \leq m'$, and hence $mxx' \leq m'xx'$, so that $(m,m')(x,x') = (mx, m'x') \precapprox (x,x')$. Therefore $(m,m')(x,x') \in L(U(X)) = X$. We have shown that $L(\{(1,1)\})X \subset X$. We conclude that $L(\{(1,1)\})X = X$. This establishes the neutrality law for multiplication. By Remark 161A,(a), it is not necessary to prove the neutrality law for addition explicitly.

3. Let $X, Y \in \mathrm{Rng}L$ be given, and assume that $X \neq \Lambda$ and $Y \neq L(\Lambda)$. Then we may choose $\eta \in Y \backslash L(\Lambda)$ and, by Lemma 161C, $k \in N^\times$ such that $(k-1,1)\eta \in X$ and $(k,1)\eta \notin X$. Since $(k-1,1)\eta + \eta \approx (k,1)\eta$, we find$(k-1,1)\eta + \eta \in (X+Y)\backslash X \subset (X \oplus Y)\backslash X$, so that $X \oplus Y \neq X$. This shows that (PS2) holds.

Let $X, Y \in \mathrm{Rng}L\backslash\{\Lambda\}$ be given. We may choose $\mu \in \Lambda\backslash X = U(X)\backslash X$ and $\nu \in \Lambda\backslash Y = U(Y)\backslash Y$. Then $\mu + \nu \in U(X+Y) = U(X \oplus Y)$. Therefore $U(X \oplus Y) \neq \emptyset = U(\Lambda)$, and hence $X \oplus Y \neq \Lambda$. This shows that (PS3) holds.

Let $X, Y \in \mathrm{Rng}L$ be given. If $Y = X \oplus Z$ for some $Z \in \mathrm{Rng}L$, then $X \subset X+Z \subset X \oplus Z = Y$ (since $Z \neq \emptyset$). Conversely, assume that $X \subset Y$, note that $\xi \precapprox \mu$ for all $\xi \in X$ and $\mu \in U(Y)$, and define

$$Z := L(U(Y) - X) \in \mathrm{Rng}L.$$

For all $\xi \in X$, $\zeta \in Z$, and $\mu \in U(Y)$ we have $\zeta \precapprox \mu - \xi$, and hence $\xi + \zeta \precapprox \mu$. Therefore $U(Y) \subset U(X+Z)$, and hence

$$X \oplus Z = L(U(X+Z)) \subset L(U(Y)) = Y. \tag{162.6}$$

On the other hand, let $\mu \in \Lambda\backslash U(Y)$ be given. We may then choose $\eta \in Y$ such that $\mu \prec \eta$. Set $\delta := (1,2)(\eta - \mu)$, and note that $(0,1) \prec \delta$, so that $\delta \in \Lambda\backslash L(\Lambda)$. We now distinguish two cases. If $\mu + \delta \in X$, then $\mu + \delta \in X + Z$; but $\mu \prec \mu + \delta$, and so $\mu \in \Lambda\backslash U(X+Z)$. We are left with the case in which $\mu + \delta \in \Lambda\backslash X = U(X)\backslash X$, and in this case $X \neq \Lambda$. By Lemma 162C we may choose $k \in N^\times$ such that $(k-1,1)\delta \in X$ and $(k,1)\delta \notin X$. Since $\mu + \delta \in U(X)$, we may set $\zeta := (\mu + \delta) - (k-1,1)\delta$ and find, using the definition of δ, that $\zeta \approx \eta - (k,1)\delta$.

For all $\xi \in X$ and $\lambda \in U(Y)$ we have $\xi \prec (k,1)\delta$ and $\eta \precapprox \lambda$, so that $\xi + \eta \prec \lambda + (k,1)\delta$, and therefore $\zeta \approx \eta - (k,1)\delta \prec \lambda - \xi$. It follows that $\zeta \in Z$, and therefore $\mu \prec \mu + \delta \approx \zeta + (k-1,1)\delta \in X + Z$. We conclude that $\mu \in \Lambda\backslash U(X+Z)$ in this case too.

We have shown that $\Lambda\backslash U(Y) \subset \Lambda\backslash U(X+Z)$. Therefore $U(X+Z) \subset U(Y)$, and hence

$$X \oplus Z = L(U(X+Z)) \supset L(U(Y)) = Y.$$

Combining this with (162.6) we find $X \oplus Z = Y$. This completes the proof of the validity of (PS4).

4. It is obvious that $L(\Lambda) = \{0\} \times N^\times \subsetneqq L(\{(1,1)\}) \subsetneqq \Lambda$, so that (PS5) holds.

Let $X \in \mathrm{Rng}L\backslash\{\Lambda, L(\Lambda)\}$ be given, and set

$$Y := L(\{(c,c') \in \Lambda \mid c \neq 0,\ (c',c) \in X\}) \in \mathrm{Rng}L.$$

Let $(x,x') \in X$ and $(y,y') \in Y$ be given. Then either $x = 0$, and in that case $(x,x')(y,y') = (0,x'y') \in L(\{(1,1)\})$; or else $x \neq 0$, and then $(y,y') \precapprox (x',x)$, so that $(x,x')(y,y') \precapprox (1,1)$ and therefore $(x,x')(y,y') \in L(\{(1,1)\})$. We conclude that $XY \subset L(\{(1,1)\})$, and therefore

$$X \odot Y = L(U(XY)) \subset L(U(L(\{(1,1)\}))) = L(\{1,1)\}). \tag{162.7}$$

On the other hand, let $(a,a') \in \Lambda\backslash U(L(\{(1,1)\}))$ be given, so that $a < a'$. Since $X \neq L(\Lambda)$, we may choose $(m,m') \in X\backslash L(\Lambda)$, so that $m \neq 0$. Since $X \neq \Lambda$, we may choose, by Lemma 162C, $k \in N^\times$ such that

$$((k-1)m, a'm') = (k-1,1)(m,a'm') \in X,$$
$$(km, a'm') = (k,1)(m,a'm') \in \Lambda\backslash X = U(X)\backslash X.$$

We have $(m,m') \prec (km,a'm')$, so that $k > a'$, and therefore $k(a'-a) \geq k > a'$, and hence

$$ka < (k-1)a'. \tag{162.8}$$

For all $(x,x') \in X$ with $x \neq 0$ we have $(x,x') \prec (km,a'm')$, and hence $(a'm',km) \prec (x',x)$. It follows that $(a'm',km) \in Y$. From this and (162.8) we find

$$(a,a') \prec (k-1,k) \approx ((k-1)m, a'm')(a'm',km) \in XY.$$

Therefore $(a,a') \in \Lambda\backslash U(XY)$. We conclude that $\Lambda\backslash U(L(\{(1,1)\})) \subset \Lambda\backslash U(XY)$. Then $U(XY) \subset U(L(\{(1,1)\}))$, and therefore

$$X \odot Y = L(U(XY)) \supset L(U(L(\{(1,1)\}))) = L(\{(1,1)\}).$$

From this and (162.7) we conclude that $X \odot Y = L(\{(1,1)\})$. Since $X \in \mathrm{Rng}L\backslash\{\Lambda, L(\Lambda)\}$ was arbitrary, we have shown that (PS6) holds. This concludes the proof. ■

163. Existence

We can now state a precise form of the equivalence of several existence problems.

163A. Theorem. *The following assertions are equivalent:*

(i): *There exists a counting system.*

(ii): *There exists a positivity system.*

(iii): *There exists a complete ordered field.*

(iv): *There exists an ordered field.*

Proof. By Theorem 162A, (i) implies (ii). By Theorem 161E, (ii) implies (iii). It is trivial that (iii) implies (iv). By Theorem 142E, (iv) implies (i). We remark again that no use was made of natural numbers in defining (complete) ordered fields or positivity systems, or in the proofs of Theorems 161E and 142E. ■

This page intentionally left blank

Chapter 17

INFINITE SETS

171. Introduction

In all mathematical work, with insignificant exceptions, we must deal with a profusion of infinite sets. Our intuition is, however, not too well equipped for all aspects of this task, and on occasion fails us altogether. The most basic agreements about sets, mappings, and relations that we have accepted and used are not quite powerful enough to settle some rather natural questions concerning infinite sets. To overcome some of these difficulties, several assertions have been proposed and used as additional agreements or "axioms": an example is the Axiom of Choice, which we have already encountered. Their acceptability has had its ups and downs since they were first made explicit, beginning at the turn of the 20th century. Later, they were shown to be compatible with the more basic agreements about sets and mappings (i.e., their acceptance would not ruin the consistency of these agreements, if they *are* consistent). Finally, it was shown that they are independent of these agreements (i.e., their rejection would not ruin such consistency either; in particular, they could not be proved from more basic principles by accepted rules of inference).

In this chapter we shall introduce the most useful of these assertions. Some have strong intuitive appeal, as, e.g., some forms of the Axiom of Choice. Others seem far less credible: a good example is the assertion that every set can be well-ordered (we cannot *effectively* well-order even the set $\mathbb{R}$). Nevertheless, all the assertions that we present are in fact equivalent: this means that any one of them can be proved, by accepted rules of inference, on the assumption of any other, together with the more basic agreements about sets and mappings.

What attitude is one to adopt with regard to the validity of such "marginal" assertions? This depends, of course, on the purpose at hand. For general mathematics as practiced by the ordinary mathematician, as well as for the educated user of mathematics, the prevailing, and probably most advisable, course is this: to accept the validity of these assertions freely, but to make unobtrusive mention of their use when engaged in careful exposition; and to avoid their use, as a matter of good style, when it is not excessively costly to do so.

Since the purpose of this chapter is, to a large extent, to acquaint the readers with these assertions and, secondarily, to convince them of their equivalence, we do *not*, in this chapter, commit ourselves to the acceptance of any one of these assertions. Instead, we show the derivation of all from a single one, Hausdorff's Maximality Principle, chosen for its simplicity and plausibility.

In broad outline this account follows J. L. Kelley, *General Topology* (Van Nostrand). The quite complicated argument required to complete the proof of the equivalence (Lemmas 177A, 177B) is included only for the record; it is essentially due to Ernst Zermelo (1871-1953), and our account follows P.R. Halmos, *Naive Set Theory* (Van Nostrand).

