ELEMENTARY
SET THEORY

初級集合論

梁鑑添
郭麗珠
著

ELEMENTARY SET THEORY

PARTS I AND II

KAM-TIM LEUNG
AND
DORIS LAI-CHUE CHEN

FOREWORD BY

YUNG-CHOW WONG

Professor of Mathematics, University of Hong Kong

HONG KONG UNIVERSITY PRESS

This combined edition first published 1967
Ninth impression 1992

ISBN 962 209 026 5

Printed in Hong Kong by Condor Production Ltd.

FOREWORD*

The most striking characteristic of modern mathematics is its greater unity and generality. In modern mathematics, the boundaries between different areas have become obscured; very often, what used to be separate and unrelated disciplines are now special cases of a single one; and, amid these far-reaching changes, there have emerged certain basic concepts, notations and terminologies that are of considerable importance and frequent use in a large portion of mathematics.

By 1959, I felt that the time was ripe for this University to introduce into its first-year mathematics course the most fundamental and the more elementary of these basic concepts, notations and terminologies (which we call 'Fundamental Concepts') to serve as a foundation on which much of the undergraduate mathematics rests. The fundamental concepts we first introduced were set theory and some basic operations of algebra. Outline notes for these were written by Dr. S. T. Tsou, who is now with the Chinese University of Hong Kong. These notes were later revised and expanded by Dr. D. Chen to include elements of symbolic logic. In 1961, we further reorganized the contents of the first-year mathematics course by bringing the basic operations of algebra into the more general framework of linear algebra, and Dr. K. T. Leung, who joined this University in 1960, wrote detailed lecture notes for it. Thus, the main topics now taught in our first-year mathematics course are fundamental concepts, linear algebra and calculus. For us, these are new experiments in the teaching of mathematics to undergraduates, but we believe in the correctness of our approach and are greatly heartened by the very encouraging results so far obtained.

In the meantime, cries for the reform of school mathematics began to echo around the world, and there is unanimous agreement that symbolic logic and set theory should be taught as part of the mathematics course in the schools. Thus, in 1962 we decided to include these two topics in the syllabuses of the Ordinary and Advanced Level Pure Mathematics in our Matriculation Examination. Unfortunately, suitable textbooks for teaching these 'sophisticated' topics to VI-form students are lacking, and we have received numerous requests from school teachers and students for advice. To meet this urgent need, Dr. K. T.

* To the 1964 edition of Part I.

Leung and Dr. D. Chen have now expanded the first part of their lecture notes into a book for use by both the undergraduates in the universities and the VI-form students in the schools.

On reading through the manuscript of this book, I find that the material is carefully chosen to suit the purpose, and the ideas are presented in a pleasingly original way. I think in fact that the authors, by using many innovations of their own, have succeeded in writing a book which brings together some of the best features of the French, German, English and Chinese exposition of mathematics. This is a timely textbook which, I am sure, will benefit many students in Hong Kong and elsewhere.

Y. C. WONG

October 1964

University of Hong Kong

PREFACE *

Elementary Set Theory is an extension of the lecture notes for the course 'Fundamental Concepts of Mathematics' given each year to first-year undergraduate students of mathematics in the University of Hong Kong since 1959. The purpose of this course, arranged in about twenty-five lectures, is to provide students of mathematics with the minimum amount of knowledge in logic and set theory needed for a profitable continuation of their studies. The addition of statement calculus and set theory to the Hong Kong University Matriculation Examination Syllabus in 1964 resulted in part of the course being taught in schools. This book, therefore, is intended for schools as well as for universities.

The book begins with a chapter on statement calculus—a brief introduction to the language of logic as used in mathematics. The remaining eight chapters are devoted to set theory. In these eight chapters, the subject is treated in a manner that is intended to be a judicious compromise between the axiomatic and the informal. While *sets* and *belonging* are accepted as undefined primitive concepts, axioms governing these concepts are postulated at different stages in the development of the theory, as circumstances demand. Throughout the book, the everyday language of mathematics is used and meta-mathematical questions are not touched upon. Exercises of varying degrees of difficulty are given at the end of each chapter, the more difficult ones being marked by an asterisk. While there is no need for the reader to work out all the exercises, he should read each one carefully to gain some idea of further possible developments of the theory in question.

The book is divided into two parts, corresponding to the needs of school pupils in Hong Kong and the needs of university students. In writing Part I, which is a revised version of the offset edition published in 1964, we have had in mind the Hong Kong University Advanced Level Examination requirements. Part II, published here for the first time, contains all the material of the course 'Fundamental Concepts of Mathematics' and further reading material for undergraduates.

Professor Y. C. Wong has read our manuscript of Part I and has made valuable suggestions for its improvement. We wish to thank him both

*To the 1967 edition of Parts I and II.

for this and for writing the Foreword to the 1964 edition of Part I. We are also grateful to our colleagues Mr. Y. M. Wong and Mr. Y. H. Au-Yeung who have given us valuable assistance. Finally, we thank Mr. K. W. Ho for the typing of the manuscript.

K. T. LEUNG
DORIS L. C. CHEN

March 1966
University of Hong Kong

CONTENTS

PART I

CHAPTER 1

STATEMENT CALCULUS

A. Statements

By a *statement* (or a *proposition* or a *declarative sentence*) we understand a sentence of which it is meaningful to say that its content is true or false. Obviously, each of the following sentences is a statement:

Geography is a science.
Confucius was a soldier.
Cheung Sam is dead and Lee Sai is in prison.
2 is smaller than 3 and 3 is a prime number.
The steering gear was loose or the driver was drunk.
If John is here, then the book is not his.

Whereas none of the following sentences can be regarded as a statement in the above sense:

The number 3 is stupid.
Friends, Romans, countrymen, lend me your ears.
Cousin of Exeter, what thinks your lordship?

Throughout this chapter, we shall mainly be concerned with statements. Here we shall briefly describe what we propose to do with them. In the *statement calculus* (or *propositional calculus*) of this chapter, with the exception of Sections **K** and **L**, we shall not concern ourselves with the relation between the subjects and the predicates of the statements. Instead we shall deal with the statements themselves as entireties, and study the modes of compounding them into further statements. Examples might explain this more clearly.

Consider the statement

(1) Geography is a science.

Being a statement, (1) is either true or false. If it is true, then we say that (1) has *truth* as its *truth value*; if it is false, we say that (1) has *falsehood* as its truth value. Now this truth value is a relation between the subject and the predicate of the sentence (1), and opinion concerning the truth of this sentence may be divided. For us the sentence 'Geography

is a science' is merely a statement to which either of the truth values can be meaningfully assigned.

The statements

(2) 2 is smaller than 3

(3) 3 is a prime number

are true statements of arithmetic. By joining (2) and (3) we can form a new statement

(4) 2 is smaller than 3 *and* 3 is a prime number.

Our main concern about (4) is its truth value in relation to the truth values of (2) and (3). As a statement of arithmetic, (4) is true since (2) and (3) are both true statements of arithmetic.

This and similar problems will be discussed more thoroughly in the next few sections.

B. Conjunctions

In ordinary speech, we frequently join two statements by the word *and*. Let us consider the statement

(1) Cheung Sam is dead *and* Lee Sai is in prison.

We say that (1) has as its *first component* the statement 'Cheung Sam is dead' and as its *second component* the statement 'Lee Sai is in prison', and moreover that (1) is formed by joining these two components by the connective *and*. Ordinarily, a statement such as (1) is accepted as true if both of its components are true; otherwise it is considered false. Corresponding to this mode of composition, which consists in joining two statements by the connective *and*, we have the concept of *conjunction* in statement calculus.

In the notation of mathematical logic, the conjunction of two statements X and Y is denoted by $X \wedge Y$, and read 'X and Y'. $X \wedge Y$ is a statement which is true if both X and Y are true; otherwise it is false. Thus the truth value of a conjunction $X \wedge Y$ is uniquely determined by the truth values of its components X and Y. This can be conveniently expressed in a *truth table*:

1st Component X	2nd Component Y	Conjunction $X \wedge Y$
T	T	T
F	T	F
T	F	F
F	F	F

where 'T' stands for truth and 'F' stands for falsehood.

C. Disjunctions

A second and equally familiar mode of composition consists in joining two statements, the components, by the connective *or*. Unfortunately, the meaning of the word *or* in English is ambiguous; sometimes it is used in the inclusive sense, meaning either one or the other or both; sometimes it is used in the exclusive sense, meaning either one or the other, but not both. Common usage probably favours the inclusive *or*. For example, in a motor-car manual, we may find the following warning:

> You will damage the engine if you run it
> when it is too low on oil *or* water.

Clearly, the driver is advised to avoid running the engine not only when oil is too low *or* when water is too low, but also when ***both*** oil ***and*** water are too low; it is therefore an example of the inclusive *or*. On the other hand, expressions like *or both* and *and/or*, especially in legal documents, suggest that *or* alone should be interpreted as exclusive. The different meanings of *or* can be expressed in the form of a truth table, as follows:

1st Component X	2nd Component Y	X or Y	
		inclusive	exclusive
T	T	T	F
F	T	T	T
T	F	T	T
F	F	F	F

In mathematical logic, the ***disjunction*** is the composition which joins two statements, the components, by the symbol $\vee$ corresponding to *or* in the inclusive sense. For any two statements X and Y, the disjunction of X and Y is written as $X \vee Y$ and read 'X *or* Y'. Thus the truth table of the disjunction is as follows:

1st Component X	2nd Component Y	Disjunction $X \vee Y$
T	T	T
F	T	T
T	F	T
F	F	F

The exclusive use of *or*, which is less important in mathematical logic, does not call for a special name and symbol. In fact, this can be

expressed as an iterated composition, which will be introduced in a later section.

REMARKS. The English *or*, the French *ou* and the German *oder* all admit the two interpretations in common usage. In Latin, the words *aut* and *vel* correspond to the exclusive and the inclusive *or* respectively. This partly explains the choice of the symbol $\vee$ for the disjunction.

D. Negations

We say that the disjunction and the conjunction are *binary* compositions because each of them combines *two* statements into a new statement. On the other hand, the *negation* (or *denial*) is called a *singulary* (or *unary*) composition because it just transforms a statement into a new one. The usual way to express a negation is to insert the word *not* into a sentence or to withdraw it from a sentence as the case may be. For example, the negation of the statement 'Confucius was a soldier' is the statement 'Confucius was *not* a soldier', and that of 'Wang is *not* here' is 'Wang is here'. However there are also other ways of doing it; for example:

'*Sometimes* it rains in Hong Kong' is transformed into 'It *never* rains in Hong Kong';

'*All* cats are vicious' is transformed into '*Some* cats are *not* vicious';

'*Every* dog is *not* obedient' is transformed into '*Some* dogs are obedient', and

'Cheung Sam *and* Lee Sai make a great deal of noise' is transformed into 'Cheung Sam *or* Lee Sai (i.e. at least one of them) does *not* make a great deal of noise'.

These examples also show that the truth value of the negation of a statement is just the opposite of that of the statement.

In mathematical logic, the negation of a statement is formed by prefixing to the statement the tilde $\sim$ which is conveniently read as *not*. Thus the negation of 'Confucius was a soldier' is written as '$\sim$ Confucius was a soldier' and can be read as 'Not, Confucius was a soldier' without disturbing the structure of the original statement. The truth table of the negation is as follows:

Component X	Negation $\sim X$
T	F
F	T

E. Conditionals

The '*if—then*' combination in ordinary speech, such as in

(1) *If* Chan was here, *then* the knife is not his

provides us with another important binary composition called the *conditional.* A conditional has two components: the first is called the *hypothesis* and the second is called the *consequent* of the conditional. 'Chan was here' and 'the knife is not his' are then the hypothesis and the consequent of the conditional (1).

Let us now investigate the truth value of a conditional. It is generally accepted in everyday life that a conditional with a true hypothesis is true (respectively false) if its consequent is true (respectively false). This is easily seen from the following examples:

(2) If $2 > 1$, then $3 > 2$

(3) If $2 > 1$, then $3 < 2$

(2) is a true statement, and (3) is a false statement. Likewise, a conditional with false hypothesis and false consequent is regarded as true. For example, in

(4) If $2 < 1$, then $3 < 2$

we have a true statement which is a conditional with a false hypothesis and a false consequent.

A conditional with a false hypothesis and a true consequent can also be regarded as true, although it is often difficult to convince some people, especially non-mathematicians, of this. Let us consider an example. Cheung Sam, seeing a car zigzagging along, came to the conclusion that

(5) If the steering gear was not loose, then the driver was drunk.

It was found later that both the steering gear was loose and the driver was drunk. However Cheung Sam's conclusion was still correct. In fact, he might have equally well re-phrased his conclusion as

(6) The steering gear was loose or the driver was drunk.

Consequently, in the case of both the steering gear being loose and the driver being drunk, the disjunction (6) is true, and hence the conditional (5) is also true.

Thus a conditional is false if the hypothesis is true and the consequent is false, but it is true in all other cases. In mathematical logic, the conditional is formed by inserting an arrow $\rightarrow$ between the components

in the order: hypothesis → consequent. The conditional (1) can be written as

(Chan was here) → (The knife is not his)

and read '*If* Chan was here, *then* the knife is not his' or 'Chan was here *only if* the knife is not his'. Finally, the truth table of the conditional is given by

Hypothesis X	Consequent Y	Conditional $X \to Y$
T	T	T
F	T	T
T	F	F
F	F	T

F. Truth functional compositions

Having introduced the four fundamental modes of composition, we now deal with a simple but important property of theirs, which will lead us further on in our investigation.