A word about terminology. The labels "Principle", "Lemma", and "Axiom" signal no difference in status; they should be regarded as quaint and colorful relics of a debate to decide which of these assertions was more "fundamental" than the rest. The attributions by name to individual mathematicians are loosely traditional, and are in some cases more grounded in habit than in history. The assertions to be introduced will be identified by Roman numerals.

It is not our aim, in this chapter, to develop the theory of infinite sets much beyond the purpose already stated. However, in Section 175 we discuss the most elementary facts concerning the comparison of sets by "size", and in Section 176 we do something similar for well-ordered sets. These facts serve as points of departure for so-called transfinite cardinal and ordinal arithmetic, respectively; but we do not pursue these topics in the present account.

172. Maximality principles

Each of the assertions discussed in this section states that some ordered set has a maximal member.

We recall from Section 61 that a collection of sets is called a **nest** if it is totally ordered by inclusion; and that an ordered subset S of an ordered set D is called a **chain of** D if S is totally ordered. In particular, if $\mathcal{D}$ is a collection of sets, a subcollection $\mathcal{S}$ of $\mathcal{D}$ is a chain of $\mathcal{D}$ ordered by inclusion if and only if $\mathcal{S}$ is a nest.

(I) (Hausdorff's Maximality Principle). *In every collection of sets there is a maximal nest; more precisely, if $\mathcal{D}$ is a collection of sets, then the collection $\{\mathcal{N} \in \mathfrak{P}(\mathcal{D}) \mid \mathcal{N}$ is a nest$\}$, ordered by inclusion, has a maximal member.*

(II)[(III)] (Set Maximality Principle). *If $\mathcal{D}$ is a [non-empty] collection of sets such that every [non-empty] nest $\mathcal{N} \in \mathfrak{P}(\mathcal{D})$ satisfies $\bigcup\mathcal{N} \in \mathcal{D}$, then $\mathcal{D}$ ordered by inclusion has a maximal member.*

(IV) (Kuratowski's Lemma). *If D is an ordered set, every chain of D is included in a maximal chain of D, i.e., in a maximal member of the collection of all chains of D, ordered by inclusion.*

(V) (Zorn's Lemma). *If D is an ordered set such that every chain of D has an upper bound, then D has a maximal member.*

172A. Lemma. *Let D be an ordered set and M a maximal chain of D. Then every upper bound of M is a maximal member of D.*

Proof. Let $m \in \mathrm{Ub}(M)$ be given, and let $x \in D$ be given such that $m \prec x$. Then $x \in \mathrm{Ub}(M)$, so that $M \cup \{x\}$ is a chain of D. In view of the maximality of M, we must have $M \cup \{x\} = M$, which implies $x \in M$, and therefore $x \prec m$. We conclude that $x = m$. ■

172B. Lemma. *Let D be an ordered set, and let the chain C of D and the nest $\mathcal{N}$ of subsets of D be given. If $C \cup S$ is a chain of D for every $S \in \mathcal{N}$, then $C \cup \bigcup\mathcal{N}$ is a chain of D.*

Proof. If $\mathcal{N} = \emptyset$, then $C \cup \bigcup\mathcal{N} = C$, and the assertion follows trivially. Assume now that $\mathcal{N} \neq \emptyset$. Let $x, y \in C \cup \bigcup\mathcal{N}$ be given. We may choose $S, T \in \mathcal{N}$ such that $x \in C \cup S$ and $y \in C \cup T$. Since $\mathcal{N}$ is a nest, we have $S \cup T \in \mathcal{N}$, and $C \cup (S \cup T)$ is a chain of D. Since this chain contains x and y, we conclude that $x \prec y$ or $y \prec x$. Since $x, y \in C \cup \bigcup\mathcal{N}$ were arbitrary, we conclude that $C \cup \bigcup\mathcal{N}$ is a chain of D. ■

172C. Proposition. *The assertions* (I), (II), (III), (IV), (V) *are equivalent.*

Proof. We shall prove the following implications:

$$\text{(I)} \Rightarrow \text{(III)} \Rightarrow \text{(II)} \Rightarrow \text{(IV)} \Rightarrow \text{(I)} \quad \text{and} \quad \text{(IV)} \Rightarrow \text{(V)} \Rightarrow \text{(II)}.$$

(I) *implies* (III). Let $\mathcal{D}$ be a non-empty collection of sets such that every non-empty nest $\mathcal{N} \in \mathfrak{P}(\mathcal{D})$ satisfies $\bigcup\mathcal{N} \in \mathcal{D}$. If (I) holds, we may choose a maximal nest $\mathcal{M} \in \mathfrak{P}(\mathcal{D})$, which is clearly not empty. Then $\bigcup\mathcal{M}$ is an upper bound of $\mathcal{M}$ in

$\mathcal{D}$ ordered by inclusion. By Lemma 172A, $\bigcup\mathcal{M}$ is a maximal member of $\mathcal{D}$ ordered by inclusion.

(III) *implies* (II). Let $\mathcal{D}$ be a collection of sets such that every nest $\mathcal{N} \in \mathfrak{P}(\mathcal{D})$ satisfies $\bigcup\mathcal{N} \in \mathcal{D}$. In particular, this holds for every *non-empty* nest, and in addition $\emptyset = \bigcup\emptyset \in \mathcal{D}$, so that $\mathcal{D}$ is not empty. If (III) holds, then $\mathcal{D}$ ordered by inclusion has a maximal member.

(II) *implies* (IV). Let D be an ordered set, and let the chain C of D be given. Set $\mathcal{D} := \{S \in \mathfrak{P}(D) \mid C \cup S \text{ is a chain of } D\}$. By Lemma 172B, every nest $\mathcal{N} \in \mathfrak{P}(\mathcal{D})$ satisfies $\bigcup\mathcal{N} \in \mathcal{D}$.

Assume now that (II) holds. We may then choose a maximal member M of $\mathcal{D}$ ordered by inclusion. Since obviously $C \cup M \in \mathcal{D}$, we have $C \subset M$. We claim that M is a maximal member of the collection of *all* chains of D, ordered by inclusion. Indeed, if N is a chain of D, such that $M \subset N$, we have $C \subset M \subset N$; hence $C \cup N = N$ is a chain of D, so that $N \in \mathcal{D}$. But M is a maximal member of $\mathcal{D}$, and hence $M = N$. Thus M is a maximal chain of D, as claimed.

(IV) *implies* (V). Let D be an ordered set such that every chain of D has an upper bound. If (IV) holds, we may choose a maximal chain of D (by taking the given chain to be $\emptyset$); this maximal chain of D has an upper bound; and by Lemma 172A such an upper bound is a maximal member of D.

(IV) *implies* (I). Apply (IV) to the case in which D is the collection of sets $\mathcal{D}$, ordered by inclusion, and the given chain of $\mathcal{D}$ is the nest $\emptyset$.

(V) *implies* (II). Apply (V) to the case in which D is the collection of sets $\mathcal{D}$, ordered by inclusion, and every nest $\mathcal{N} \in \mathfrak{P}(\mathcal{D})$ has its union as an upper bound in $\mathcal{D}$. ■

172D. Remark. Every non-empty finite ordered set has a maximal member, and every collection of subsets of a finite set is finite (Proposition 105A — applied to the reverse order — and Corollaries 103J and 101F). Therefore the special cases of (I), (II), and (III) in which $\mathcal{D}$ is *finite* and those of (IV) and (V) in which D is *finite*, are valid regardless of the status of the assertions (I), (II), (III), (IV), (V) themselves. ■

173. Collections of finitary character

A collection of sets $\mathcal{D}$ is said to be **of finitary character** if

$$\text{for every set } S, \qquad S \in \mathcal{D} \;\Leftrightarrow\; \mathfrak{F}(S) \subset \mathcal{D};$$

$\mathcal{D}$ is said to be **of binary character** if

$$\text{for every set } S, \qquad S \in \mathcal{D} \;\Leftrightarrow\; (\forall x, y \in S,\;\; \{x, y\} \in \mathcal{D}).$$

173A. Remarks. (a): If $\mathcal{D}$ is a collection of finitary character or of binary character, then

$$\forall S \in \mathcal{D}, \qquad \mathfrak{P}(S) \subset \mathcal{D}. \tag{173.1}$$

(b): $\{x, y\}$ is a finite set for every x, y. This observation, together with (a), shows that every collection of binary character is also of finitary character.

(c): If a collection of finitary character is not empty, it contains $\emptyset$. Every collection of binary character contains $\emptyset$, and hence is not empty.

(d): In verifying that a given collection $\mathcal{D}$ is of finitary character or of binary character, it is possible to avoid the quantification "for every set S", as follows. If D is a set that includes all members of $\mathcal{D}$ (e.g., $D := \bigcup\mathcal{D}$), then $\mathcal{D}$ is of finitary character if and only if

$$\forall S \in \mathfrak{P}(D), \qquad S \in \mathcal{D} \;\Leftrightarrow\; \mathfrak{F}(S) \subset \mathcal{D}, \tag{173.2}$$

and $\mathcal{D}$ is of binary character if and only if

$$\forall S \in \mathfrak{P}(D), \qquad S \in \mathcal{D} \;\Leftrightarrow\; (\forall x, y \in S,\;\; \{x, y\} \in \mathcal{D}). \tag{173.3}$$

The "only if" parts of these equivalences are trivial; to prove the "if" parts, it is enough to observe that if S is a set such that $\{x\} \in \mathcal{D} \subset \mathfrak{P}(D)$ for all $x \in S$, then $S \subset D$. ■

We now formulate maximality principles for collections of finitary character or of binary character.

(VI) (Tukey's Lemma). *Every non-empty collection of sets that is of finitary character has a maximal member when ordered by inclusion.*

(VII) (Tukey's Lemma, Binary Version). *Every collection of sets that is of binary character has a maximal member ordered by inclusion.*

173B. Proposition. (III) *implies* (VI), *and* (VI) *implies* (VII).