All four compositions have the property:

(TF) *The truth value of the compound of the composition is determined in all cases by the truth values of the components.*

Because of this, we define formally: a mode of composition is said to be *truth functional* if it has the property (TF). Thus each truth functional composition can be described by a truth table. The conjunction, the disjunction, the negation, and the conditional are all truth functional compositions. Statement calculus deals only with truth functional compositions.

G. Iterated compositions

We are now in a position to study more complicated statements. The statement

(1) If it rains, then John will drive Jack home or Jack will stay with John.

has three components, namely, 'it rains', 'John will drive Jack home' and 'Jack will stay with John'. Let us denote these three statements by X, Y and Z respectively. To obtain (1), we first form the disjunction $Y \vee Z$ and then the conditional $X \to (Y \vee Z)$. This is an example of an iterated composition.

In general, an *iterated composition* is a mode of composition which consists in applying truth functional compositions a number of times. Clearly, an iterated composition is also a truth functional composition. Thus the truth value of (1) can also be given in the form of a truth table.

X	Y	Z	$Y \vee Z$	$X \rightarrow (Y \vee Z)$
T	T	T	T	T
F	T	T	T	T
T	F	T	T	T
T	T	F	T	T
F	F	T	T	T
T	F	F	F	F
F	T	F	T	T
F	F	F	F	T

To set up this truth table, we write all possible combinations of truth values of the components X, Y and Z into the first three columns; then we calculate the truth values for the fourth column from those of the second and third columns, and the truth values for the fifth column from those of the first and fourth columns. Then the truth values of statement (1) are given in the fifth column.

In order to avoid unnecessary repetition, especially in more complicated cases, we can set up the truth table in the following form:

$$X \rightarrow (Y \vee Z)$$

T T T T T
F T T T T
T T F T T
T T T T F
F T F T T
T F F F F
F T T T F
F T F F F

The following is the standard procedure:

(i) Fill in the columns below the components; here, these are the first, third and fifth columns. If a component appears more than once, then in each row the same value (T or F) should be entered for this component.

(ii) Calculate row by row the truth value of the first composition and put it into the column under its representative logical symbol; here, this is the fourth column.

(iii) Calculate row by row the truth value of the next composition and put it into the column under its representative logical symbol, and so forth; here, this is the second column, which gives the truth value of the statement (1).

As a second example of iterated composition, we study a statement of the form

(2) $(X \to Y) \wedge (Y \to X)$

where X, Y denote any two arbitrary statements. The truth table of (2) is easily calculated:

$$
\begin{array}{cccccccc}
(X & \to & Y) & \wedge & (Y & \to & X) \\
\mathrm{T} & \mathrm{T} & \mathrm{T} & \mathrm{T} & \mathrm{T} & \mathrm{T} & \mathrm{T} \\
\mathrm{F} & \mathrm{T} & \mathrm{T} & \mathrm{F} & \mathrm{T} & \mathrm{F} & \mathrm{F} \\
\mathrm{T} & \mathrm{F} & \mathrm{F} & \mathrm{F} & \mathrm{F} & \mathrm{T} & \mathrm{T} \\
\mathrm{F} & \mathrm{T} & \mathrm{F} & \mathrm{T} & \mathrm{F} & \mathrm{T} & \mathrm{F}
\end{array}
$$

This iterated composition occurs very frequently in mathematics, and we find it convenient to call it the ***biconditional*** and use the symbol $\leftrightarrow$ to represent it. Thus the biconditional (2) is abbreviated to

(3) $X \leftrightarrow Y$

and read '*X if and only if Y*'. (The American mathematician P. R. HALMOS has introduced the abbreviation *iff* for 'if and only if' which is now widely accepted.) The truth table of (3) takes a simple form

$$
\begin{array}{ccc}
X & \leftrightarrow & Y \\
\mathrm{T} & \mathrm{T} & \mathrm{T} \\
\mathrm{F} & \mathrm{F} & \mathrm{T} \\
\mathrm{T} & \mathrm{F} & \mathrm{F} \\
\mathrm{F} & \mathrm{T} & \mathrm{F}
\end{array}
$$

Thus the biconditional is true if both its components have the same truth value; otherwise it is false.

H. Equivalent formulae

Without doubt the main thing that distinguishes secondary school algebra from primary school arithmetic is its use of letters to represent

numerals. We have now reached the point where we can treat the statements in the same way as numerals are treated in algebra. In this section, we shall represent statements by capital letters X, Y, Z, . . . , and their truth functional compositions by *statement formulae* such as $X \to Y$ and $(\sim Y) \leftrightarrow ((\sim X) \wedge Z)$. Therefore the compositions $\wedge$, $\vee$, $\to$, $\sim$, play a role in statement calculus similar to that played by $+$, $-$, $\times$, in algebra. Corresponding to the equality $(=)$ in algebra there is the concept of *equivalence* which we shall introduce now.

A statement formula, which represents a truth functional composition, can also be regarded as a *truth function*. For example, the formula

(1) $\quad X \to Y$

is a truth function in the indeterminates X and Y. As a truth function, (1) has the value F if the indeterminates X, Y take the values T and F respectively, and has the value T otherwise.

We say that two statement formulae are *equivalent* if, when regarded as truth functions, they are identical. For example, the formula (1) and the formula

(2) $\quad (\sim X) \vee Y$

are equivalent. To show this, we have only to compare their truth tables

$X \to Y$			$(\sim X) \vee Y$			
T	T	T	F	T	T	T
F	T	T	T	F	T	T
T	F	F	F	T	F	F
F	T	F	T	F	T	F

From these tables, we see clearly that the functions (1) and (2) yield identical values on each substitution of values of the indeterminates, and this means that the two functions are identical. Therefore statement formulae (1) and (2) are equivalent, and we write

(3) $\quad X \to Y \text{ eq } (\sim X) \vee Y.$

The following equivalences will be very useful later on:

(4) $\quad \sim(\sim(X)) \text{ eq } X$

(5) $\quad X \wedge Y \text{ eq } Y \wedge X$ *(the commutative law of conjunction)*

(6) $\quad X \wedge (Y \wedge Z) \text{ eq } (X \wedge Y) \wedge Z$ *(the associative law of conjunction)*

(7) $\quad X \vee X \text{ eq } X$

(8) $\quad X \vee Y \text{ eq } Y \vee X$ *(the commutative law of disjunction)*

(9) $\quad X \vee (Y \vee Z) \text{ eq } (X \vee Y) \vee Z$ *(the associative law of disjunction)*

(10) $X \vee (Y \wedge Z)$ eq $(X \vee Y) \wedge (X \vee Z)$
(11) $X \wedge (Y \vee Z)$ eq $(X \wedge Y) \vee (X \wedge Z)$ } *(the distributive laws)*

(12) $\sim(X \wedge Y)$ eq $(\sim X) \vee (\sim Y)$
(13) $\sim(X \vee Y)$ eq $(\sim X) \wedge (\sim Y)$ } *(De Morgan's laws)*

(14) $X \to Y$ eq $(\sim Y) \to (\sim X)$
(15) $X \leftrightarrow Y$ eq $Y \leftrightarrow X$
(16) $X \leftrightarrow Y$ eq $(\sim X) \leftrightarrow (\sim Y)$.

Each of these equivalences can be verified by setting up the truth tables of the formulae on both sides of it. In some cases, this method may involve long computations, then we use some other more convenient methods. As an illustration, we prove (14) as follows:

$X \to Y$ eq $(\sim X) \vee Y$ [by (3)]
$(\sim X) \vee Y$ eq $Y \vee (\sim X)$ [by (8)]
$Y \vee (\sim X)$ eq $(\sim(\sim Y)) \vee (\sim X)$ [by (4)]
$(\sim(\sim Y)) \vee (\sim X)$ eq $(\sim Y) \to (\sim X)$ [by (3)]

Therefore, $X \to Y$ eq $(\sim Y) \to (\sim X)$.

The pattern of this proof is similar to that of proving an algebraic identity. In fact, 'eq' is treated here in the same way as '=' is in algebra. The conclusion is preceded by a sequence of equivalences. The left-hand side of (14) appears as the left-hand side of the first equivalence, and the right-hand side of (14) appears as the right-hand side of the last equivalence, whereas the right-hand side of each equivalence in the sequence appears as the left-hand side of the next equivalence. Therefore, when regarded as truth functions, all members of the equivalences of the sequence are identical; in particular, the left-hand side of the first equivalence and the right-hand side of the last equivalence are identical truth functions, proving the equivalence (14). To avoid repetition, we may also write the sequence of equivalences as

$X \to Y$ eq $(\sim X) \vee Y$ eq $Y \vee (\sim X)$
eq $(\sim(\sim Y)) \vee (\sim X)$ eq $(\sim Y) \to (\sim X)$

I. Valid formulae

The observation in the last paragraph of the preceding section provides a method to simplify statement formulae. Here we shall give another method which corresponds to the cancellation laws of algebra. We say that a statement formula is a *valid formula* if it represents a

constant truth function having T as its value. The following are all valid formulae:

(1) $(\sim X) \vee X$
(2) $\sim(X \wedge (\sim X))$
(3) $(X \wedge Y) \rightarrow X$
(4) $X \rightarrow (X \vee Y)$
(5) $((X \rightarrow Y) \wedge (Y \rightarrow Z)) \rightarrow (X \rightarrow Z)$
(6) $(X \wedge Y) \rightarrow (X \rightarrow Y)$

A further collection of valid formulae can be obtained if we replace 'eq' by '$\leftrightarrow$' (or '$\rightarrow$') in (3)–(18) of the last section; for example,

(7) $(X \rightarrow Y) \leftrightarrow (\sim X) \vee Y$
(8) $(\sim(\sim X)) \rightarrow X.$

When equivalences of statement formulae are our main concern, we shall denote valid formulae by the bold-faced capital letter **T**, since all the valid formulae are equivalent to each other. The negation of a valid formula is a formula which represents a constant truth function taking F as its value. We shall denote such formulae by **F**. The following equivalences provide some rules of cancellation:

(9) $X \wedge \mathbf{T}$ eq X
(10) $X \wedge \mathbf{F}$ eq $\mathbf{F}$
(11) $X \vee \mathbf{T}$ eq $\mathbf{T}$
(12) $X \vee \mathbf{F}$ eq X
(13) $\mathbf{T} \rightarrow X$ eq X
(14) $X \rightarrow \mathbf{T}$ eq $\mathbf{T}$
(15) $X \leftrightarrow \mathbf{F}$ eq $\sim X$
(16) $\sim \mathbf{T}$ eq $\mathbf{F}$
(17) $\sim \mathbf{F}$ eq $\mathbf{T}$

These results can be used to simplify statement formulae. Consider, for example, the statement formula

$$X \wedge \{[Z \rightarrow (X \vee Y)] \leftrightarrow [(Z \rightarrow X) \vee (Z \rightarrow Y)]\}$$

For the expression within the braces, we get

$$\begin{aligned} &[Z \rightarrow (X \vee Y)] \leftrightarrow [(Z \rightarrow X) \vee (Z \rightarrow Y)] \\ \text{eq } &[(\sim Z) \vee (X \vee Y)] \leftrightarrow [((\sim Z) \vee X) \vee ((\sim Z) \vee Y)] \\ \text{eq } &[(\sim Z) \vee X \vee Y] \leftrightarrow [(\sim Z) \vee X \vee (\sim Z) \vee Y] \\ \text{eq } &[(\sim Z) \vee X \vee Y] \leftrightarrow [(\sim Z) \vee X \vee Y]. \end{aligned}$$

The last formula is obviously a valid formula; therefore
$[Z \to (X \vee Y)] \leftrightarrow [(Z \to X) \vee (Z \to Y)]$ is a valid formula. Hence

$$X \wedge \{[Z \to (X \vee Y)] \leftrightarrow [(Z \to X) \vee (Z \to Y)]\} \text{ eq } X \wedge \mathbf{T} \text{ eq } X.$$

J. Names of objects and names of names

In this section, we want to clarify the uses of names in daily discourse and introduce an appropriate notation to deal with them. The title of this section may suggest a hair-splitting pedantry, but before we form such a judgement, let us consider the following two sentences:

(1) Hong Kong is populous.

(2) Hong Kong is disyllabic.

Sentence (1) is intended to mean that the city in question, which has a population of nearly four million, is populous, while sentence (2) is intended to mean that the name of the city in question has two syllables. These two sentences have therefore different subjects, a city in (1) and a name in (2), though these two different subjects are written in the same way. Obviously this situation is unsatisfactory, and therefore it is desirable to have some way of distinguishing them.

In (1), we have a statement about a city, and since it is physically impossible to put a city into a sentence, we are compelled to use its name to represent it there. In fact, it is general practice *to use a name of an object to talk about that object.* Accordingly, if the object in question is itself a name, then, *to talk about this name, a name of this name is used.* Employing inverted commas to indicate a name of a name of an object, we rewrite (2) in the correct form

(3) 'Hong Kong' is disyllabic.

Now the subject of sentence (3) is a name of the subject of sentence (1).

To summarize: in (1), the name of a city is used and thus a city is mentioned; in (3), the name of a name of a city is used and thus a name of a city is mentioned. We may also say that

'Hong Kong' is used in (1) to mention Hong Kong, and
' 'Hong Kong' ' is used in (3) to mention 'Hong Kong'.

K. Implications

In the previous sections, we have given a rudimentary theory of statement calculus. We shall now study some of its applications to mathematics. Mathematics has a vocabulary of its own, which we must

learn. First of all, the words ***statements*** and ***compositions of statements*** in mathematics must be understood to have the same meaning as defined in statement calculus. The next words we shall study in this vocabulary are the verbs ***imply*** and ***follow***.

In saying

(1) Victoria Peak is 1809 feet above sea level,

we attribute the property of being 1809 feet above sea level to Victoria Peak. Analogously, we can attribute the property of being true to statements. For example, we may attribute this property to (1) and say

(2) 'Victoria Peak is 1809 feet above sea level' is true.

More generally, when 'X' is a statement, in saying that

(3) 'X' is true,

we mean that 'X' has the truth value T; in other words, we mean that the content of 'X' is true.

An ***implication*** consists in attributing the property of being true to a conditional; thus for the conditional

(4) $X \to Y$

we have the implication

(5) '$X \to Y$' is true.

As a truth functional composition, (4) has the truth value T or F, depending on the truth values of its components, and this is shown in this truth table:

X	Y	$X \to Y$
T	T	T
F	T	T
T	F	F
F	F	T

However, the implication (5) excludes the case where the truth value of (4) is F. In conformity with the notation and terminology used in mathematics, we write (5) into any of the following forms:

(6) 'X' implies 'Y'.

(7) 'Y' follows from 'X'.

(8) 'X' $\Rightarrow$ 'Y'.

We notice that the symbol $\Rightarrow$ does not represent a composition of statements, since, by definition, it represents the attribution of T to the composition (4).

If 'X' $\Rightarrow$ 'Y', then by definition we have the following mutually exclusive cases: (i) 'X' is true and 'Y' is true, (ii) 'X' is false and 'Y' is true, (iii) 'X' is false and 'Y' is false. Provided that, in addition, 'X' is true, then the cases (ii) and (iii) must be excluded, and we conclude that 'Y' is true. Therefore we have the following *law of inference*:

(9) If 'X' $\Rightarrow$ 'Y' and 'X' is true, then 'Y' is true.

This is known as the *modus ponens*. Other laws of inference can be proved in a similar way; for example:

(10) If 'X' $\Rightarrow$ 'Y' and 'Y' $\Rightarrow$ 'Z', then 'X' $\Rightarrow$ 'Z'.
(hypothetical syllogism)

(11) If 'X' $\Rightarrow$ 'Y', then '$\sim Y$' $\Rightarrow$ '$\sim X$'.
(contrapositive inference or *modus tollens)*

Similarly, from a biconditional

(12) $X \leftrightarrow Y$

we have

(13) '$X \leftrightarrow Y$' is true

which may be written in either of the forms:

(14) 'X' is *equivalent* to 'Y'

(15) 'X' $\Leftrightarrow$ 'Y'.

Here again the symbol $\Leftrightarrow$ does not represent a composition of statements.

In mathematics, *theorems*, *lemmas*, *propositions* or *corollaries* assert the truth of statements. Strictly speaking, also in terms of grammatical structure, a theorem should have the name of a statement as its subject and 'is true', 'holds', or something of the kind, as its predicate. But in general practice, the predicate is usually not written out, so that a theorem generally consists of just the statement whose truth it asserts. For example,

Theorem. If $a \perp c$ and $b \perp c$ then $a \parallel b$

should be understood as

'If $a \perp c$ and $b \perp c$ then $a \parallel b$' is true.

Strictly speaking, the verbs *imply* and *follow* should be used as in (6) and (7); however, the invarted commas are usually omitted. Thus instead of writing correctly

'2 is greater than 1' implies '3 is greater than 2'

and

$$'X' \Rightarrow 'Y',$$

we usually write (incorrectly!)

2 is greater than 1 implies that 3 is greater than 2

and

$$X \Rightarrow Y$$

respectively. Finally, in a proof we write (incorrectly)

$$A \Rightarrow B \Rightarrow C \Rightarrow \ldots \Rightarrow X \Rightarrow Y$$

we mean that

$$'A' \Rightarrow 'B',\ 'B' \Rightarrow 'C',\ \ldots,\ \text{and}\ 'X' \Rightarrow 'Y'.$$

From now on, we shall drop the inverted commas when no confusion is possible.

L. The symbols ∀ and ∃

In mathematics, we are familiar with the process of classifying objects according to some properties they share. For instance, in the arithmetic of integers, we classify integers which are divisible by 2 as even integers. Let us denote by $\mathscr{E}$ the class of all even integers and by $\mathscr{F}$ the class of all integers divisible by 4, and study some relations between $\mathscr{E}$ and $\mathscr{F}$, for example,

(1) All integers of $\mathscr{F}$ are integers of $\mathscr{E}$

(2) No integers of $\mathscr{F}$ are not integers of $\mathscr{E}$

(3) Not all integers of $\mathscr{E}$ are not integers of $\mathscr{F}$

(4) Some integers of $\mathscr{E}$ are integers of $\mathscr{F}$

(5) Not all integers of $\mathscr{E}$ are integers of $\mathscr{F}$

(6) Some integers of $\mathscr{E}$ are not integers of $\mathscr{F}$.

In order to handle this type of sentence efficiently, we must introduce the symbols ∀ and ∃; these are, respectively, the *universal quantifier* and the *existential quantifier* of the *predicate calculus* of mathematical logic.

Let us represent by E the property of being an even integer, and write $E(x)$ to represent the sentence

(7) x is an even integer.

This sentence is not a statement in the sense laid down in Section **A**, since the subject x is unspecified. But in substituting for x any definite object, say the number 4, we have $E(4)$ or

(8) 4 is an even integer.

Then $E(4)$, i.e. (8), is a statement. Thus we say that $E(x)$ is a *statement function*. Similarly, if F is the property of being an integer divisible by 4, we define a statement function $F(x)$ as

(9) x is an integer divisible by 4.

Consider now the sentences:

Everything is an even integer.

Every individual thing in the universe is an even integer.

For every x, x is an even integer.

For all x, x is an even integer.

All these sentences have the same meaning. In the predicate calculus they are represented by the expression

(10) $\forall(x)\,[E(x)]$.

In general, if $H(x)$ is a statement function representing

x has the property H,

then the sentences

Everything has the property H, and

For every x, x has the property H,

are represented by the expression $\forall(x)\,[H(x)]$. In this way, statements (1), (3) and (5) are represented respectively by

(1′) $\forall(x)\,[F(x) \rightarrow E(x)]$

(3′) $\sim\forall(x)\,[E(x) \rightarrow \sim F(x)]$

(5′) $\sim\forall(x)\,[E(x) \rightarrow F(x)]$

Consider now another group of sentences:

Something is an even integer.

There exists an even integer.

There is at least one x such that x is an even integer.

There is an x such that x is an even integer.

Again, all these sentences have the same meaning. In the predicate calculus they are represented by the expression

(11) $\exists(x)\,[E(x)]$

Using the symbol $\exists$ we can write statements (2), (4) and (6) respectively as

(2′) $\sim\exists(x)\,[F(x) \wedge \sim E(x)]$

(4′) $\exists(x)\,[E(x) \wedge F(x)]$

(6′) $\exists(x)\,[E(x) \wedge \sim F(x)]$

Before we conclude this section, let us consider the statement function $E(x)$: x is an even integer. In applying the negation, we have

$\sim E(x)$: x is not an even integer.

Clearly the sentences

Not everything is an even integer, and

Something is not an even integer,

have the same meaning and the sentences

Everything is not an even integer, and

Nothing an even integer,

also have the same meaning. Therefore we have

$$\sim \forall(x)[E(x)] \text{ eq } \exists(x)[\sim E(x)]$$

and

$$\forall(x)[\sim E(x)] \text{ eq } \sim \exists(x)[E(x)]$$

In general, for any statement function $H(x)$, the following rule is valid

(12) $$\sim \forall(x)[H(x)] \text{ eq } \exists(x)[\sim H(x)]$$

From this we see that (1′) eq (2′), (3′) eq (4′) and (5′) eq (6′). Therefore (1) eq (2), (3) eq (4) and (5) eq (6).

M. Exercises

1. Set up the truth tables of the following iterated compositions:
   (i) $(X \to Y) \vee (X \to Z)$
   (ii) $(\sim(X \wedge Y)) \to Z$
   (iii) $(X \leftrightarrow Y) \wedge (Z \wedge X)$
2. Turn the statement 'either X or Y' into an iterated composition.
3. Find the iterated compositions of negations and disjunctions that are equivalent to $X \wedge Y$, $X \to Y$, and $X \leftrightarrow Y$.
4. Use the results of Exercise 3 to simplify the following formulae:
   (i) $[X \vee [(Y \to Z) \leftrightarrow X]] \to (Y \to X)$
   (ii) $X \leftrightarrow \{Y \vee [(\sim X) \to Z]\}$
5. Corresponding to the statement 'neither X nor Y', the *joint negation* $X \downarrow Y$ is defined by the truth table

X	Y	$X \downarrow Y$
T	T	F
F	T	F
T	F	F
F	F	T

Show that (i) $X \downarrow Y$ eq $\sim(X \vee Y)$ and (ii) X eq $(X \downarrow X) \downarrow (X \downarrow X)$.

6. Show that (i) $\sim X$ eq $X \downarrow X$ and (ii) $X \vee Y$ eq $(X \downarrow Y) \downarrow (X \downarrow Y)$. From these results find the equivalent formulae of $X \wedge Y$, $X \to Y$ and $X \leftrightarrow Y$ as iterated compositions of joint negations.
7. Prove the equivalences (4)–(16) of Section **H**.
8. Prove that (1)–(8) of Section **I** are valid formulae.
9. Prove the laws of inference (9), (10) and (11) of Section **K**.
10. Prove that every truth function of the indeterminates X and Y is an iterated composition of negations and disjunctions. A hint: of the sixteen distinct truth functions in the indeterminates X and Y, eight are given below; the rest are their negations.

X	Y	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
T	T	T	T	T	T	F	T	T	T
F	T	T	T	T	F	T	T	F	F
T	F	T	T	F	T	T	F	F	T
F	F	T	F	T	T	T	F	T	F

Now f_1 eq $(\sim X) \vee X$, f_2 eq $X \vee Y$.

11. (a) How many distinct truth functions in the indeterminates $X_1, X_2, \ldots, X_n$ exist?

    (b) If g_1 and g_2 are truth functions in the indeterminates $X_1, X_2, \ldots, X_{n-1}$, show that the truth function

    $$(X_n \wedge g_1) \vee \{(\sim X_n) \wedge g_2\}$$

    has the same value as g_1 when X_n is true and the same value as g_2 when X_n is false.

    (c) Prove that every truth function in the indeterminates $X_1, X_2, \ldots, X_n$ is an iterated composition of negations and disjunctions.

12. Let $H(x)$ be a statement function and S a statement in which x does not appear. Show that

    $$\text{`}\forall(x)[S \vee H(x)]\text{'} \Leftrightarrow \text{`}S \vee \{\forall(x)[H(x)]\}\text{'}$$
    $$\text{`}\exists(x)[S \wedge H(x)]\text{'} \Leftrightarrow \text{`}S \wedge \{\exists(x)[H(x)]\}\text{'}.$$

CHAPTER 2

SETS

A. Sets

A fundamental concept in mathematics is that of a set. This concept can be used as a foundation of all known mathematics. In this and the following chapters, we shall develop some of the basic properties of sets. In set theory, we shall be dealing with sets of objects. Here we take *objects* to be simply the individual things of our intuition and our thoughts. In what follows, objects are referred to by their names, usually letters. Thus in saying 'an object is denoted by x' or 'x is an object', we mean that 'x' is a name of the object in question.

We also use *equality* between objects in its common-sense meaning. Equality between two objects x and y is denoted by $x = y$; it means that 'x' and 'y' are different names of the same object. The negation of $x = y$ is written as $x \neq y$. We emphasize that for any objects x and y, either $x = y$ or $x \neq y$.

Having introduced the words *object* and *equality* into our vocabulary, we shall now study the word *set*. One may try to define the notion of a set in terms of an act of collecting many objects together into a whole. However it has been found that any such definition is not only unsatisfactory in itself, but also leads to certain logical difficulties. For this reason, we take the axiomatic approach to elementary geometry, which has proved very successful, as the example to be followed. In elementary geometry, concepts such as points and lines are undefined; let us therefore *accept the notion of a set as undefined.* Thus no definition of sets will be given and we are only interested in what we can do with them. Though the concept of sets is undefined, the sets themselves must be taken as objects in the sense described above.

What is the analogous situation in elementary geometry? In the commonly used text-books, sentences such as 'a point has position, but is said to have no magnitude' or 'a line has length, but not breadth' can only serve as intuitive descriptions of a point and a line, not as exact definitions of these objects, unless more information about the concepts of position, magnitude and breadth in terms of which they might be defined, is available. Therefore, we may either (i) attempt to define the concepts of position, magnitude and breadth, or (ii) accept

the concepts of position, magnitude and breadth as undefined primitive concepts and define line and point in terms of these undefined concepts, or (iii) accept the concepts of point and line as undefined primitive concepts. The first alternative is unsatisfactory, for in defining these concepts, we cannot avoid using other concepts which have yet to be defined. Of the second and the third alternatives, the third is more profitable, because in geometry points and lines are the objects we work with.