Proof. (III) *implies* (VI). Let $\mathcal{D}$ be a non-empty collection of sets that is of finitary character. Let $\mathcal{N}$ be a non-empty nest included in $\mathcal{D}$; we claim that $\bigcup\mathcal{N} \in \mathcal{D}$. Let $T \in \mathfrak{F}(\bigcup\mathcal{N})$ be given. By Proposition 102F we may choose $S \in \mathcal{N}$ such that $T \subset S$. Since $S \in \mathcal{D}$ and $T \in \mathfrak{F}(S)$ and $\mathcal{D}$ is of finitary character, we conclude that $T \in \mathcal{D}$.

Since $T \in \mathfrak{F}(\bigcup\mathcal{N})$ was arbitrary, we have $\mathfrak{F}(\bigcup\mathcal{N}) \subset \mathcal{D}$. Since $\mathcal{D}$ is of finitary character, we conclude that $\bigcup\mathcal{N} \in \mathcal{D}$.

If (III) holds, it follows that $\mathcal{D}$ has a maximal member.

(VI) *implies* (VII). This is trivial, since a collection of binary character is nonempty and of finitary character (Remarks 173A,(b),(c)). ■

▼ **173C. Remark**. Let $n \in \mathbb{N}^\times$ be given. A collection of sets $\mathcal{D}$ is said to be **of $n$ary character** if

$$\text{for every set } S, \qquad S \in \mathcal{D} \;\Leftrightarrow\; \bigcup_{m \in n^\sqsupset} \mathfrak{F}_m(S) \subset \mathcal{D}.$$

The following facts are easily established:

(a): For every $n \in \mathbb{N}^\times$, every collection of $n$ary character satisfies (173.1) and contains $\emptyset$.

(b): For all $m, n \in \mathbb{N}^\times$, if $m \leq n$ then every collection of $m$ary character is of $n$ary character and of finitary character as well.

(c): A collection of sets $\mathcal{D}$ is of 1-ary character if and only if $\mathcal{D} = \mathfrak{P}(S)$ for some set S; $\mathcal{D}$ is of 2-ary character if and only if $\mathcal{D}$ is of binary character.

On account of these remarks, a version of (VII) in which "binary" is replaced by ▲ "$n$ary" is implied by (VI) and implies (VII), for each $n \in \mathbb{N}^\times\backslash\{1\}$. ■

Tukey's Lemma (VI) may be used to yield a similar conclusion for certain collections of sets that are not of finitary character themselves, but are included in collections of finitary character, as we now show.

173D. Lemma. *Let a collection of sets $\mathcal{C}$ and a set A be given. If $\mathcal{C}$ is of finitary character, then so is the subcollection*

$$\mathcal{E} := \{S \in \mathcal{C} \mid A \cup S \in \mathcal{C}\}. \tag{173.4}$$

Proof. Let $S \in \mathcal{E}$ be given. For every $T \in \mathfrak{F}(S)$ we have $T \in \mathcal{C}$ (since $S \in \mathcal{C}$) and $A \cup T \in \mathfrak{P}(A \cup S) \subset \mathcal{C}$ (by Remark 173A,(a)); hence $T \in \mathcal{E}$. Thus $\mathfrak{F}(S) \subset \mathcal{E}$.

Conversely, let S be a set such that $\mathfrak{F}(S) \subset \mathcal{E}$. Then $\mathfrak{F}(S) \subset \mathcal{C}$, and therefore $S \in \mathcal{C}$. Let $T \in \mathfrak{F}(A \cup S)$ be given. Then $T\backslash A \in \mathfrak{F}(S) \subset \mathcal{E}$, and therefore $A \cup T = A \cup (T\backslash A) \in \mathcal{C}$. It follows that $T \in \mathfrak{F}(A \cup T) \subset \mathcal{C}$. Since $T \in \mathfrak{F}(A \cup S)$ was arbitrary, we conclude that $A \cup S \in \mathcal{C}$, and therefore $S \in \mathcal{E}$. ■

173E. Proposition. *Let a collection of sets $\mathcal{D}$ and sets A, B be given. If $\mathcal{D}$ is of finitary character and* (VI) *holds, then the subcollection*

$$\mathcal{F} := \{S \in \mathcal{D} \mid A \subset S \subset B\}, \tag{173.5}$$

ordered by inclusion, has a maximal member if and only if $A \subset B$ and $A \in \mathcal{D}$.

Proof. The collection defined in (173.5) is not empty if and only if $A \subset B$ and $A \in \mathcal{D}$ (by Remark 173A,(a)). Assume that these conditions hold, and note that the collection $\mathcal{C} := \mathcal{D} \cap \mathfrak{P}(B)$ is of finitary character. With $\mathcal{E}$ and $\mathcal{F}$ as defined in (173.4) and (173.5), respectively, we obviously have $\mathcal{F} \subset \mathcal{E}$; moreover, $S \subset A \cup S \in \mathcal{F}$ for

every $S \in \mathcal{E}$. It follows from these remarks that a set is a maximal member of $\mathcal{F}$ if and only if it is a maximal member of $\mathcal{E}$. By Lemma 173D, $\mathcal{E}$ is of finitary character; and $\mathcal{E}$ is not empty, since it contains A. It follows from (VI) that $\mathcal{E}$, and therefore also $\mathcal{F}$, has a maximal member. ■

174. The Axiom of Choice

The idea embodied in the assertions that we shall presently consider is that it is possible to choose *simultaneously* one member from each term of a family of sets — provided the sets are not empty, of course — no matter how large the index set is. (For finite index sets such a choice is always possible, by the Principle of Finite Choice; see Remark 174D,(b).)

We formulate four versions of this kind of assertion. In the first two versions, the family of sets is disjoint; the third version is the "general" one; in the fourth, the choice is made from all non-empty subsets of a given set.

If S is a set, a mapping $\gamma : \mathfrak{P}^{\times}(S) \to S$ is called a **choice-mapping for** S if $\gamma(A) \in A$ for all $A \in \mathfrak{P}^{\times}(S)$.

(VIII) (Surjection Axiom). *Every surjection is right-invertible.*

(IX) (Zermelo's Axiom Of Choice). *If $\mathcal{P}$ is a partition of a set S, there exists a subset K of S such that $K \cap E$ is a singleton for each $E \in \mathcal{P}$.*

(X) (Axiom Of Choice, General Version). *The Cartesian product of a family of non-empty sets is a non-empty set.*

(XI) (Axiom Of Choice, Special Version). *For every set there is a choice-mapping.*

174A. Lemma. *Let the surjection $f\colon D \to C$ and the choice-mapping γ for D be given. Then $(y \mapsto \gamma(f^{<}(\{y\}))) : C \to D$ is a right-inverse of f.*

174B. Proposition. *The assertions* (VIII), (IX), (X), (XI) *are equivalent, and all follow from* (VII).

Proof. We shall prove the following implications:

$$\text{(VII)} \Rightarrow \text{(VIII)} \Rightarrow \text{(X)} \Rightarrow \text{(XI)} \Rightarrow \text{(VIII)} \quad \text{and} \quad \text{(VIII)} \Leftrightarrow \text{(IX)}.$$

We could have saved one implication in this scheme, but Lemma 174A is useful elsewhere.

(VII) *implies* (VIII). Let the surjection $f\colon D \to C$ be given, and set $\mathcal{D} := \{A \in \mathfrak{P}(D) \mid f|_A \text{ is injective}\}$. For a given subset S of D, we have $S \in \mathcal{D}$ if and only if $f|_{\{x,y\}}$ is injective for all $x, y \in S$, i.e., if and only if $\{x, y\} \in \mathcal{D}$ for all $x, y \in S$. We have shown that $\mathcal{D}$ satisfies (173.3) and hence is a collection of binary character.

Assume now that (VII) holds. We may then choose a maximal member M of $\mathcal{D}$ ordered by inclusion. If the injective mapping $f|_M : M \to C$ were not surjective, we could choose $y \in C \backslash f_{>}(M)$ and $x \in f^{<}(\{y\})$, since f was surjective, and we should find that $x \notin M$ but $M \cup \{x\} \in \mathcal{D}$, which would contradict the maximality of M. Therefore $f|_M$ is surjective, and hence invertible. Then $(f|_M)^{\leftarrow}|^D = 1_{M \subset D} \circ (f|_M)^{\leftarrow}$ is a right-inverse of f: indeed, $f \circ 1_{M \subset D} \circ (f|_M)^{\leftarrow} = (f|_M) \circ (f|_M)^{\leftarrow} = 1_C$.

(VIII) *implies* (IX). If $\mathcal{P}$ is a partition of the set S, and the surjection $\Omega_{\mathcal{P}} : S \to \mathcal{P}$ has a right-inverse, the range of each such right-inverse meets every member of $\mathcal{P}$ in a singleton.

(IX) *implies* (VIII). Assume that (IX) holds, and let the surjection $f\colon D \to C$ be given. We apply (IX) to the partition Part$f$ of D, and choose a subset K of D

such that $K \cap f^<(\{y\})$ is a singleton for every $y \in C$. Then the mapping $g\colon C \to D$ defined by $g(y) :\in K \cap f^<(\{y\})$ for all $y \in C$ is a right-inverse of f.

(VIII) *implies* (X). Let $(A_i \mid i \in I)$ be a family of non-empty sets. Consider the mappings $\lambda : \dot{\bigcup_{i\in I}} A_i \to I$ and $\mu : \dot{\bigcup_{i\in I}} A_i \to \bigcup_{i\in I} A_i$ defined by

$$\lambda((j,x)) := j \quad \text{and} \quad \mu((j,x)) := x \qquad \text{for all } (j,x) \in \dot{\bigcup_{i\in I}} A_i.$$

Note that

$$\mu(s) \in A_{\lambda(s)} \qquad \text{for all } s \in \dot{\bigcup_{i\in I}} A_i \tag{174.1}$$

and that λ is surjective, since $A_i \neq \emptyset$ for all $i \in I$.

Assume now that (VIII) holds and choose a right-inverse of λ, say $\nu : I \to \dot{\bigcup_{i\in I}} A_i$. Then (174.1) implies

$$\mu(\nu(j)) \in A_{\lambda(\nu(j))} = A_j \qquad \text{for all } j \in I.$$

Therefore the family $(\mu(\nu(i)) \mid i \in I)$ is a member of $\bigtimes_{i\in I} A_i$, and this product is not empty.

(X) *implies* (XI). Let the set S be given, and assume that (X) holds. Then $\bigtimes_{A\in\mathfrak{P}^\times(S)} A \neq \emptyset$, and we may choose $c \in \bigtimes_{A\in\mathfrak{P}^\times(S)} A$. Then $\gamma : \mathfrak{P}^\times(S) \to S$, defined by $\gamma(A) := c_A$ for all $A \in \mathfrak{P}^\times(S)$, is a choice-mapping for S.

(XI) *implies* (VIII). This is an immediate consequence of Lemma 174A. ■

Closely connected to the Axiom of Choice are some of the general distributive laws for families of sets (see Section 45).

(XII)[(XIII)] (General Distributive Law For Products [Intersections]). *Let a set I [and a set X], a family of sets $(J_i \mid i \in I)$, and a family of families of sets $((A_{i,j} \mid j \in J_i) \mid i \in I)$ be given [such that $\bigcup_{i\in I}\bigcup_{j\in J_i} A_{i,j} \subset X$]. Set $P := \bigtimes_{i\in I} J_i$. Then*

$$\bigtimes_{i\in I}(\bigcup_{j\in J_i} A_{i,j}) = \bigcup_{k\in P}(\bigtimes_{i\in I} A_{i,k_i}) \qquad [\bigcap_{i\in I}{}^X(\bigcup_{j\in J_i} A_{i,j}) = \bigcup_{k\in P}(\bigcap_{i\in I}{}^X A_{i,k_i})].$$

174C. Proposition. *The assertions* (X), (XII), (XIII) *are equivalent.*

Proof. The implications (X) ⇒ (XII) and (X) ⇒ (XIII) were proved in Theorem 45A (proofs of (45.5) and (45.3)). To prove the converse implications, we assume that (XII)[(XIII)] holds and that the family of non-empty sets $(J_i \mid i \in I)$ is given. We set $P := \bigtimes_{i\in I} J_i$ and define the family of families of sets $((A_{i,j} \mid j \in J_i) \mid i \in I)$ by $A_{i,j} := \{\emptyset\}$ for all $i \in I$ and $j \in J_i$ [and the set X by $X := \{\emptyset\}$]. Then $\bigcup_{j\in J_i} A_{i,j} = \{\emptyset\}$ for all $i \in I$, and $\bigtimes_{i\in I} A_{i,k_i} = \{\emptyset\}^I$ $[\bigcap_{i\in I}{}^X A_{i,k_i} = \{\emptyset\}]$ for all $k \in P$.

Therefore

$$\bigcup_{k\in P}\{\emptyset\}^I = \bigtimes_{i\in I}\{\emptyset\} = \{\emptyset\}^I \neq \emptyset \quad [\bigcup_{k\in P}\{\emptyset\} = \bigcap_{i\in I}^{\{\emptyset\}}\{\emptyset\} = \{\emptyset\} \neq \emptyset]$$

and consequently $P \neq \emptyset$. ■

174D. Remarks. (a): If in (XIII) we replace $\bigcup$ by $\bigcap^X$ and vice versa, we obtain still another equivalent assertion; the equivalence follows by "taking complements with respect to X" (cf. Theorem 45A, proof of (45.4)).

(b): The special cases of (VIII) with finite codomain, of (IX) with finite partition, of (XI) with finite set, and of (X), (XII), (XIII) with finite index set are valid — regardless of the status of the general assertions (VIII)-(XIII) — by virtue of the Principle of Finite Choice (Theorem 103L) . ■

175. Comparison of sets

We recall that sets S and T are said to be **equinumerous** if there exists a bijection from S to T or, equivalently, a bijection from T to S. For finite sets, this word is well chosen, since finite sets are equinumerous if and only if they have the same cardinal number (Corollary 101D). If S and T are finite sets, then $\#S \leq \#T$ if and only if there exists an injection from S to T (Proposition 101E). We use this observation to motivate the following terminology. Let the sets S and T be given. Then S is said to be **outnumbered by** T, and T is said to **outnumber** S, if there exists an injection from S to T. It is clear that if S, T, U are sets and U outnumbers T and T outnumbers S, then U outnumbers S. If T outnumbers S but S does not outnumber T, then S is said to be **strictly outnumbered by** T, and T is said to **strictly outnumber** S.

If the sets S and T are equinumerous, it is obvious that T outnumbers S and S outnumbers T. It is reasonable to ask whether the reverse implication holds. It is also reasonable to ask whether, for a given pair of sets, one set in the pair must outnumber the other. The answer to the former question is affirmative, as we shall presently recall. An affirmative answer to the latter question, however, is an assertion that will eventually be seen to be equivalent to the Axiom of Choice.

175A. Theorem. (Schröder-Bernstein Theorem). *Let the sets S and T be given. Then S and T are equinumerous if (and only if) T outnumbers S and S outnumbers T.*

Proof. This is a restatement of Theorem 75C. ■

175B. Proposition. *Every set is strictly outnumbered by its power set.*

Proof. Let the set S be given. The mapping $(s \mapsto \{s\}) : S \to \mathfrak{P}(S)$ is obviously injective. If there existed an injection from $\mathfrak{P}(S)$ to S, we could choose one, say f. Since $\mathfrak{P}(S) \neq \emptyset$, we could choose a left-inverse of f, and this left-inverse would be a surjection from S to $\mathfrak{P}(S)$. But there is no surjection from S to $\mathfrak{P}(S)$, as was shown in Proposition 32E. ■

We formulate the following assertions.

(XIV) (Principle Of Comparability). *If S and T are sets, then either T outnumbers S or S outnumbers T.*

(XV) (Principle Of Comparability For Surjections). *If S and T are non-empty sets, then there exists either a surjection from S to T or a surjection from T to S.*

175C. Proposition. (VII) *implies* (XIV) *and* (XIV) *implies* (XV).

Proof. (VII) *implies* (XIV). Assume that (VII) holds. Let the sets S and T be given, and consider the collection

$$\mathcal{G} := \{G \in \mathfrak{P}(S \times T) \mid \forall (x, y), (x', y') \in G, \quad x = x' \;\Leftrightarrow\; y = y'\}.$$

It is plain that $\mathcal{G}$ is a collection of binary character. By (VII) we may therefore choose a maximal member M of $\mathcal{G}$.

We claim that

(175.1) for every $(x, y) \in S \times T$ there exists $(x', y') \in S \times T$ such that $(x, y') \in M$ or $(x', y) \in M$.

Indeed, suppose that $(u, v) \in S \times T$ were such that $(u, y') \notin M$ and $(x', v) \notin M$ for all $(x', y') \in S \times T$. Then we should find that $(u, v) \notin M$ and that $M \cup \{(u, v)\} \in \mathcal{G}$, contradicting the maximality of M. This establishes the claimed validity of (175.1).

It follows from (175.1) that either

(175.2) for every $x \in S$ there is at least one $y \in T$ such that $(x, y) \in M$

or

(175.3) for every $y \in T$ there is at least one $x \in S$ such that $(x, y) \in M$.

Moreover, since $M \in \mathcal{G}$, we also have, in every case,

(175.4) for every $x \in S$ there is at most one $y \in T$ such that $(x, y) \in M$, and for every $y \in T$ there is at most one $x \in S$ such that $(x, y) \in M$.

If (175.2) holds, then (175.2) and (175.4) show that M is the graph of an injection from S to T. If, on the other hand, (175.3) holds, then (175.3) and (175.4) show that $\{(y, x) \in T \times S \mid (x, y) \in M\}$ is the graph of an injection from T to S. Hence either T outnumbers S or S outnumbers T.

(XIV) *implies* (XV). Every injection with non-empty domain is left-invertible, and each left-inverse is surjective. The desired implication is an immediate consequence of this observation. ■

175D. Remarks. (a): Theorem 175A makes it appear desirable to assign to every set S, whether finite or infinite, an object $\#S$, the *cardinal of* S, in such a way that the following condition is satisfied: If S and T are sets, then $\#S = \#T$ if and only if S and T are equinumerous. The objects that occur as cardinals of sets would be the *cardinal numbers*; and we could define a "relation" $\leq$ among cardinal numbers by requiring $\#S \leq \#T$ if and only if T outnumbers S. This "relation" is obviously reflexive and transitive, and Theorem 175A shows that it is also antisymmetric. It is therefore an "order". Assertion (XIV) is then equivalent to the assertion that this "order" is total; (XIV) is therefore also known as the *Law of Trichotomy* (because it asserts that either $\#S < \#T$ or $\#S = \#T$ or $\#T < \#S$, but no two at the same time, for any sets S, T; here $<$ is the "strict-order" corresponding to $\leq$).

The question that must be answered to make this idea effective is, What objects are the cardinal numbers to be? We shall not explore this matter any further, but we mention that a satisfactory answer can indeed be given, and that one can arrange to have the natural numbers as the "finite" cardinal numbers, i.e., the cardinals of finite sets. The cardinal numbers, both finite and infinite (or "transfinite", as they

are sometimes called), constitute the subject matter of a branch of set theory called cardinal arithmetic. They can be added and multiplied and raised to powers in such a way that all the results proved in Section 103 remain valid with the assumption of finiteness deleted. A glimpse of some perhaps unexpected facts of cardinal arithmetic can be obtained from results such as Corollary 175F.

Proposition 175B shows that there is no "greatest" cardinal number, since $\#S < \#\mathfrak{P}(S)$ for every set S. This suggests that there are "many" cardinal numbers. There are indeed so "many" that it is not proper to speak of "the set of all cardinal numbers", just as it is not proper to speak of "the set of all sets". That is why the words "relation", "order", "strict-order", and "greatest" above were put in quotation marks; for every *set* of cardinal numbers, however, the "restriction" of $\leq$ to that set *is* an order, and this order is total if (XIV) holds.