Thus, in the axiomatic approach to geometry, we leave the points and the lines as undefined objects susceptible to a number of *relations* among themselves, such as *incidence* and *betweenness*. These relations are also undefined but are subjected to a number of conditions, called *axioms* or *postulates*. The axioms are accepted as true statements from which the theorems of elementary geometry can be deduced by purely logical argument. Take for example the incidence relation between points and lines. For '*a point P is incident to a line g*', we also say '*P lies on g*', or '*g passes through P*', or '*g contains P*' etc. The first four axioms of incidence are as follows:

1. Through two points, there passes a line.
2. Through two distinct points, there passes at most one line.
3. On a line there lie at least two points.
4. There are at least three points which do not lie on any one line.

From these four axioms, we can immediately deduce the theorem:

> Two lines are equal if and only if there are two distinct points that are incident to each of them.

Relating sets to objects, there is the concept of *belonging*, also undefined. This is like relating points and lines by the undefined concept of incidence used in elementary geometry. If an object x belongs to a set A, we write

$$x \in A.$$

Its negation is written as $x \notin A$. Thus, given any object x and any set A, either $x \in A$ or $x \notin A$. For $x \in A$, we sometimes say that *x is an element of A*, *x is a member of A*, *x is contained in A*, or *A contains x*. In the example given below, the set is a concrete set of objects, and 'belonging' has an intuitive meaning. If Z is the set of all integers, $x \in Z$ means x is an integer, and so $2 \in Z$, $3 \in Z$, $\frac{1}{2} \notin Z$ and triangle $ABC \notin Z$. However, it should be emphasized that the concept of belonging is itself *undefined* to the same extent as incidence is undefined in elementary geometry.

REMARKS. For want of an exact definition of integers, we are as yet unable to prove that they form a set. This, however, will be done in Section **6 F** of this book. At present, for the sake of convenience in giving examples, the reader is asked to accept such concepts as the set of all integers, the set of all real numbers and the set of all complex numbers.

The concept of sets and the concept of belonging are the only undefined concepts in our theory. They will be subjected to certain requirements which we call *axioms*; these are postulated true statements. In set theory, we shall be dealing with properties of sets within this system of axioms. Of course, we might have introduced the whole system of axioms right at the beginning and deduced the properties of sets from them, but instead we shall begin with a few axioms and bring the rest into our theory only gradually, as they are needed.

To discuss sets at all, we must first assume that they exist. We therefore postulate our first axiom.

AXIOM OF EXISTENCE. *There is at least one set.*

B. The axiom of extension

We have seen in the last section that, from the axioms of incidence in elementary geometry, a relation between the undefined concept of incidence and the equality of lines can be deduced; i.e. two lines are equal if and only if there are two distinct points that are incident with each of them.

Now, in the vocabulary of set theory, we already have the words *objects*, *sets*, *equality* and *belonging*. Our next step is to postulate as an axiom a basic relation between equality and belonging.

AXIOM OF EXTENSION. *Two sets A and B are equal if and only if A contains every element of B and B contains every element of A.*

The axiom of extension therefore states that

$$A = B \text{ if and only if, for all } x,\ x \in A \Leftrightarrow x \in B$$

or

$$A = B \text{ if and only if } x \in A \text{ for all } x \in B \text{ and } x \in B \text{ for all } x \in A.$$

In other words, a set is uniquely determined by the elements belonging to it. Therefore we can say that all the objects which are elements of a set A form a set which is equal to A.

The axiom of extension is to be regarded as a law governing the undefined concept of belonging with respect to the concept of equality. We notice that in the previous example, where belonging has an intuitive meaning, the axiom of extension is satisfied. However, belonging need not be confined to this meaning. Consider the model where an object

is a set if and only if it is a positive integer and an object x belongs to a set Y (Y is a positive integer) if and only if x is a positive integer dividing Y. According to our understanding of equality, the axiom of extension is satisfied.

As a law, the axiom of extension has a prohibitive aspect which ensures that the symbol $\in$ is 'well-behaved'. In a model where integers are sets and an object x belongs to the integer Y (as a set) if and only if x is an integer dividing Y, the axiom of extension states 'Two integers are equal if and only if they have the same factors'. Now according to our understanding of equality, 6 and –6 are different objects and hence different sets, but they contain the same elements, namely the integers ± 1, ± 2, ± 3, ± 6. Thus this is a model in which the axiom of extension is not satisfied and must therefore be excluded from our theory.

C. Subsets and the empty set

From the two axioms of the last two sections, sets exist and each set is uniquely determined by its elements. We now consider sets whose elements are elements of another set. For example, elements of the set E of all even integers are elements of the set Z of all integers; in this case we say that E is a subset of the set Z. We formulate the general situation as a definition.

DEFINITION 2.1. *Let A and B be sets. B is a* subset *of A if and only if every element of B is an element of A.*

To indicate that a set B is a subset of a set A, we use the symbols $\subset$ or $\supset$ and write $B \subset A$ or equivalently, $A \supset B$. If B is not a subset of A, then we write $B \not\subset A$ or $A \not\supset B$. For $B \subset A$, we sometimes say that B *is included in* A (as a subset) or A *includes* B (as a subset). Using the symbols $\subset$ and $\in$ we may write 2.1 as

$B \subset A$ if and only if A and B are sets and for all x, '$x \in B$' $\Rightarrow$ '$x \in A$'

or

$B \subset A$ if and only if A and B are sets such that $x \in A$ for all $x \in B$.

An immediate consequence of the definition of subset and the axiom of extension is the following theorem, the proof of which is left to the reader as an exercise.

THEOREM 2.2. *Two sets A and B are equal if and only if A is a subset of B and B is a subset of A.*

Because of this theorem, the proof of the statement that two sets A and B are equal is split into two parts; first prove that $A \subset B$ and then prove that $B \subset A$.

The inclusion has the following properties:

(a) *For every set* A, $A \subset A$.

(b) *If* $C \subset B$ *and* $B \subset A$, *then* $C \subset A$.

Statement (a) follows from the fact that for all x, $x \in A \Rightarrow x \in A$. To prove statement (b), we have by hypothesis, for all x, $x \in C \Rightarrow x \in B$, and $x \in B \Rightarrow x \in A$. Applying the hypothetical syllogism, we get for all x, $x \in C \Rightarrow x \in A$; i.e. $C \subset A$. This proves (b). Statement (a) means that every set is a subset of itself. Whenever $B \subset A$ and $A \neq B$, we say that B is a *proper subset* of A.

It is important to distinguish between $\in$ and $\subset$. $\in$ (belonging) is undefined while $\subset$ (inclusion) is defined in terms of $\in$. $B \subset A$ means that B and A are sets and B is a subset of A. $x \in A$ means that A is a set and the object x is an element of A, although x might itself be a set (see Section **I**).

In plane geometry, where the objects of study are points and lines, we often consider the locus of all points specified by some particular common property. For example, given two distinct points A and B, in specifying all those points P of the plane equidistant from A and B, we obtain the perpendicular bisector of the line segment joining A and B. Can we carry out similar constructions in sets? We have at our disposal the axiom of existence which postulates the existence of sets, the axiom of extension which ensures that belonging well-behaves and definition 2.1 which decides whether one of two given sets is a subset of the other. But we are still unable to construct subsets from elements of a given set. In order to be able to do this, we need an axiom that is constructive.

AXIOM OF SPECIFICATION. *Given any set* A *and any statement* $P(x)$ *on elements* x *of* A, *there exists a set* B *whose elements are exactly those elements* x *of* A *for which '*$P(x)$*' is true.*

By the axiom of extension, there is only one such set B which is also a subset of A; and we shall write

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

For example, if Z is the set of all integers and $P(x)$: *x is positive*, then $\{x \in Z : P(x)\}$ is the set of all positive integers. We notice that in applying the axiom of specification to a set and a statement function $P(x)$, we have to make sure that for each $x \in A$ the sentence $P(x)$ is a statement in the sense given in **1 A**. For example the statement function $Q(x)$: *x is dry* does not give rise to a statement on elements x of Z. Therefore $\{x \in Z : Q(x)\}$ is not a subset of Z.

An immediate consequence of the axiom of specification is the existence of a set containing no element. In fact, let A be any set and

$P(x)$ the statement function: $x \neq x$. Then we get the subset $\emptyset = \{x \in A: x \neq x\}$ of A. $\emptyset$ is a set which contains no element. If we assume that $\emptyset$ contains an element x, then the object x, as an element of the set $\emptyset$, must have the properties (i) $x \in A$ and (ii) $x \neq x$. But as an object, x is the same as itself, i.e. $x = x$, contradicting (ii). Therefore $x \notin \emptyset$ for all x, and $\emptyset$ contains no element. By the axiom of extension, $\emptyset$ is the only set which contains no element. So we have the following theorem and definition.

THEOREM 2.3. *There is one and only one set which contains no element.*

DEFINITION 2.4. *The unique set which contains no element is called the* empty set.

The empty set is also called the *void set* or the *null set*, and is denoted throughout this book by the symbol $\emptyset$. Any set which is not the empty set is called a *non-empty set*.

Clearly the only subset of $\emptyset$ is $\emptyset$ itself. Moreover, $\emptyset$ is also characterized by the property that $\emptyset$ is a subset of any set. This means that

(i) for any set A, $\emptyset \subset A$, and

(ii) if a set B is such that $B \subset A$ for every set A, then $B = \emptyset$.

Since $x \in \emptyset$ is always false, the conditional $x \in \emptyset \to x \in A$ for every set A is always true. Therefore $\emptyset \subset A$, proving (i). If B is a subset of every set, then it must be a subset of $\emptyset$. But $\emptyset$ is the only subset of $\emptyset$. Therefore $B = \emptyset$, proving (ii).

Now that we have proved the existence of a unique set which contains no object as its element, we may ask whether there is a set which contains every object as its element. This question is answered in the negative by the following theorem.

THEOREM 2.5. *Given a set A, there exists an object B such that $B \notin A$.*

PROOF. Consider the statement function

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

By the axiom of specification, we obtain a subset

$$B = \{x \in A: x \text{ is a set and } x \notin x\}$$

of the set A. We shall show that this object B does not belong to the set A. Let us assume that $B \in A$. For the object B, either (i) $B \in B$ or (ii) $B \notin B$. If (i) is true, we have $B \in A$ and $B \in B$. Therefore, by the definition of B, we have $B \notin B$, contradicting (i). If (ii) is true, we have $B \in A$ and $B \notin B$. Therefore, by the definition of B, we have $B \in B$, contradicting (ii). Hence the assumption that $B \in A$ is necessarily false and the theorem is proved. ∎

Since the object B constructed in the proof of theorem 2.5 above is itself a set, we have the following corollary:

COROLLARY 2.6. *Given a set A, there exists a set B such that $B \notin A$.*

On account of 2.5 and 2.6, we have to exclude from our theory the concepts of the set of all objects and the set of all sets; in other words, *there is no universal set*, and *all the sets do not form a set.*

We conclude this section with a point concerning notation. Let $P(x)$ be a statement function. If the x's for which '$P(x)$' is true constitute a set, then we may denote that set by

$$\{x: P(x)\}.$$

For example, if A is a set and $P(x)$ is the statement function: $x \in A$, then

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

If A is any set and $P(x)$ any statement on the elements x of A, then

$$\{x: x \in A \text{ and } P(x)\} = \{x \in A: P(x)\}.$$

The empty set may be written as $\{x: x \neq x\}$ or $\{x: P(x)\}$ where $P(x)$ is the negation of any valid statement about x. However, if $P(x)$ is the statement function: $x = x$ or any valid statement about x, then '$P(x)$' is true for every object x, and in this case, the x's for which '$P(x)$' is true do not constitute a set.

D. Venn diagrams

A convenient way of illustrating sets and relations between them is provided by the Venn diagrams. The idea is to represent a set by a simple plane area, usually bounded by a circle.

In the following diagrams, Figure 1 illustrates a set A and Figure 2 illustrates the situation $A \subset B$, where B is the area inside the larger circle and A is the area inside the smaller circle.

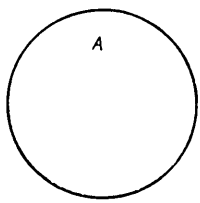


Fig. 1

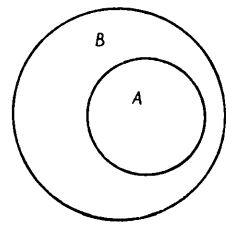


Fig. 2

Each of the following three diagrams illustrates the situation $A \not\subset B$, meaning A is not a subset of B.

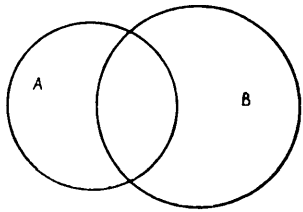


Fig. 3

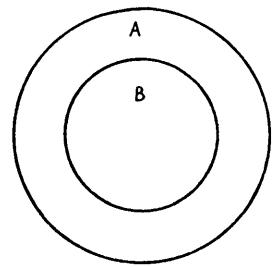


Fig. 4

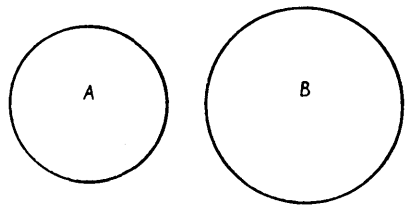


Fig. 5

Diagrams are useful in helping one to understand the theory. It must be remembered, however, that although they may illustrate certain facts they cannot replace the proofs.

E. Unordered pairs and singletons

If a set has only a few elements, then it is frequently denoted by listing its elements, enclosed within braces. Thus a set whose only elements are a, b, c, d is written as $\{a,b,c,d\}$, and the set whose only elements are a and b is written as $\{a,b\}$. The question is whether in fact such sets exist. If the objects a and b are elements of a set A, then the sets $\{a, b\}$, $\{a\}$ and $\{b\}$ can be constructed by the axiom of specification as subsets of A. However, for any given objects a and b, the existence of $\{a,b\}$ does not follow from the three axioms we have so far postulated. In fact from these three axioms, we can only know for sure that one set exists, and this after all may be the empty set $\emptyset$. Our theory would be extremely poor if we could not be sure that there were any other sets at all. For this reason, we shall assume that it is possible to construct sets consisting of only two elements and postulate this as our next axiom.

AXIOM OF PAIRING. *Given any objects a and b, there exists a set containing a and b as its only elements.*

Applying the axiom of extension to this case, we conclude that this set is uniquely determined by the objects a and b, and is denoted by $\{a,b\}$ in accordance with the notation described above. Obviously $\{a,b\} = \{b,a\}$; in other words, the order in which a and b appear is immaterial. Thus we introduce the following definition:

DEFINITION 2.7. *Let a and b be objects. Then the set* $\{a,b\}$ *is called the* unordered pair *of a and b.*

In particular, for any object a, $\{a,a\}$ is a set. By virtue of the axiom of extension, $\{a,a\} = \{a\}$. Therefore $\{a\}$ is a set containing a as its only element and we call this set the *singleton* of a.

The axiom of pairing, useful though it is, still does not enable us to construct sets containing more than two elements. But we are not going to postulate similar axioms for three, four, . . . elements since a single axiom, which we shall postulate later in Section **G**, will render this step unnecessary.

F. Intersections

Let A and B be any two sets. It follows from the axiom of specification that the elements of A which are also elements of B form a set, i.e. the subset $\{x \in A : x \in B\}$ of A. Similarly $\{x \in B : x \in A\}$ is a subset of B. Applying the axiom of extension, we see that $\{x \in A : x \in B\} = \{x \in B : x \in A\}$. This set is formed by all the objects belonging to both A and B, and therefore it can be written as $\{x : x \in A \text{ and } x \in B\}$. Thus we may formulate the following definition:

DEFINITION 2.8. *Let A and B be sets. The* intersection *of A and B is the set* $A \cap B = \{x : x \in A \text{ and } x \in B\}$. *When* $A \cap B = \emptyset$, *we say that the sets A and B are* disjoint.

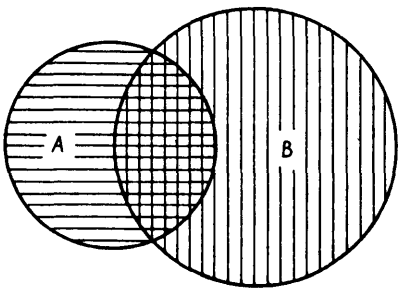


Fig. 6

In the above diagram, A is represented by the area shaded by horizontal lines and B by the area shaded by vertical lines, while $A \cap B$ is represented by the area shaded by both horizontal and vertical lines.

THEOREM 2.9. *For any three sets A, B and C, the following statements are true*:

(a) $A \cap A = A$

(b) $A \cap B = B \cap A$ *(commutative law)*

(c) $A \cap (B \cap C) = (A \cap B) \cap C$ *(associative law)*

(d) $A \cap B \subset A$, $A \cap B \subset B$

(e) $A \cap \emptyset = \emptyset$.

PROOF. (a) Since for all x, $x \in A \Leftrightarrow (x \in A$ and $x \in A)$, it follows that $A \cap A = A$.

(b) This follows from the equivalence $(x \in A$ and $x \in B) \Leftrightarrow (x \in B$ and $x \in A)$.

(c) Since for all x, $[x \in A \cap (B \cap C)] \Leftrightarrow [x \in A$ and $x \in (B \cap C)] \Leftrightarrow [x \in A$ and $(x \in B$ and $x \in C)] \Leftrightarrow [(x \in A$ and $x \in B)$ and $x \in C] \Leftrightarrow [x \in (A \cap B)$ and $x \in C] \Leftrightarrow [x \in (A \cap B) \cap C]$, therefore $A \cap (B \cap C) = (A \cap B) \cap C$.

(d) This follows from the definition of intersection.

(e) Putting $B = \emptyset$ in (d), we get A $\cap \emptyset \subset \emptyset$. Since $\emptyset$ is the only subset of $\emptyset$, $A \cap \emptyset = \emptyset$. ∎

It follows from (c) of this theorem that $A \cap (B \cap C)$ may be written as $A \cap B \cap C$. Intersection of more than three sets may be obtained by iteration:

$$A \cap B \cap \ldots \cap C \cap D = (A \cap B \cap \ldots \cap C) \cap D.$$

THEOREM 2.10. *If A, B, C, D are sets such that $A \subset C$ and $B \subset D$, then $A \cap B \subset C \cap D$.*

COROLLARY 2.11. *If C is a subset of the sets A and B, then $C \subset A \cap B$.*

The proofs are straightforward and are left as exercises.

It should be noted that $A \cap B$ includes every common subset of A and B, and is itself a common subset of A and B; in the sense of inclusion, it is the largest common subset of A and B.

G. Unions

For any two sets A and B, we have constructed the intersection $A \cap B$ as the largest set (in the sense of inclusion) included in both A and B as a subset. Dual to this, we wish to find a set which is the

smallest set (in the sense of inclusion) including both A and B as subsets. If such a set exists, it must consist of all those objects which are elements of A or of B. If A and B are both subsets of a set E, the subset $\{x \in E\colon x \in A \text{ or } x \in B\}$ of E satisfies our conditions. Furthermore, if F is another set which includes both A and B as subsets, then it can readily be proved that $\{x \in F\colon x \in A \text{ or } x \in B\}$ and $\{x \in E\colon x \in A \text{ or } x \in B\}$ are equal. The set in question does not depend on the particular choice of E or F, and therefore it is permissible to write it as $\{x\colon x \in A \text{ or } x \in B\}$.

However, if A and B are arbitrary sets, then the above construction cannot be performed since the existence of a set including both A and B as subsets cannot be deduced from the axioms we have so far postulated. We propose to fill this gap by a further axiom for constructing sets.

AXIOM OF INCLUSION. *Given any sets A and B, there exists a set which includes both A and B as subsets.*

Now the construction explained in the last paragraph is again applicable, and so we are able to adopt the following definition.

DEFINITION 2.12. *The* union *of the sets A and B is the set*

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

In the following diagram, A is shaded by horizontal lines, B by vertical lines and $A \cup B$ is the whole shaded area.

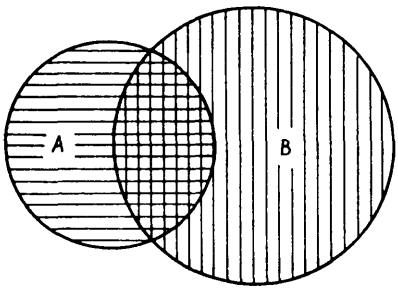


Fig. 7

THEOREM 2.13. *For any sets A, B and C, the following statements are true:*

(a) $A \cup A = A$

(b) $A \cup B = B \cup A$ *(commutative law)*

(c) $A \cup (B \cup C) = (A \cup B) \cup C$ *(associative law)*

(d) $A \subset A \cup B$, *and* $B \subset A \cup B$

(e) $A \cup \emptyset = A$.

The proof is similar to that of Theorem 2.9.

As in the case of intersection, the brackets in $A \cup (B \cup C)$ may be omitted, and similarly we can obtain the union of more than three sets by iteration.

THEOREM 2.14. *If A, B, C, D are sets such that $A \subset C$ and $B \subset D$, then $A \cup B \subset C \cup D$.*

COROLLARY 2.15. *If A and B are subsets of a set C, then $A \cup B$ is a subset of C.*

The set $A \cup B$ is clearly the smallest set (in the sense of inclusion) that includes both A and B as subsets.

The conjunction and disjunction of statements are used respectively in the definitions of the intersection and the union of sets. Therefore, corresponding to the symbols $\wedge$ and $\vee$ of statement calculus, we have used the symbols $\cap$ and $\cup$ to denote respectively the intersection and the union in set theory. Using the distributive laws in statement calculus, we prove the distributive laws of intersection and union:

THEOREM 2.16. *For any three sets A, B and C,*

(a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,

(b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

PROOF. (a) For all x, $[x \in A \text{ and } x \in B \cup C] \Leftrightarrow [x \in A \text{ and } (x \in B \text{ or } x \in C)] \Leftrightarrow [(x \in A \text{ and } x \in B) \text{ or } (x \in A \text{ and } x \in C)] \Leftrightarrow [x \in A \cap B \text{ or } x \in A \cap C]$.
Hence $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
The proof of (b) is left to the reader as an exercise.

Finally, note that the axiom of inclusion enables us to construct sets containing three, four, . . . elements; thus

$$\{a,b,c\} = \{a\} \cup \{b\} \cup \{c\},$$
$$\{a,b,c,d, \ldots e\} = \{a\} \cup \{b\} \cup \{c\} \cup \{d\} \cup \ldots \cup \{e\}$$

This justifies our statement at the end of Section **E**.

H. Complements

Given any two sets, let us consider a set whose elements belong to one of these two sets but not to the other. This concept is formulated in the following two definitions 2.17 and 2.18.

DEFINITION 2.17. *Let E be a set and A a subset of E. Then the* (absolute) complement *of A in E is the set $E \setminus A$ defined as*

$$E \setminus A = \{x \in E : x \notin A\}.$$

$E \setminus A$ is also written as $\mathbf{C}_E A$, or simply $\mathbf{C}A$ if the dependence on E is clear from the context. It is also permissible to write

$$E \setminus A = \{x: x \in E \text{ and } x \notin A\}.$$

Obviously, $E \setminus A$ is a subset of E. In the diagram, A and E are the areas bounded by the smaller and the bigger circles respectively, and $E \setminus A$ is the shaded area.

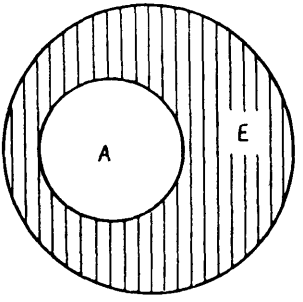


Fig. 8

DEFINITION 2.18. *Let A and B be any two sets. Then the* (relative) complement of A *in* B *is the set* $B \setminus A$ *defined as*

$$B \setminus A = \{x \in B: x \notin A\}.$$

It is also permissible to write

$$B \setminus A = \{x: x \in B \text{ and } x \notin A\}.$$

In the following diagram, A is the area bounded by the circle on the left, B is the area bounded by the circle on the right and $B \setminus A$ is the shaded area.

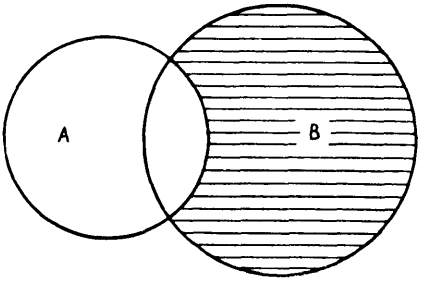


Fig. 9

Notice that in definition 2.18 it is not required that $A \subset B$, but we still have $B \setminus A \subset B$. If $A \subset B$, then the relative complement of A in B is the same as the absolute complement of A in B.

THEOREM 2.19. *If A and B are any two sets, then*

(a) $A \setminus A = \emptyset$,

(b) $A \setminus \emptyset = A$,

(c) $\emptyset \setminus A = \emptyset$,

(d) $B \setminus A = \emptyset$ *if and only if* $B \subset A$,

(e) $(A \setminus B) \cap (B \setminus A) = \emptyset$, *and*

(f) $A \cap (B \setminus A) = \emptyset$.

PROOF. (a) For any object x, $x \in A \setminus A \Leftrightarrow x \in A$ and $x \notin A$. But the conjunction on the right-hand side is always false, and so the statement on the left-hand side is also false. Therefore the set $A \setminus A$ contains no element.

(b) We already know that $A \setminus \emptyset \subset A$; therefore we need only prove that $A \subset A \setminus \emptyset$. Now for all objects x, $x \notin \emptyset$ is always true; therefore $x \in A \Rightarrow x \in A$ and $x \notin \emptyset$. This shows that $A \subset A \setminus \emptyset$.

(c) Since the only subset of $\emptyset$ is $\emptyset$, and $\emptyset \setminus A \subset \emptyset$, therefore $\emptyset \setminus A = \emptyset$.