(b): If the Axiom of Countable Choice holds (it is a consequence of (X), and hence of (VII) by Proposition 174B), then every infinite set outnumbers every countable set (Corollary 121W). Consequently, $\#\mathbb{N}$ is the least infinite cardinal number; it is often denoted by $\aleph_0$. ■

The rest of this section contains some useful additional results concerning the size of infinite sets. They all depend on assertion (VII) or (III), as indicated in each case.

175E. Theorem. *If* (VII) *holds, then every infinite set has a partition whose members are countably infinite sets.*

Proof. Let the infinite set S be given, and consider the collection Δ of all disjoint collections of countably infinite subsets of S. It is obvious that Δ is a collection of binary character. By (VII) we may therefore choose a maximal member $\mathcal{M}$ of Δ; thus $\mathcal{M}$ is a maximal disjoint collection of countably infinite subsets of S.

By (VII) and Proposition 174B, the Axiom of Choice (X) holds, and consequently we may apply Theorem 121V. If $S\backslash(\bigcup\mathcal{M})$ were infinite, that theorem shows that we might choose a countably infinite subset N of $S\backslash(\bigcup\mathcal{M})$; but then $\mathcal{M}\cup\{N\} \in \Delta$, and this would contradict the maximality of $\mathcal{M}$. We conclude that $S\backslash(\bigcup\mathcal{M})$ is finite.

If $\mathcal{M}$ were empty, we should find that $S = S\backslash(\bigcup\mathcal{M})$ is finite, contrary to the assumption. Therefore $\mathcal{M}$ is not empty, and we may choose $M \in \mathcal{M}$. Now $K := S\backslash(\bigcup(\mathcal{M}\backslash\{M\})) = (S\backslash(\bigcup\mathcal{M})) \cup M$, and this union of a finite set and a countably infinite set is countably infinite (this follows from Corollary 121O, but can easily be shown directly from the definitions). We conclude that $(\mathcal{M}\backslash\{M\}) \cup \{K\}$ is the desired partition of S into countably infinite subsets. ■

175F. Corollary. *If* (VII) *holds, and if S is an infinite set and T a countable non-empty set, then the sets S, $S \times T$, and $T \times S$ are equinumerous.*

Proof. By Theorem 175E we may choose a partition $\mathcal{P}$ of S whose members are countably infinite sets. By Corollary 121F, each member of $\mathcal{P}$ is equinumerous with $\mathbb{N}$. By (X), which follows from (VII) by Proposition 174B, we may choose a family

of invertible mappings $(\phi_E \mid E \in \mathcal{P}) \in \bigtimes_{E\in\mathcal{P}} \mathrm{Inv}(E, \mathbb{N})$. Each of the mappings

$$(x \mapsto (\Omega_{\mathcal{P}}(x), \phi_{\Omega_{\mathcal{P}}(x)}(x))) : S \to \mathcal{P} \times \mathbb{N}$$

and

$$((E, n) \mapsto (\phi_E)^{\leftarrow}(n)) : \mathcal{P} \times \mathbb{N} \to S$$

is the inverse of the other. We conclude that

(175.5) S and $\mathcal{P} \times \mathbb{N}$ are equinumerous;

it follows that

(175.6) $S \times T$ and $(\mathcal{P} \times \mathbb{N}) \times T$ are equinumerous.

Now $\mathbb{N} \times T$ is infinite (since $T \neq \emptyset$) and is outnumbered by $\mathbb{N} \times \mathbb{N}$ (since T is countable); but $\mathbb{N} \times \mathbb{N}$ is countable, and hence so is $\mathbb{N} \times T$ (Corollary 121L and Proposition 121B). Thus $\mathbb{N} \times T$ is countably infinite, and hence equinumerous to $\mathbb{N}$ (Corollary 121F). Hence

(175.7) $\mathcal{P} \times \mathbb{N}$ and $\mathcal{P} \times (\mathbb{N} \times T)$ are equinumerous.

Since $(\mathcal{P} \times \mathbb{N}) \times T$ and $\mathcal{P} \times (\mathbb{N} \times T)$ are obviously equinumerous, it follows from (175.5), (175.7), and (175.6) that S and $S \times T$ are equinumerous. On the other hand, $S \times T$ and $T \times S$ are of course equinumerous. ■

175G. Lemma. *Assume that* (VII) *holds and let the sets K and L be given. Let the family of sets $(A_k \mid k \in K)$ be given and assume that L outnumbers A_k for every $k \in K$. Then $K \times L$ outnumbers both $\dot{\bigcup_{k\in K}} A_k$ and $\bigcup_{k\in K} A_k$.*

Proof. By Proposition 174B, (VII) implies (VIII) and (X). By (X) we may choose a family of injections $(f_k \mid k \in K) \in \bigtimes_{k\in K} \mathrm{Map}(A_k, L)$. The mapping

$$((k, x) \mapsto (k, f_k(x))) : \dot{\bigcup_{k\in K}} A_k \to K \times L$$

is injective. Therefore $K \times L$ outnumbers $\dot{\bigcup_{k\in K}} A_k$.

The mapping $((k, x) \mapsto x) : \dot{\bigcup_{k\in K}} A_k \to \bigcup_{k\in K} A_k$ is surjective. By (VIII) we may choose a right-inverse of this mapping, and this right-inverse is injective. Hence $\dot{\bigcup_{k\in K}} A_k$ outnumbers $\bigcup_{k\in K} A_k$. ■

175H. Corollary. *Assume that* (VII) *holds, and let the infinite set I be given. Then:*

(a): *If $(A_j \mid j \in J)$ is a countable family of sets such that I outnumbers A_j for every $j \in J$, then I outnumbers $\dot{\bigcup_{j\in J}} A_j$ and $\bigcup_{j\in J} A_j$.*

(b): *If* $(A_i \mid i \in I)$ *is a family of countable sets, then* I *outnumbers* $\dot{\bigcup_{i\in I}} A_i$ *and* $\bigcup_{i\in I} A_i$.

Proof. We apply Lemma 175G with $K := J$ and $L := I$ for Part (a) (the case $J = \emptyset$ is trivial), and with $K := I$ and $L := \mathbb{N}$ for Part (b). The proof is completed by using Corollary 175F to show that I outnumbers $J \times I$ and $I \times \mathbb{N}$, respectively. ■

175I. Corollary. *Assume that* (VII) *holds. If* $(A_j \mid j \in J)$ *is a countable family of sets and* $k \in J$ *is such that* A_k *is infinite and outnumbers* A_j *for every* $j \in J$*, then the sets* A_k, $\dot{\bigcup_{j\in J}} A_j$, *and* $\bigcup_{j\in J} A_j$ *are equinumerous.*

Proof. By Corollary 175H,(a) with $I := A_k$, we conclude that A_k outnumbers $\dot{\bigcup_{j\in J}} A_j$ and $\bigcup_{j\in J} A_j$. On the other hand, the mapping $\sigma_k : A_k \to \dot{\bigcup_{j\in J}} A_j$ and the inclusion mapping of A_k into $\bigcup_{j\in J} A_j$ are injective, so that $\dot{\bigcup_{j\in J}} A_j$ and $\bigcup_{j\in J} A_j$ each outnumbers A_k. The conclusion follows by Theorem 175A. ■

175J. Theorem. *Assume that* (III) *holds. Let the infinite set* S *be given. Then* S *and* $\mathfrak{F}(S)$ *are equinumerous.*

Proof. 1. Since (III) holds, we also have (VII) (Proposition 173B), (X) (Proposition 174B), and (XIV) (Proposition 175C).

2. Consider the subcollection

$$\mathcal{G} := \{\mathrm{Gr}\phi \mid \phi \in \mathrm{Inv}(A, \mathfrak{F}(A)) \text{ for some infinite } A \in \mathfrak{P}(S)\}$$

of $\mathfrak{P}(S \times \mathfrak{F}(S))$. We shall use the mappings $\pi_1 : S \times \mathfrak{F}(S) \to S$ and $\pi_2 : S \times \mathfrak{F}(S) \to \mathfrak{F}(S)$ defined by

$$\pi_1(x, F) := x \quad \text{and} \quad \pi_2(x, F) := F \qquad \text{for all } (x, F) \in S \times \mathfrak{F}(S).$$

We note that

$$(\pi_2)_>(G) = \mathfrak{F}((\pi_1)_>(G)) \qquad \text{for all } G \in \mathcal{G}. \tag{175.8}$$

Our aim in this part in this part of the proof is to show that $\mathcal{G}$ ordered by inclusion has a maximal member. To this end it will be sufficient to prove, since (III) holds, that

$$G \neq \emptyset \tag{175.9}$$

$$\bigcup \mathcal{N} \in \mathcal{G} \qquad \text{for every non-empty nest } \mathcal{N} \in \mathfrak{P}(\mathcal{G}). \tag{175.10}$$

Since S is infinite, we may choose a countably infinite subset N of S (Theorem 121V, which depends on the Axiom of Countable Choice, which follows from (X)). Since N and $\mathfrak{F}(N)$ are equinumerous (Corollary 121F and Theorem 121J), we may choose a bijection $\nu : N \to \mathfrak{F}(N)$. Then $\mathrm{Gr}\nu \in \mathcal{G}$, and (175.9) is proved.

To prove (175.10), let the non-empty nest $\mathcal{N} \in \mathfrak{P}(\mathcal{G})$ be given. It is clear that $\bigcup\mathcal{N}$ is the graph of a bijection from $M := (\pi_1)_>(\bigcup\mathcal{N})$ to $\mathcal{M} := (\pi_2)_>(\bigcup\mathcal{N})$ (cf. Theorem 43D). The proof of (175.10) will be complete when we show that

$$\mathcal{M} = \mathfrak{F}(M). \tag{175.11}$$

Using (175.8), we find

$$M = (\pi_1)_>(\bigcup\mathcal{N}) = \bigcup((\pi_1)_>)_>(\mathcal{N}) \tag{175.12}$$

$$\begin{aligned}\mathcal{M} = (\pi_2)_>(\bigcup\mathcal{N}) \ &= \bigcup((\pi_2)_>)_>(\mathcal{N}) = \bigcup\{\mathfrak{F}((\pi_1)_>)(G)) \mid G \in \mathcal{N}\} \subset \\ &\subset \mathfrak{F}(\bigcup((\pi_1)_>)_>(\mathcal{N})) = \mathfrak{F}(M).\end{aligned} \tag{175.13}$$

Conversely, let $F \in \mathfrak{F}(M)$ be given. By (175.12) and the Principle of Finite Choice (Theorem 103L), we may choose a non-empty finite subcollection $\mathcal{F}$ of $\mathcal{N}$ such that $F \in \bigcup((\pi_1)_>)_>(\mathcal{F})$. Since $\mathcal{F}$ is a non-empty finite nest, we have $\bigcup\mathcal{F} \in \mathcal{F} \subset \mathcal{G}$ (Corollary 105C), and $F \subset (\pi_1)_>(\bigcup\mathcal{F})$. Therefore, by (175.8),

$$F \in \mathfrak{F}((\pi_1)_>(\bigcup\mathcal{F})) = (\pi_2)_>(\bigcup\mathcal{F}) \subset (\pi_2)_>(\bigcup\mathcal{N}) = \mathcal{M}.$$

Since $F \in \mathfrak{F}(M)$ was arbitrary, we conclude that $\mathfrak{F}(M) \subset \mathcal{M}$. Together with (175.13) this establishes (175.11).

This completes the proof of the assertion that $\mathcal{G}$ ordered by inclusion has a maximal member.

3. Choose a maximal member H of $\mathcal{G}$, and set $M : (\pi_1)_>(H)$, so that, by (175.8), $\mathfrak{F}(M) = (\pi_2)_>(H)$. By the definition of $\mathcal{G}$, we note that M is infinite. Our aim in this part of the proof is to show that M outnumbers $S\backslash M$.

Suppose not; then $S\backslash M$ outnumbers M, by (XIV), and we may choose a subset P of $S\backslash M$ such that M and P are equinumerous. It follows from Corollary 175I that M and $M \cup P$ are equinumerous, and therefore $\mathfrak{F}(M), \mathfrak{F}(P)$, and $\mathfrak{F}(M \cup P)$ are equinumerous. Now P is infinite and $\mathfrak{F}(P) = \{\emptyset\} \cup \mathfrak{F}^\times(P)$. By Corollary 175I, $\mathfrak{F}(P)$ and $\mathfrak{F}^\times(P)$ are equinumerous. But

$$\mathfrak{F}^\times(P) \subset \mathfrak{F}(M \cup P)\backslash\mathfrak{F}(M) \subset \mathfrak{F}(M \cup P);$$

therefore $\mathfrak{F}(M \cup P)\backslash\mathfrak{F}(M)$ and $\mathfrak{F}^\times(P)$ are equinumerous (Theorem 175A). But M and $\mathfrak{F}(M)$ are equinumerous (H is the graph of a bijection from M to $\mathfrak{F}(M)$). We conclude that P and $\mathfrak{F}(M \cup P)\backslash\mathfrak{F}(M)$ are equinumerous, and we may choose a bijection $\psi : P \to \mathfrak{F}(M \cup P)\backslash\mathfrak{F}(M)$. But then $H \cup \mathrm{Gr}\psi$ is the graph of a bijection from $M \cup P$ to $\mathfrak{F}(M \cup P)$, hence a member of $\mathcal{G}$ that properly includes H, contradicting the maximality of H. Our supposition that M does not outnumber $S\backslash M$ is thus untenable.

4. Thus M outnumbers $S\backslash M$. By Corollary 175I it follows that M and $S = M \cup (S\backslash M)$ are equinumerous. Hence $\mathfrak{F}(M)$ and $\mathfrak{F}(S)$ are equinumerous. But M and $\mathfrak{F}(M)$ are equinumerous, as noted before. We conclude that S and $\mathfrak{F}(S)$ are equinumerous, as was to be proved. ■

175K. Corollary. *Assume that* (III) *holds. Let the infinite set S be given. Then S and $S \times S$ are equinumerous.*

Proof. $S \times S$ obviously outnumbers S. On the other hand, the mapping

$$(x, y) \mapsto \{\{x\}, \{x, y\}\} : S \times S \to \mathfrak{F}(\mathfrak{F}(S))$$

is injective (cf. Remark 17D), so that $\mathfrak{F}(\mathfrak{F}(S))$ outnumbers $S \times S$; but by Theorem 175J (applied twice), S and $\mathfrak{F}(\mathfrak{F}(S))$ are equinumerous, so that S outnumbers $S \times S$. The conclusion then follows by Theorem 175A. ■

175L. Corollary. *Assume that* (III) *holds. Let the finite family of non-empty sets $(A_i \mid i \in I)$ and $k \in I$ be given, and assume that A_k is infinite and outnumbers A_i for every $i \in I$. Then A_k and $\bigtimes_{i\in I} A_i$ are equinumerous.*

Proof. By the Principle of Finite Choice (Theorem 103L) we may choose a family of injections $(\phi_i \mid i \in I) \in \bigtimes_{i\in I} \mathrm{Map}(A_i, A_k)$. Then $\bigtimes_{i\in I} \phi_i : \bigtimes_{i\in I} A_i \to (A_k)^I$ is injective; hence $(A_k)^I$ outnumbers $\bigtimes_{i\in I} A_i$. Now A_k is infinite; by Corollary 175K and an obvious proof by special induction, it follows that A_k and $(A_k)^I$ are equinumerous; hence A_k outnumbers $\bigtimes_{i\in I} A_i$.

Since $A_i \neq \emptyset$ for all $i \in I$, it follows form Theorem 103L that $\bigtimes_{i\in I} A_i \neq \emptyset$. Now $\pi_k : \bigtimes_{i\in I} A_i \to A_k$ is surjective (Proposition 44B); hence $\bigtimes_{i\in I} A_i$ outnumbers A_k. The conclusion follows by Theorem 175A. ■

176. Well-ordered sets

In earlier days it appeared desirable to apply inductive-proof schemes to the proof of properties of all members of an infinite set of arbitrary "size". In order to do so, the set had to be provided with an order such that the set would be well-ordered. The assertion that this can always be done is the Well-Ordering Principle, and this was shown to be equivalent to the Axiom of Choice. In current mathematical practice, the use of the Well-Ordering Principle has largely been superseded by the more practical use of maximality principles; it retains its usefulness, however, in special fields of mathematics (e.g., general topology), and also in its own set-theoretical context.

We begin by obtaining a result (Theorem 176C), of interest in its own right, which says that well-ordered sets can be compared very precisely as to their "length".

We recall from Section 81 that, if D is an ordered set and $x \in D$ is given, we denote by $\mathrm{Spr}(x)$ the set of all members of D that strictly precede x; this is also the set of strict lower bounds of $\{x\}$.

Let the well-ordered sets D and D' be given. A mapping $f\colon D \to D'$ is called a **matching from D to D'** if it satisfies

$$f_>(\mathrm{Spr}(x)) = \mathrm{Spr}'(f(x)) \qquad \text{for all } x \in D. \tag{176.1}$$

(We use "primed" symbols for the ordered set D'.)

176A. Remarks. (a): The definition of *matching* would be meaningful for any ordered sets D and D', but we shall use it only for well-ordered sets.

(b): Every matching is strictly isotone, and hence injective. A mapping $f\colon D \to D'$ is an order-isomorphism if (and only if) it is both a matching and a surjection (Proposition 62D).

(c): For every well-ordered set D and every $x \in D$, the inclusion mapping $1_{\mathrm{Spr}(x)\subset D}$ is a matching from the (well-)ordered subset $\mathrm{Spr}(x)$ of D to D.

(d): A composite of matchings is a matching. ■

176B. Lemma. *Let the well-ordered sets D and D' be given. If $f\colon D \to D'$ is a matching but not an order-isomorphism, then there is exactly one $x' \in D'$ such that $f|^{\mathrm{Rng}}$ is an order-isomorphism from D to $\mathrm{Spr}'(x')$.*

Proof. By Remark 176A,(b), f is not surjective. For every $y' \in \mathrm{Rng}f$ we have $\mathrm{Spr}'(y') \subset \mathrm{Rng}f$, on account of (176.1). By Lemma 82F, there is exactly one $x' \in D'$ such that $\mathrm{Rng}f = \mathrm{Spr}'(x')$. Since f is strictly isotone and D is totally ordered, $f|^{\mathrm{Rng}} : D \to \mathrm{Spr}'(x')$ is an order-isomorphism. ■

176C. Theorem. *Let the well-ordered sets D and D' be given. Then there exists a matching from D to D' or a matching from D' to D, and there is at most one matching of each kind. If there is a matching from D to D' and also a matching from D' to D, then each is the inverse of the other, and they are order-isomorphisms.*

Proof. 1. We first prove the uniqueness assertion. Let the matchings f and g from D to D' be given. Let $x \in D$ be given and assume that $f|_{\mathrm{Spr}(x)} = g|_{\mathrm{Spr}(x)}$. Then $\mathrm{Spr}'(f(x)) = f_>(\mathrm{Spr}(x)) = g_>(\mathrm{Spr}(x)) = \mathrm{Spr}'(g(x))$. Since D' is totally ordered, this

implies $f(x) = g(x)$. Since $x \in D$ was arbitrary, we have proved by induction that $f(x) = g(x)$ for all $x \in D$, so that $f = g$. This shows that there is at most one matching from D to D'. Repeating the argument with D and D' interchanged, we conclude that there is at most one matching from D' to D.

Assume that $f\colon D \to D'$ and $f' : D' \to D$ are matchings. By Remark 176A,(d), $f' \circ f$ is a matching from D to D, but so is 1_D. Since there is at most one matching from D to D (by the argument of the preceding paragraph), we must have $f' \circ f = 1_D$. In a similar manner we conclude that $f \circ f' = 1_{D'}$. We have shown that each of f and f' is the inverse of the other. Since they are isotone, they are order-isomorphisms.