(d) If $B \subset A$, then for all objects x, $x \in B \Rightarrow x \in A$. In other words, for all objects x, the statement '$x \notin B$ or $x \in A$' is true, so that its negation is false. This negation is the statement '$x \in B$ and $x \notin A$', which is equivalent to $x \in B \setminus A$. Therefore $B \setminus A = \emptyset$. To prove the converse we need only reverse the argument.

(e) For all x, $[x \in (A \setminus B) \cap (B \setminus A)] \Leftrightarrow [x \in A \setminus B$ and $x \in B \setminus A] \Leftrightarrow [(x \in A$ and $x \notin B)$ and $(x \in B$ and $x \notin A)] \Leftrightarrow [(x \in A$ and $x \notin A)$ and $(x \in B$ and $x \notin B)] \Leftrightarrow [x \in A \setminus A$ and $x \in B \setminus B] \Leftrightarrow x \in \emptyset$. Therefore $(A \setminus B) \cap (B \setminus A) = \emptyset$.

The proof of (f) is left to the reader as an exercise.

THEOREM 2.20. *Let E be a set. For any two subsets A and B of E,*

(a) $E \setminus (E \setminus A) = A$,

(b) $A \subset B$ *if and only if* $E \setminus B \subset E \setminus A$.

PROOF. (a) For all x, $[x \in E \setminus (E \setminus A)] \Leftrightarrow [x \in E$ and $x \notin E \setminus A] \Leftrightarrow [x \in E$ and $\sim (x \in E$ and $x \notin A)] \Leftrightarrow [x \in E$ and $(x \notin E$ or $x \in A)] \Leftrightarrow [(x \in E$ and $x \notin E)$ or $(x \in E$ and $x \in A)] \Leftrightarrow [x \in E$ and $x \in A] \Leftrightarrow x \in E \cap A \Leftrightarrow x \in A$. Therefore $E \setminus (E \setminus A) = A$.

(b) To prove (b) we have to prove the two parts:

(1) if $A \subset B$, then $E \setminus B \subset E \setminus A$,

(2) if $E \setminus B \subset E \setminus A$, then $A \subset B$.

If $A \subset B$, then for all x, $x \in A \Rightarrow x \in B$, so that we have $x \notin B \Rightarrow x \notin A$. On the other hand, we have $x \in E \Rightarrow x \in E$. Therefore ($x \in E$ and $x \notin B$) $\Rightarrow$ ($x \in E$ and $x \notin A$), showing that $E \setminus B \subset E \setminus A$. This proves (1).
$E \setminus B$ and $E \setminus A$ are subsets of E. If $E \setminus B \subset E \setminus A$, then we can apply part (1) to get $E \setminus (E \setminus A) \subset E \setminus (E \setminus B)$. It follows from (a) that $A \subset B$. This proves (2). ∎

With $\wedge$, $\vee$ and $\sim$ corresponding to $\cap$, $\cup$ and $\setminus$ respectively, we have De Morgan's laws for complements.

THEOREM 2.21. (De Morgan's Laws) *Let E be a set. If A and B are subsets of E, then the following statements are true:*

(a) $E \setminus (A \cap B) = (E \setminus A) \cup (E \setminus B)$,

(b) $E \setminus (A \cup B) = (E \setminus A) \cap (E \setminus B)$.

PROOF. (a) For all objects x, $[x \in E \setminus (A \cap B)] \Leftrightarrow [x \in E$ and $x \notin (A \cap B)] \Leftrightarrow [x \in E$ and $(x \notin A$ or $x \notin B)] \Leftrightarrow [(x \in E$ and $x \notin A)$ or $(x \in E$ and $x \notin B)] \Leftrightarrow [x \in E \setminus A$ or $x \in E \setminus B] \Leftrightarrow x \in (E \setminus A) \cup (E \setminus B)$. Therefore $E \setminus (A \cap B) = (E \setminus A) \cup (E \setminus B)$.

(b) Substituting $E \setminus A$ for A and $E \setminus B$ for B in (a), we get $E \setminus [(E \setminus A) \cap (E \setminus B)] = [E \setminus (E \setminus A)] \cup [E \setminus (E \setminus B)] = A \cup B$ [by 2.20 (a)].
Taking the absolute complements of both sides in E and applying 2.20(a) again, we get $(E \setminus A) \cap (E \setminus B) = E \setminus (A \cup B)$. ∎

I. Power sets

In studying a given set A, we often have to consider its subsets. It is important for us to know, therefore, whether there exists a set that contains all the subsets of a given set A as elements. If A contains only a finite number of elements, then it has only a finite number of subsets, and these subsets constitute a set according to the construction at the end of Section G. However, if the set A is not such a set, then the existence of a set of all subsets of A cannot be deduced from the axioms we have at our disposal. In order to overcome this difficulty, we now postulate another axiom for constructing sets.

AXIOM OF POWER. *The totality of all subsets of any given set constitutes a set.*

If we accept this axiom, the following definition is meaningful:

DEFINITION 2.22. *Let A be a set. The set of all subsets of A is the* power set *of A and is denoted by $\mathfrak{P}(A)$. $\mathfrak{P}(A)$ is defined as*

$$\mathfrak{P}(A) = \{X : X \subset A\}.$$

From the very definition of $\mathfrak{P}(A)$, we see that $\emptyset \in \mathfrak{P}(A)$ and $A \in \mathfrak{P}(A)$. Clearly $\emptyset \subset \mathfrak{P}(A)$; but in general $A \subset \mathfrak{P}(A)$ is not true, although it is true when $A = \emptyset$.

For sets with only a finite number of elements, we are able to construct their power sets explicitly. Thus,

$$\mathfrak{P}(\emptyset) = \{\emptyset\},$$
$$\mathfrak{P}(\{a\}) = \{\emptyset, \{a\}\},$$
$$\mathfrak{P}(\{a,b\}) = \{\emptyset, \{a\}, \{b\}, \{a,b\}\}.$$

In general, the power set of a set containing n elements has 2^n elements.

J. Unions and intersections of subsets

Within the power set $\mathfrak{P}(A)$ of a set A, unions and intersections of subsets of A can be constructed under more general conditions than those of Section **G.**

If $\mathfrak{S}$ is a subset of $\mathfrak{P}(A)$, then each element B of $\mathfrak{S}$ is, by definition, a subset of A. Using the axiom of specification, we can construct in A the subset $\{x \in A : x \in B \text{ for some } B \in \mathfrak{S}\}$, which is precisely the set of all members x of A belonging to some element B of $\mathfrak{S}$. This set is called the *union of all* $B \in \mathfrak{S}$, (or over $\mathfrak{S}$) and written

$$\bigcup_{B \in \mathfrak{S}} B = \{x : x \in B \text{ for some } B \in \mathfrak{S}\}.$$

It is not difficult to show that the concept of union as formulated in Section **G** is a special case of the above; therefore the name is justified. Clearly

$$\bigcup_{B \in \mathfrak{P}(A)} B = A \text{ and } \bigcup_{B \in \emptyset} B = \emptyset.$$

Similarly, the *intersection of all* $B \in \mathfrak{S}$ (or over $\mathfrak{S}$) is defined as

$$\bigcap_{B \in \mathfrak{S}} B = \{x \in A : x \in B \text{ for all } B \in \mathfrak{S}\}.$$

Consequently

$$\bigcap_{B \in \mathfrak{P}(A)} B = \emptyset \text{ and } \bigcap_{B \in \emptyset} B = A.$$

To conclude this section, we introduce the concept of partition of a set, which will be useful later.

DEFINITION 2.23. *Given a set A, a* partition *of A is a subset $\mathfrak{S}$ of $\mathfrak{P}(A) \setminus \{\emptyset\}$, such that*

(i) $\bigcup_{B \in \mathfrak{S}} B = A$, *and*

(ii) $B' \cap B = \emptyset$ *for any two distinct elements B and B' of $\mathfrak{S}$.*

In other words, a subset $\mathfrak{S}$ of $\mathfrak{P}(A)\setminus\{\emptyset\}$ is a partition of the set A if and only if (iii) each element of A is contained in ***at least one*** element of $\mathfrak{S}$ and (iv) each element of A is contained in ***at most one*** element of $\mathfrak{S}$.

Clearly the singleton $\{A\}$ is trivially a partition of $A \neq \emptyset$, and so is the set of all singletons of elements of A. We shall encounter more interesting examples in the later chapters.

K. Exercises

1. Let $A = \{\{\emptyset\}, \emptyset\}$. State whether each of the following statements is true: $\{\{\emptyset\}\} \in A$, $\emptyset \in A$, $\{\emptyset\} \in A$, $\{\{\emptyset\}\} \subset A$, $\emptyset \subset A$, $\{\emptyset\} \subset A$.

2. If a, b, c, d are distinct objects, determine which of the five sets $\{a,b,c\}$, $\{b,c,a,b\}$, $\{c,a,c,b\}$, $\{b,c,b,a\}$, $\{a,b,c,d\}$ are equal.

3. Let $A = \{a,b,c\}$, $B = \{a,b\}$, $C = \{a,b,d\}$, $D = \{a\}$ and $E = \{b,c\}$, where a,b,c are distinct objects. State whether each of the following statements is true or false:
   (a) $B \subset A$, (b) $E \neq C$, (c) $D \not\subset B$, (d) $D \subset A$, (e) $A = B$.

4. Let $A = \{1,2,3,4,5,6\}$, $B = \{4,5,6,7,8,9\}$, $C = \{2,4,6,8\}$, $D = \{4,5\}$, $E = \{5,6\}$, $F = \{4,6\}$, and X a set which satisfies the following conditions: $X \subset A$, $X \subset B$ and $X \not\subset C$. Determine which of the sets A, B, C, D, E, F can equal X.

5. Which of the following sets is the empty set?
   (a) $\{x$: *x is an odd integer and* $x^2 = 4\}$,
   (b) $\{x$: *x is an integer and* $x + 8 = 8\}$,
   (c) $\{x$: *x is a positive integer and* $x < 1\}$.

6. Consider the following sets of figures in a Euclidean plane S: $A = \{x$: *x is a quadrilateral in* $S\}$, $B = \{x$: *x is a parallelogram in* $S\}$, $C = \{x$: *x is a rhombus in* $S\}$, $D = \{x$: *x is a rectangle in* $S\}$, $E = \{x$: *x is a square in* $S\}$.
   (a) Determine which sets are subsets of others.
   (b) Find the union and intersection of each pair of sets.

7. Let $A = \{a,b,c,d\}$, $B = \{b,d,f,h\}$, $C = \{c,d,e,f\}$. Find
   (a) $A \cap B$, $A \cap C$, $B \cap C$,
   (b) $A \cup B$, $A \cup C$, $B \cup C$,
   (c) $A \setminus B$, $B \setminus A$, $B \setminus C$, $C \setminus B$, $A \setminus C$, $C \setminus A$.

8. Let R be the set of real numbers, $A = \{x \in R\colon 1 \leqslant x \leqslant 3\}$ and $B = \{x \in R\colon 2 \leqslant x \leqslant 4\}$. Find
   (a) $A \cup B$, (b) $A \cap B$, (c) $(R \setminus A) \cap B$, (d) $(R \setminus B) \cap A$,
   (e) $(R \setminus A) \cap (R \setminus B)$, (f) $(R \setminus A) \cup (R \setminus B)$,
   (g) $B \cup [A \cap (R \setminus B)]$, (h) $[(R \setminus A) \cap B] \cup [(R \setminus B) \cap A]$.

9. Let Z be the set of all integers, $A = \{x \in Z$: *x is a multiple of* $10\}$, and $B = \{x \in Z$: *x is a multiple of* $15\}$. What is $A \cap B$? Can you generalize this result?

10. What is the power set of (a) the set $\{a,b,c\}$, and (b) the set $\{a,\{b,c\}\}$?

11. A, B, C are three sets. Prove that
    (a) if $A \subset B$ and $A \not\subset C$, then $B \not\subset C$,
    (b) if $A \cap B = A$ and $A \cap C \neq \emptyset$, then $B \cap C \neq \emptyset$.

12. Show that the statement

    If $A \subset B$ *and* $B \not\subset C$, *then* $A \not\subset C$

is not always true, by giving an example in which $A \subset B$ and $B \not\subset C$ are true but $A \not\subset C$ is false.

13. Prove 2.10 and 2.11.

14. Prove 2.13, 2.14, 2.15 and part *(b)* of 2.16.

15. Prove that for any two sets A and B,
    (a) $A \cup (A \cap B) = A$,
    (b) $A \cap (A \cup B) = A$,
    (c) $A \cap (B \setminus A) = \emptyset$.

16. In the following statements, A and B are sets:
    (a) $A \subset A \cup B$.
    (b) $A \supset A \cup B$.
    (c) $A \subset A \cap B$.
    (d) $A \supset A \cap B$.
    (e) If $A \supset B$, then $A \cap B = A$.
    (f) If $A \supset B$, then $A \cap B = B$.
    (g) If $A \supset B$, then $A \cup B = A$.
    (h) If $A \supset B$, then $A \cup B = B$.

    Determine in each case whether the statement is *always* true or not, and, if the statement is not always true, determine when it is true. Illustrate each case by a Venn diagram.

17. Prove that $A \setminus B = B \setminus A$ is not always true, and decide when $A \setminus B$ and $B \setminus A$ are both empty.