2. It remains to prove the existence assertion. We use a method of recursive definition with possible "break-down" of the recursion rule. We apply Theorem 82E with $I := D$, $A_x := D'$ for all $x \in D$, and

$$U_x := \{u \in D'^{\mathrm{Spr}(x)} \mid \mathrm{Rng}u \neq D'\} \qquad \text{for all } x \in D,$$

$$\phi_x(u) := \min(D' \backslash \mathrm{Rng}u) \qquad \text{for all } x \in D \text{ and } u \in U_x.$$

By the theorem just quoted we conclude that there exists a (unique) subset K of D and family $a \in D'^K$ such that

$$x \in K \;\Leftrightarrow\; (\mathrm{Spr}(x) \subset K \;\text{ and }\; a_>(\mathrm{Spr}(x)) \neq D') \qquad \text{for all } x \in D, \tag{176.2}$$

$$a_x = \min(D' \backslash a_>(\mathrm{Spr}(x))) \qquad \text{for all } x \in K. \tag{176.3}$$

Let $x \in K$ be given. For every $y \in \mathrm{Spr}(x)$ we have $\mathrm{Spr}(y) \subset \mathrm{Spr}(x)$, and (176.2) and (176.3) yield $y \in K$ and $\mathrm{Spr}'(a_y) \subset a_>(\mathrm{Spr}(y)) \subset a_>(\mathrm{Spr}(x))$. We apply Lemma 82F,((i)⇒(ii)), to the well-ordered set D' and the subset $a_>(\mathrm{Spr}(x))$; this subset is not D' in view of (176.2), and therefore $a_>(\mathrm{Spr}(x)) = \mathrm{Spr}'(\min(D' \backslash a_>(\mathrm{Spr}(x))))$. Combining this with (176.3) and noting that $x \in K$ was arbitrary, we find

$$\mathrm{Spr}'(a_x) = a_>(\mathrm{Spr}(x)) \qquad \text{for all } x \in K. \tag{176.4}$$

Two cases must now be distinguished. In the first case, $K = D$, and (176.4) shows that the mapping $(x \mapsto a_x) : D \to D'$ is a matching from D to D'. In the second case, $K \neq D$; we now examine this case. From (176.2) and Lemma 82F we have $K = \mathrm{Spr}(k)$ for some (unique) $k \in D$. Since $k \notin K$ but $\mathrm{Spr}(k) = K$, it follows from (176.2) that $\mathrm{Rng}a = a_>(\mathrm{Spr}(k)) = D'$. Since (176.4) shows that a is strictly isotone, hence injective, the mapping $(x \mapsto a_x) : K \to D'$ is bijective. If $f' : D' \to K$ is its inverse, we have by (176.4),

$$f'_>(\mathrm{Spr}'(x')) = \mathrm{Spr}(f'(x')) \qquad \text{for all } x' \in D'.$$

Therefore the mapping $f'|^D$ is a matching from D' to D. ■

176D. Corollary. *Let the well-ordered sets D and D' be given. Then exactly one of the following three statements holds:*

(i): *D and D' are order-isomorphic.*

(ii): *D is order-isomorphic to the ordered subset* $\mathrm{Spr}'(x')$ *of D' for exactly one* $x' \in D'$.

(iii): *D' is order-isomorphic to the ordered subset* $\mathrm{Spr}(x)$ *of D for exactly one* $x \in D$.

Moreover, there is exactly one order-isomorphism in each case.

Proof. The assertion follows from Theorem 176C with the help of Lemma 176B and Remark 176A,(b). ■

The next theorem shows that there are many "different" well-ordered sets. If we knew that every set can be well-ordered, this result would be an almost immediate consequence of Proposition 175B. Our purpose, however, is to use the next theorem to derive the Well-Ordering Principle from assertion (XV).

176E. Theorem (Hartogs's Theorem). *For every collection of well-ordered sets there exists a well-ordered set that is not order-isomorphic to any member of the collection.*

Proof. 1. Since we shall have to deal with many well-ordered sets at once, we shall temporarily have to be especially precise. Thus let $(A;\alpha)$ be a well-ordered set. For every $x \in A$ we indicate by $\mathrm{Spr}_\alpha(x)$ the set consisting of all members of A that strictly precede x in the ordered set $(A;\alpha)$; and for each $x \in A$ we use the abbreviation $\alpha_x := \alpha|_{\mathrm{Spr}_\alpha(x)}$ for the appropriate restriction of the order α.

2. Let the Λ_0 be the given collection of well-ordered sets, and consider the collection

$$\Lambda := \Lambda_0 \cup \{(\mathrm{Spr}_\alpha(x);\alpha_x) \mid (A;\alpha) \in \Lambda_0,\ x \in A\}$$

of well-ordered sets. It is clear that Λ satisfies the following condition:

$$(\mathrm{Spr}_\alpha(x);\alpha_x) \in \Lambda \qquad \text{for all } (A;\alpha) \in \Lambda \text{ and all } x \in A. \tag{176.5}$$

We define the relation σ in Λ by the rule

$$\forall (A;\alpha),(B;\beta) \in \Lambda, \quad (A;\alpha)\ \sigma\ (B;\beta) \ :\Leftrightarrow\ (\text{there is a matching from } (A;\alpha) \text{ to } (B;\beta)). \tag{176.6}$$

It follows form (176.6) and Remark 176A,(d) that σ is reflexive and transitive. From Theorem 176C it follows that σ is total and that

$$\forall (A;\alpha),(B;\beta) \in \Lambda, \quad ((A;\alpha)\ \sigma\ (B;\beta) \text{ and } (B;\beta)\ \sigma\ (A;\alpha)) \ \Leftrightarrow \ \Leftrightarrow ((A;\alpha) \text{ and } (B;\beta) \text{ are order-isomorphic}). \tag{176.7}$$

Let $\mathcal{P}$ be the partition of Λ associated with the equivalence relation "is order-isomorphic to" (Theorem 57C), and define the relation $\prec$ in $\mathcal{P}$ by the rule

$$\forall \Gamma,\Delta \in \mathcal{P}, \quad \Gamma \prec \Delta \ :\Leftrightarrow\ (\exists (A;\alpha) \in \Gamma, \exists (B;\beta) \in \Delta, \quad (A;\alpha)\ \sigma\ (B;\beta)).$$

It follows from (176.7) and Proposition 57H that $\prec$ is a total order in $\mathcal{P}$ and that

$$\forall (A;\alpha),(B;\beta) \in \Lambda, \quad (A;\alpha)\ \sigma\ (B;\beta) \ \Leftrightarrow\ \Omega_{\mathcal{P}}((A;\alpha)) \prec \Omega_{\mathcal{P}}((B;\beta)). \tag{176.8}$$

3. Let $(A;\alpha) \in \Lambda$ be given. It is clear from Remark 176A,(c), Theorem 176C, and (176.5) that

$$\forall x \in A, \quad (\mathrm{Spr}_\alpha(x);\alpha_x)\ \sigma\ (A;\alpha) \tag{176.9}$$

$$\forall x, y \in A, \quad x\ \alpha\ y\ \Leftrightarrow\ (\mathrm{Spr}_\alpha(x);\alpha_x)\ \sigma\ (\mathrm{Spr}_\alpha(y);\alpha_y). \tag{176.10}$$

(Note that if $x \overset{\alpha}{\neq} y$ then $x \in \mathrm{Spr}_\alpha(y)$ and $\mathrm{Spr}_\alpha(x) = \mathrm{Spr}_{\alpha_y}(x)$.)

On account of (176.8) and (176.9) we may define the mapping $\Psi_\alpha : A \to \mathrm{Spr}_\prec(\Omega_\mathcal{P}((A;\alpha)))$ by the rule

$$\Psi_\alpha(x) := \Omega_\mathcal{P}((\mathrm{Spr}_\alpha(x);\alpha_x)) \qquad \text{for all } x \in A.$$

By (176.8) and (176.10) it follows that Ψ_α is strictly α-$\prec$-isotone. We shall now show that Ψ_α is surjective.

Let $\Gamma \in \mathrm{Spr}_\prec(\Omega_\mathcal{P}((A;\alpha)))$ be given, and choose $(B;\beta) \in \Gamma$. Then $\Omega_\mathcal{P}((B;\beta)) = \Gamma \supsetneqq \Omega_\mathcal{P}((A;\alpha))$. By (176.8) we find that $(B;\beta)\ \sigma\ (A;\alpha)$, so that there is a matching from $(B;\beta)$ to $(A;\alpha)$ but $(B;\beta)$ and $(A;\alpha)$ are not order-isomorphic. By Lemma 176B there is exactly one $b \in A$ such that $(B;\beta)$ is order-isomorphic to $(\mathrm{Spr}_\alpha(b);\alpha_b)$. Then $\Gamma = \Omega_\mathcal{P}((B;\beta)) = \Omega_\mathcal{P}((\mathrm{Spr}_\alpha(b);\alpha_b)) = \Psi_\alpha(b) \in \mathrm{Rng}\Psi_\alpha$. Since $\Gamma \in \mathrm{Spr}_\prec(\Omega_\mathcal{P}((A;\alpha)))$ was arbitrary, we conclude that Ψ_α is surjective. Since it is strictly isotone from the well-ordered set $(A;\alpha)$ to $\mathrm{Spr}_\prec(\Omega_\mathcal{P}((A;\alpha)))$ ordered by $\prec$, it is an order-isomorphism; and consequently $\mathrm{Spr}_\prec(\Omega_\mathcal{P}((A;\alpha)))$ is well-ordered by $\prec$.

Since $(A;\alpha) \in \Lambda$ was arbitrary and $\Omega_\mathcal{P} : \Lambda \to \mathcal{P}$ is surjective, we conclude that $\mathrm{Spr}_\prec(\Gamma)$ is well-ordered by $\prec$ for every $\Gamma \in \mathcal{P}$. A simple argument, which we omit, then shows that $\mathcal{P}$ itself is well-ordered by $\prec$.

4. For every $(A;\alpha) \in \Lambda$ we have just shown that there is an order-isomorphism (namely Ψ_α) from $(A;\alpha)$ to $\mathrm{Spr}_\prec(\Omega_\mathcal{P}((A;\alpha)))$ ordered by $\prec$. By Corollary 176D we infer that $(A;\alpha)$ is not order-isomorphic to $(\mathcal{P};\prec)$. Thus $(\mathcal{P};\prec)$ is a well-ordered set that is not order-isomorphic to any member of Λ, let alone to any member of the given collection Λ_0. ■

176F. Corollary. *For every set S there exists a well-ordered set $(W;\omega)$ such that no mapping from S to W is surjective.*

Proof. Let the S be given. Consider the collection Λ of all well-ordered sets whose underlying sets are partitions of S (If $\mathrm{Ord}(A)$ denotes the set of all orders in the set A, and $\mathrm{Part}(S)$ denotes the set of all partitions of S, then Λ may be regarded as a subset of $\bigcup_{\mathcal{P}\in\mathrm{Part}(S)} \mathrm{Ord}(\mathcal{P})$.) By theorem 176E we may choose a well-ordered set $(W;\omega)$ that is not order-isomorphic to any member of Λ.

Now let $f\colon S \to W$ be given, and suppose that f were surjective. By Corollary 36D, there would be a bijection from the partition $\mathrm{Part}f$ of S to the set W. Consequently there would be an order π in $\mathrm{Part}f$ such that the ordered set $(\mathrm{Part}f;\pi)$ is order-isomorphic to $(W;\omega)$, and hence is well-ordered. But then $(\mathrm{Part}f;\pi) \in \Lambda$, and

this would contradict the choice of $(W;\omega)$. Hence the supposition that f is surjective is untenable. ■

We are now ready to formulate the Well-Ordering Principle and to derive it from assertion (XV).

(XVI) (Well-Ordering Principle). *Every set can be well-ordered; more precisely, for every set S there exists an order $\prec$ in S such that $(S;\prec)$ is well-ordered.*

176G. Lemma. *If D is a well-ordered set, then $(A \mapsto \min A) : \mathfrak{P}^{\times}(D) \to D$ is a choice-mapping for the set D.*

176H. Proposition. (XV) *implies* (XVI), *and* (XVI) *implies* (XI).

Proof. (XV) *implies* (XVI). Let the set S be given. By Corollary 176F we may choose a well-ordered set $(W;\omega)$ such that there is no surjection from S to W. If (XV) holds, we may then choose a surjection from W to S. Since there is a choice mapping for W by Lemma 176G, we may apply Lemma 174A and infer that we may choose a right-inverse $f\colon S \to W$ of this surjection; and f is injective. We may therefore define an order $\prec$ in S in such a way that the bijection $f|^{\mathrm{Rng}}$ is an order-isomorphism from S ordered by $\prec$ to $\mathrm{Rng}f$ ordered by ω. Since the latter ordered set is well-ordered, so is the former.

(XVI) *implies* (XI). this is an immediate consequence of Lemma 176G. ■

176I. Remark. The set $\mathbb{N}$ is well-ordered by $\leq$. It follows that every countable set can be well-ordered, and this assertion is valid regardless of the status of the Well-Ordering Principle (XVI). ■

177. Completing the proof of equivalence

This is a purely technical section, in which we shall complete the proof of the fact that the assertions numbered with Roman numerals in this chapter are all equivalent.

177A. Lemma. *Let $\mathcal{D}$ be a collection of sets such that every nest $\mathcal{N}$ included in $\mathcal{D}$ satisfies $\bigcup\mathcal{N} \in \mathcal{D}$. Let $\Phi : \mathcal{D} \to \mathcal{D}$ be a mapping such that, for every $S \in \mathcal{D}$, we have $S \subset \Phi(S)$ and $\Phi(S)\backslash S$ is empty or a singleton. Then there exists $S \in \mathcal{D}$ such that $\Phi(S) = S$.*

Proof. 1. We require a definition. A subcollection $\mathcal{T}$ of $\mathcal{D}$ is called a **tower** if it satisfies the following conditions:

(177.1) every nest $\mathcal{N}$ included in $\mathcal{T}$ satisfies $\bigcup\mathcal{N} \in \mathcal{T}$,

(177.2) $\Phi_>(\mathcal{T}) \subset \mathcal{T}$.

We observe that $\mathcal{D}$ is itself a tower. It follows form this and the definition that the collection of all towers is intersection-stable. Let $\mathcal{S}$ be the intersection of the whole collection, i.e., the smallest of all towers. We intend to show that $\mathcal{S}$ is in fact a nest.

2. Set $\mathcal{C} := \{S \in \mathcal{S} \mid T \subset S \text{ or } S \subset T \text{ for all } T \in \mathcal{S}\}$; and for every $S \in \mathcal{C}$ set $\mathcal{B}(S) := \{T \in \mathcal{S} \mid T \subset S \text{ or } \Phi(S) \subset T\}$.

Let $S \in \mathcal{C}$ be given. We claim that $\mathcal{B}(S)$ is a tower. If $\mathcal{N}$ is a nest and $\mathcal{N} \subset \mathcal{B}(S)$, then either $T \subset S$ for all $T \in \mathcal{N}$, and then $\bigcup\mathcal{N} \subset S$, or there is $T \in \mathcal{N}$ such that $\Phi(S) \subset T$, and hence $\Phi(S) \subset \bigcup\mathcal{N}$. Thus in either case $\bigcup\mathcal{N} \in \mathcal{B}(S)$, and we have shown that (177.1) holds for $\mathcal{T} := \mathcal{B}(S)$. Let $T \in \mathcal{B}(S)$ be given; since $S \in \mathcal{C}$, we have either $\Phi(T) \subset S$ or $S \subsetneqq \Phi(T)$. In the former case, $\Phi(T) \in \mathcal{B}(S)$. In the latter case, the fact that $T \in \mathcal{B}(S)$ implies that either $T \subset S \subsetneqq \Phi(T)$, which requires $T = S$ (since $\Phi(T)\backslash T$ is a singleton) and hence $\Phi(T) = \Phi(S)$; or else $\Phi(S) \subset T \subset \Phi(T)$; so that $\Phi(T) \in \mathcal{B}(S)$ in this case as well. Since $T \in \mathcal{B}(S)$ was arbitrary, (177.2) also holds for $\mathcal{T} := \mathcal{B}(S)$. We have shown that $\mathcal{B}(S)$ is a tower, as claimed; but $\mathcal{B}(S) \subset \mathcal{S}$, and hence $\mathcal{B}(S) = \mathcal{S}$. This conclusion is valid for all $S \in \mathcal{C}$.

We next claim that $\mathcal{C}$ is a tower. (177.1) holds for $\mathcal{T} := \mathcal{C}$, by an argument quite similar to that of the preceding paragraph. If $S \in \mathcal{C}$ and $T \in \mathcal{S} = \mathcal{B}(S)$, we have either $T \subset S \subset \Phi(S)$ or $\Phi(S) \subset T$, so that $\Phi(S) \in \mathcal{C}$. Thus (177.2) also holds for $\mathcal{T} := \mathcal{C}$, and $\mathcal{C}$ is a tower, as claimed. But $\mathcal{C} \subset \mathcal{S}$, and hence $\mathcal{C} = \mathcal{S}$. It follows from the definition of $\mathcal{C}$ that $\mathcal{C} = \mathcal{S}$ is a nest.

3. Since $\mathcal{S}$ is both a nest and a tower, we have $\bigcup\mathcal{S} \in \mathcal{S}$ by (177.1). From (177.2) we obtain $\Phi(\bigcup\mathcal{S}) \in \mathcal{S}$ and therefore $\bigcup\mathcal{S} \subset \Phi(\bigcup\mathcal{S}) \subset \bigcup\mathcal{S}$, so that equality must hold. Thus $S := \bigcup\mathcal{S} \in \mathcal{D}$ satisfies $\Phi(S) = S$. ∎

177B. Lemma. *Let the ordered set D be given, and assume that there is a choice mapping for D. Then every chain of D is included in a maximal chain of D.*

Proof. Choose a choice-mapping γ for the set D. Let C be a given chain of D, and set $\mathcal{D} := \{S \in \mathfrak{P}(D) \mid C \cup S \text{ is a chain of } D\}$. By Lemma 172B, every nest $\mathcal{N}$ included in $\mathcal{D}$ satisfies $\bigcup\mathcal{N} \in \mathcal{D}$.

We define the mapping $\Phi : \mathcal{D} \to \mathcal{D}$ by the rule

$$\Phi(S) := \begin{cases} S \cup \{\gamma(\{x \in D \backslash S \mid C \cup S \cup \{x\} \text{ is a chain of } D\})\} & \text{if } C \cup S \text{ is not a maximal chain of } D \\ S & \text{if } C \cup S \text{ is a maximal chain of } D, \end{cases}$$

for all $S \in \mathcal{D}$. Then Lemma 177A is applicable to $\mathcal{D}$ and Φ, and we conclude that there exists $S \in \mathcal{D}$ such that $\Phi(S) = S$, hence such that $C \cup S$ is a maximal chain that includes C. ■

177C. Theorem. *All the assertions* (I)-(XVI) *are equivalent.*

Proof. By Propositions 172C, 173B, 174B, 174C, 175C, and 176H we have established the implications

$$\text{(IV)} \Rightarrow \text{(III)} \Rightarrow \text{(VI)} \Rightarrow \text{(VII)} \Rightarrow \text{(XIV)} \Rightarrow \text{(XV)} \Rightarrow \text{(XVI)} \Rightarrow \text{(XI)}$$

and the equivalences

$$\text{(I)} \Leftrightarrow \text{(II)} \Leftrightarrow \text{(IV)} \Leftrightarrow \text{(V)} \quad \text{and} \quad \text{(VIII)} \Leftrightarrow \text{(IX)} \Leftrightarrow \text{(X)} \Leftrightarrow \text{(XI)} \Leftrightarrow \text{(XII)} \Leftrightarrow \text{(XIII)}.$$

To complete the proof, we note that the implication (XI) ⇒(IV) is an immediate
▲ consequence of Lemma 177B. ■

INDEXES

Index of terms

Index of names

Index of conditions

Index of symbols

Symbols standing for generic sets, mappings, relations, numbers, etc., are omitted when consistent with intelligibility, and do not affect alphabetic order when present.