18. Prove the following statements for any two subsets A and B of a set E:
    (a) $B \setminus A \subset E \setminus A$.
    (b) $B \setminus (E \setminus A) = B \cap A$.
    (c) If $A \cap B = \emptyset$, then $B \cap (E \setminus A) = B$ and $A \cup (E \setminus B) = E \setminus B$.
    (d) $(E \setminus A) \setminus (E \setminus B) = B \setminus A$.
    (e) If $A \cup B = E$ and $A \cap B = \emptyset$ then $B = E \setminus A$.

19. Is the statement

$$A \cap B = \emptyset \text{ } \textit{iff} \text{ } (E \setminus A) \cap (E \setminus B) = \emptyset$$

    true for any two subsets A and B of a set E? If not, illustrate by a counter example.

20. Determine whether the following statements are true for any three sets A, B, C:
    (a) $A \setminus (B \setminus C) = (A \setminus B) \setminus C$.
    (b) $A \cap (B \setminus C) = (A \cap B) \setminus (A \cap C)$.
    (c) If $A \subset B$, then $(C \setminus A) \subset (C \setminus B)$.

21. If A, B, C, D are sets such that $\{A,B\} = \{C,D\}$, prove that $A \cap B = C \cap D$ and $A \cup B = C \cup D$.

22. A, B and C are subsets of a set E. Express $A \setminus (B \setminus C)$ as a union of sets of the form

$$X_A \cap X_B \cap X_C$$

where X_A (respectively X_B, X_C) is equal to A (respectively B, C) or $E \setminus A$ (respectively $E \setminus B$, $E \setminus C$).

23. Prove the following statements for any three sets A, B and C:
    (a) $(A \cup B) \setminus C = (A \setminus C) \cup (B \setminus C)$.
    (b) $A \setminus (B \setminus C) = (A \setminus B) \cup (A \cap C)$
    (c) $A \setminus (B \cup C) = (A \setminus B) \setminus C$.

*24. For two sets A and B, we denote by $A \triangle B$ the set $(A \setminus B) \cup (B \setminus A)$. Prove that for any three sets A, B and C,
    (a) $A \triangle B = (A \cup B) \setminus (A \cap B)$.
    (b) $A \triangle B = B \triangle A$.
    (c) $A \triangle (B \triangle C) = (A \triangle B) \triangle C$.
    (d) $A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$.

*25. A subset $\mathfrak{T}$ of the power set $\mathfrak{P}(X)$ of a set X such that $\bigcup_{B \in \mathfrak{T}} B = X$ is said to be a *topology* for X iff the following two axioms are satisfied:
    (T.1) If A and B are elements of $\mathfrak{T}$ then $A \cap B \in \mathfrak{T}$.
    (T.2) For any subset $\mathfrak{S}$ of $\mathfrak{T}$, $\bigcup_{B \in \mathfrak{S}} B \in \mathfrak{T}$.

    A *topological space* is a set X together with a topology $\mathfrak{T}$ for X. Elements of X are called *points* and elements of $\mathfrak{T}$ *open sets* of the topological space X. If no confusion about the topology is possible, we may also say that X is a topological space.
    (a) Show that $\emptyset$ and X are open sets of a topological space X.
    (b) Show that $\mathfrak{P}(X)$ and $\{\emptyset, X\}$ are topologies for a set X.
    (c) If X is the set of all real numbers, show that the set consisting of all unions of open intervals is a topology for X, called the *usual topology* of the real line.
    (d) Let $X = \{a,b,c,d\}$. Construct ten topologies for X.

*26. Let X be a topological space. A subset U of X is said to be a *neighbourhood* of a point x of X if and only if U includes an open set to which x belongs. Prove that a subset A of X is open if and only if it is a neighbourhood of each of its points.

*27. Let $\mathfrak{U}_x$ be the set of neighbourhoods of a point x of a topological space X. Prove that $\mathfrak{U}_x$ has the following properties:
    (a) If $B \subset X$ and $B \supset U$ for some $U \in \mathfrak{U}_x$, then $B \in \mathfrak{U}_x$.
    (b) $X \in \mathfrak{U}_x$.
    (c) For all sets A and B, $A \in \mathfrak{U}_x$ and $B \in \mathfrak{U}_x$ implies $A \cap B \in \mathfrak{U}_x$.
    (d) For all B, $B \in \mathfrak{U}_x$ implies $x \in B$.
    (e) If $U \in \mathfrak{U}_x$, then there exists $V \in \mathfrak{U}_x$ such that, for all $y \in V$, $U \in \mathfrak{U}_y$.

*28. A subset of a topological space X is said to be a *closed set* if and only if its complement in X is open. If $\mathfrak{F}$ is a set of subsets of a set X satisfying the conditions:

(i) if $A \in \mathfrak{F}$ and $B \in \mathfrak{F}$, then $A \cup B \in \mathfrak{F}$,

(ii) $\emptyset \in \mathfrak{F}$,

(iii) for any subset $\mathfrak{S}$ of $\mathfrak{F}$, $\bigcap_{B \in \mathfrak{S}} B \in \mathfrak{F}$,

and if $\mathfrak{T} = \{A \in \mathfrak{P}(X) : X \setminus A \in \mathfrak{F}\}$, prove that $\mathfrak{T}$ is a topology for X such that $\mathfrak{F}$ is precisely the set of closed sets for this topology.

CHAPTER 3

RELATIONS

A. Ordered pairs

We have seen in Section **2 E** that, given any two objects x and y, there is a set $\{x,y\}$ which has x and y as its only elements. Moreover, $\{x,y\} = \{y,x\}$; in other words, the order in which the objects x and y appear is immaterial to the construction of the set $\{x,y\}$. For this reason the set $\{x,y\}$ is called an unordered pair.

Let us recall a well-known technique used in plane analytic geometry. With respect to a fixed rectangular coordinate system, each point P in the plane is uniquely represented by a pair (x,y) of numbers called the coordinates of P. Two such pairs (x_1,y_1) and (x_2,y_2) are equal, and hence represent one and the same point of the plane if and only if $x_1 = x_2$ and $y_1 = y_2$. For example, $(1,3) \neq (3,1)$. We see, therefore, that the order in which the numbers 1 and 3 appear in (1,3), unlike in {1,3}, is essential. The coordinates of a point in the plane is an example of the so-called *ordered pair* which we shall define presently.

Let x and y be two objects. Then by the results of Section **2 E** we can construct the sets

(1) $\quad X = \{x\}$ and $Y = \{x,y\}$.

Regarding X and Y as objects, we can then further construct the set

(2) $\quad \{X,Y\} = \{\{x\},\{x,y\}\}$.

Clearly the set (2) is a well-defined set; and we shall see later (in 3.2) that it has the property of the coordinates of a point in the plane, described briefly in the last paragraph.

DEFINITION 3.1. *Let x and y be objects. The set $(x,y) = \{\{x\},\{x,y\}\}$ is called the* ordered pair *of the objects x and y.*

In the ordered pair (x,y), x and y are called respectively the *first* and the *second coordinates* of (x,y). We shall now justify the name *ordered pair* by proving the following theorem:

THEOREM 3.2. *$(x,y) = (s,t)$ if and only if $x = s$ and $y = t$.*

PROOF. Let $X = \{x\}$, $Y = \{x,y\}$, $S = \{s\}$, $T = \{s,t\}$. By definition $(x,y) = \{X,Y\}$ and $(s,t) = \{S,T\}$.

If $x = s$ and $y = t$, then $X = S$ and $Y = T$. Therefore $\{X,Y\} = \{S,T\}$; i.e. $(x,y) = (s,t)$.

Conversely, if $(x,y) = (s,t)$, then $\{X,Y\} = \{S,T\}$, and we have the following cases:

Case 1. $X = S$ and $Y = T$. It follows from this assumption that $x = s$ and $\{x,y\} = \{s,t\}$. Therefore $x = s$ and $y = t$.

Case 2. $X = T$ and $Y = S$. It follows from this assumption that $x = s = t$ and $x = y = s$. Therefore $x = s$ and $y = t$. ∎

Using the concept of ordered pairs, we can now introduce the concept of *ordered triples*.

DEFINITION 3.3. *x, y and z are three objects. The* ordered triple (x,y,z) *is the ordered pair* $((x,y),z)$.

It follows from 3.2 that two ordered triples (x,y,z) and (s,t,u) are equal if and only if $x = s$, $y = t$ and $z = u$.

Similarly, *ordered n-tuples* can be defined by iteration.

B. Cartesian products of sets

By restricting the choice of the coordinates of ordered pairs to members of given sets, we obtain the *cartesian product*. The name is derived from the well-known method of coordinating the points in a plane first used by the French philosopher and mathematician RENÉ DESCARTES (1596–1650).

Now the important question arises: given two sets A and B, do the ordered pairs (a,b), where $a \in A$ and $b \in B$, constitute a set? The answer to this is *yes* and the proof is as follows:

According to the results of Section **2 G**, we can construct the union $A \cup B$, which is a set. By applying the axiom of power to the set $A \cup B$ twice, we obtain the sets $\mathfrak{P}(A \cup B)$ and $\mathfrak{P}(\mathfrak{P}(A \cup B))$. Consider an ordered pair $(a,b) = \{\{a\}, \{a,b\}\}$ where $a \in A$ and $b \in B$. Since $\{a\}$ and $\{a,b\}$ are both elements of the set $\mathfrak{P}(A \cup B)$ the ordered pair $(a,b) = \{\{a\}, \{a,b\}\}$ is a subset of the set $\mathfrak{P}(A \cup B)$. Therefore, it is an element of the power set $\mathfrak{P}(\mathfrak{P}(A \cup B))$. We can now construct the subset $\{x \in \mathfrak{P}(\mathfrak{P}(A \cup B)) : x = (a,b)$ *for some* $a \in A$ *and* $b \in B\}$ which, by the axiom of extension, is the unique set consisting exactly of all the ordered pairs

(a,b), where $a \in A$ and $b \in B$. This set will be denoted by $A \times B$ and can be written as

$$A \times B = \{(a,b)\colon a \in A \text{ and } b \in B\}.$$

DEFINITION 3.4. *Given two sets A and B, the* cartesian product $A \times B$ *is the set of all ordered pairs (a,b) where $a \in A$ and $b \in B$.*

It is not difficult to see that $A \times B \neq B \times A$ unless $A = \emptyset$ or $B = \emptyset$ or $A = B$; in the last case we may also write A^2 in place of $A \times A$.

The reader will have no difficulty in seeing that the cartesian plane is the cartesian product of the set of all real numbers with itself.

C. Relations

Usually a relation consists in associating objects of one kind with objects of another kind. As an example, let us consider the relation called *marriage*, which may be expressed in the following way: x is a man, y is a woman, and x is a husband of y. Using the concept of ordered pairs, we set up an ordered pair (x,y) where x is a man, y is a woman, and x is a husband of y. Suppose now that we are to investigate this relation among the population of Hong Kong. We can do this by going through the entire Hong Kong marriage register and constructing a set G of all such ordered pairs. G is then a subset of the cartesian product of the set of all Hong Kong men and the set of all Hong Kong women. Thus *Cheung Sam* is married to *Lee Sai* if and only if the ordered pair (*Cheung Sam*, *Lee Sai*) belongs to G. Now the marriage relation among the population of Hong Kong consists of (i) the set M of all Hong Kong men, (ii) the set F of all Hong Kong women, and (iii) the subset G of $M \times F$.

DEFINITION 3.5. *A* relation *R from a set A to a set B is an ordered triple (A,B,G) where $G \subset A \times B$; G is called the* graph *of R; A and B are respectively called the* set of departure *and the* set of destination *of R.*

It follows from this definition and 3.2 that two relations are equal if and only if they have the same set of departure, the same set of destination and the same graph. Consequently two different relations may sometimes have identical graphs. A relation is therefore not to be identified with its graph.

DEFINITION 3.6. *Given a relation $R = (A,B,G)$, the first and the second* projections *of the relation R are respectively the subsets of A and B defined as*

$$\mathrm{pr}_1 R = \{a \in A\colon (a,b) \in G \textit{ for some } b \in B\},$$
$$\mathrm{pr}_2 R = \{b \in B\colon (a,b) \in G \textit{ for some } a \in A\}.$$

These two subsets of A and B are characterized by a minimal property:

THEOREM 3.7. *If $R = (A,B,G)$ is a relation, then $\mathrm{pr}_1 R$ and $\mathrm{pr}_2 R$ are the smallest subsets of A and B respectively such that $G \subset \mathrm{pr}_1 R \times \mathrm{pr}_2 R$.*

PROOF. It follows immediately from the definition that $\mathrm{pr}_1 R \subset A$, $\mathrm{pr}_2 R \subset B$ and $G \subset \mathrm{pr}_1 R \times \mathrm{pr}_2 R$. Therefore it remains to be shown that if C and D are respectively subsets of A and B such that $G \subset C \times D$, then $\mathrm{pr}_1 R \subset C$ and $\mathrm{pr}_2 R \subset D$. If $a \in \mathrm{pr}_1 R$, then there exists $b \in B$ such that $(a,b) \in G$. By assumption, $G \subset C \times D$; therefore $(a,b) \in C \times D$ and hence $a \in C$, proving $\mathrm{pr}_1 R \subset C$. Similarly $\mathrm{pr}_2 R \subset D$. ∎

If no confusion about the sets of departure and destination is possible, we may write aRb (read 'a is R-related to b') for $(a,b) \in G$, where $R = (A,B,G)$.

D. Inverses and compositions

Intersection and union of two relations $R_1 = (A,B,G_1)$ and $R_2 = (A,B,G_2)$ having the same set of departure and the same set of destination are defined as follows:

$$R_1 \cap R_2 = (A,B,G_1 \cap G_2)$$
$$R_1 \cup R_2 = (A,B,G_1 \cup G_2)$$

But we shall not discuss these further, as they are seldom used in mathematics. The more important operations on relations are the forming of *inverses* and *compositions*, which we introduce now.

DEFINITION 3.8. *If $R = (A,B,G)$ is a relation, then the* inverse relation *of the relation R is the relation $R^{-1} = (B,A,G^{-1})$ where*

$$G^{-1} = \{(b,a) \in B \times A : (a,b) \in G\}$$

Notice that the set of departure of R is the set of destination of R^{-1} and that the set of destination of R is the set of departure of R^{-1}. Furthermore $\mathrm{pr}_1 R = \mathrm{pr}_2 R^{-1}$ and $\mathrm{pr}_2 R = \mathrm{pr}_1 R^{-1}$.

EXAMPLE 3.9. Let X be the set of all real numbers. Then the cartesian product $X \times X$ may be regarded as the ordinary x-y-plane. If L is a straight line in the plane, then we can define a relation R as (X,X,L). Now it is not difficult to see that the graph L^{-1} of R^{-1} is the straight line symmetric to L with respect to the line $x - y = 0$ (see Fig. 10).

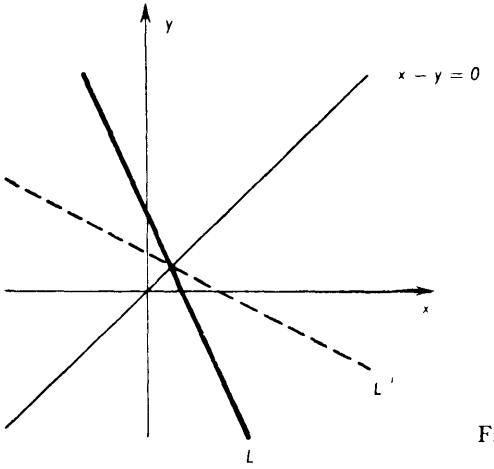


Fig. 10

DEFINITION 3.10. *Given two relations* $R = (A,B,G)$ *and* $S = (B,C,H)$, *the* composition *of* R *and* S *is defined as the relation* $S \circ R = (A,C,H \circ G)$, *where* $H \circ G = \{(a,c) \in A \times C$: $(a,b) \in G$ *and* $(b,c) \in H$ *for some* $b \in B\}$.

Notice that among the sets of departure and destination the set of departure of S is the set of destination of R, the set of departure of $S \circ R$ is the set of departure of R, and the set of destination of $S \circ R$ is the set of destination of S. Furthermore, $\mathrm{pr}_1 R \supset \mathrm{pr}_1 S \circ R$, $\mathrm{pr}_2 S \supset \mathrm{pr}_2 S \circ R$.

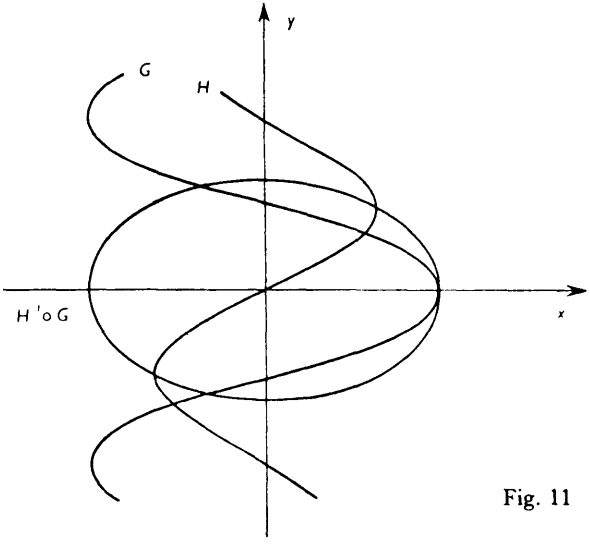


Fig. 11

EXAMPLE 3.11. (notation as in 3.9). Let $G = \{(3\ cos\ \theta, \theta)\colon \theta \in X\}$, and $H = \{(2\ sin\ \theta, \theta)\colon \theta \in X\}$, so that G is a cosine-curve and H is a sine-curve. Then $R = (X,X,G)$ and $S = (X,X,H)$ are relations, and furthermore we can form the composition $S^{-1} \circ R$. It is easy to show that the graph of $S^{-1} \circ R$ is an ellipse with its major axis equal to 3 and minor axis equal to 2 (see Fig. 11).

THEOREM 3.12. *For any three relations* $R = (A,B,G)$, $S = (B,C,H)$ *and* $T = (C,D,K)$, *the following equalities hold:*

(a) $(R^{-1})^{-1} = R$

(b) $(T \circ S) \circ R = T \circ (S \circ R)$

(c) $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$

The proof of 3.12 is sufficiently simple to be left to the reader as an exercise.

E. Equivalence relations

A relation R is said to be defined in a set A if both the set of departure and the set of destination of R are identical with A. For example, the relations we have given and constructed in 3.9 and 3.11 are relations defined in the set of all real numbers.

Among the relations defined in a set there is one that is particularly simple and useful.

DEFINITION 3.13. *The* diagonal *of a set* A *is the relation* D_A *in* A *whose graph is the set* $\{(a,a) \in A \times A\colon a \in A\}$.

With the introduction of the subscript A, no confusion about the sets of departure and destination will arise; we may therefore conveniently denote the graph of D_A by the same symbol D_A.

Owing to their frequent appearance in set theory and especially in pure mathematics, the following types of relation are of particular importance:

DEFINITION 3.14. $R = (A,A,G)$ *is a relation defined in the set* A.

(i) R *is* reflexive *iff* $D_A \subset G$.

(ii) R *is* symmetric *iff* $R = R^{-1}$.

(iii) R *is* transitive *iff* $G \circ G \subset G$.

(iv) R *is an* equivalence relation *in* A *iff it is reflexive, symmetric and transitive.*

The diagonal D_A and the relation $(A,A, A \times A)$ are clearly equivalence relations.

The properties (i), (ii) and (iii) are independent of one another in the sense that none of them is implied by the other two (see Exercise 8).

Let us now study equivalence relations in detail. Given an equivalence relation $R = (A,A,G)$, we can construct for each element a of A the set

$$a/R = \{b \in A : (a,b) \in G\}$$

a/R is clearly a subset of A, and therefore an element of the power set $\mathfrak{P}(A)$ of A. Consequently, the set

$$A/R = \{p \in \mathfrak{P}(A) : p = a/R \text{ for some } a \in A\}$$

is a well-defined subset of the power set $\mathfrak{P}(A)$ of A.

DEFINITION 3.15. *Let R be an equivalence relation in a set A. For each element a of A, the subset a/R of A is called the* equivalence class *of a by R. The subset A/R of $\mathfrak{P}(A)$ is called the* quotient set *of A by R.*

The main properties of equivalence classes and the quotient set are given in the following two theorems:

THEOREM 3.16. *Let $R = (A,A,G)$ be an equivalence relation in the set A. Then* (a) *for every element $a \in A$, $a \in a/R$; and* (b) *for any two elements a, b of A, $a/R = b/R$ if and only if $(a,b) \in G$.*

PROOF. To prove (a), we need only observe that since R is reflexive, $(a,a) \in G$ for every $a \in A$.

To prove (b), we first assume that $a/R = b/R$. Then by (a) $b \in b/R = a/R$, and therefore $(a,b) \in G$. Conversely, let us assume that $(a,b) \in G$. Since R is transitive, if $(a,b) \in G$ and $(b,x) \in G$, then $(a,x) \in G$; i.e. if $(a,b) \in G$ and $x \in b/R$, then $x \in a/R$. In other words, if $(a,b) \in G$, then $b/R \subset a/R$. Since R is symmetric, the assumption $(a,b) \in G$ implies that $(b,a) \in G$; interchanging the roles of a and b, we obtain $a/R \subset b/R$. ∎

Any element of an equivalence class is called a *representative of the equivalence class*. Using this terminology, we may restate 3.16 as: a is a representative of a/R; a and b are representatives of one and the same equivalence class if and only if they are R-related.

THEOREM 3.17. *The quotient set A/R of a set A by an equivalence relation R is a partition of the set A.*

PROOF. It is sufficient to show that the conditions (iii) and (iv) following 2.23 are satisfied. The first of these two conditions follows immediately from 3.16 (a). To prove the second condition, let us assume that a and b are elements of A and that there exists an element $x \in a/R \cap b/R$.

Then we have $(a,x)\in G$ and $(b,x)\in G$, where G is the graph of R. From these it follows by the symmetry and the transitivity of R that $(a,b)\in G$. Therefore by 3.16 (b) we have $a/R = b/R$. ∎

It is interesting to notice that the converse of 3.17 also holds.

THEOREM 3.18. *If $\mathfrak{S}$ is a partition of a set A and $G = \{(a,b)\in A \times A: a\in B$ and $b\in B$ for some element $B\in\mathfrak{S}\}$, then the relation $R = (A,A,G)$ is an equivalence relation. Furthermore $A/R = \mathfrak{S}$.*

PROOF. The first statement follows immediately from our definition of the graph G. To prove the second statement, we observe that if B is any element of $\mathfrak{S}$ and a is any element of B, then $B = a/R$. The assertion that $A/R = \mathfrak{S}$ now follows from the definition 3.15 of quotient sets. ∎

Theorems 3.17 and 3.18 show that, in a sense, equivalence relation and partition are 'equivalent' concepts.

To conclude this section, we give some examples to illustrate the concept of equivalence relation.

EXAMPLE 3.19. We consider the set A of all straight lines in the ordinary plane of elementary geometry. Recall that two straight lines ζ and ζ' on the plane are *parallel* iff they coincide or they have no point in common. We define a relation R in A in the following way: for any two straight lines ζ and ζ'

$$\zeta R\zeta' \text{ iff } \zeta \text{ and } \zeta' \text{ are parallel.}$$

Then R is obviously an equivalence relation in A, and the equivalence class ζ/R consists of all straight lines on the plane that are parallel to the given straight line ζ. We can call the equivalence class ζ/R the direction of the straight line ζ and see that any two straight lines ζ and ζ' are parallel iff they have the same direction.

EXAMPLE 3.20. Let Z be the set of all integers and p an arbitrary element of Z. We say that two integers x and y are *congruent modulo p* and write $x \equiv y \bmod p$ iff $x - y$ is a multiple of p. An equivalence relation R is then defined in Z in the following way:

$$xRy \text{ iff } x \equiv y \bmod p.$$

For further properties of this equivalence relation see Exercise 9.

F. Exercises

1. A, B, C and D are sets. Prove that
   (a) $(A \cup B) \times C = (A \times C) \cup (B \times C)$
   (b) $(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$
   (c) $(A \setminus B) \times C = (A \times C) \setminus (B \times C)$
2. What are the elements of $\{a,b,c\} \times \{b,c,d,e\}$? Find all possible relations from $\{a,b,c\}$ to $\{b,c,d,e\}$.
3. A, B and C are sets. The set $A \times B \times C = (A \times B) \times C$ is called the cartesian product of A, B and C. For each $a \in A$, each $b \in B$ and each $c \in C$, the element $(a,b,c) = ((a,b), c)$ of $A \times B \times C$ is called an ordered triple, whose first, second and third coordinates are a, b and c respectively. Show that $(a,b,c) = (a',b',c')$ if and only if $a = a'$, $b = b'$ and $c = c'$.
4. Use the results of Exercise 3 to define *ordered n-tuples* by induction. Formulate and prove a similar necessary and sufficient condition for two ordered n-tuples being equal.
5. Using the method discussed at the beginning of Section **C**, formulate the *ternary relation* of parentage among the population of Hong Kong in terms of ordered triples.
6. Prove Theorem 3.12 of Section **D**.
7. A is a set. Show that if $A \neq \emptyset$ the *null relation* $O = (A, A, \emptyset)$ is not reflexive but is symmetric and transitive.
8. Construct relations R_i $(i = 1, 2, \ldots, 8)$ in such a way that

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
A. reflexive	+	+	+	−	+	−	−	−
B. symmetric	+	+	−	+	−	+	−	−
C. transitive	+	−	+	+	−	−	+	−

   where a + (respectively −) sign under R_i and at the right-hand side of A, B or C indicates that R_i has (respectively does not have) the property A, B or C respectively.

*9. Z is the set of integers and p is a fixed prime number. Show that $R = (Z,Z,G)$, where $G = \{(a,b) \in Z \times Z: a - b$ is a multiple of $p\}$ is an equivalence relation in Z.
   (a) How many elements does Z/R contain?
   (b) Give a necessary and sufficient condition for $a/R = b/R$.
   (c) Show that if $a/R = b/R$ and $a'/R = b'/R$, then
   $$(a \pm a')/R = (b \pm b')/R \text{ and } (aa')/R = (bb')/R.$$
   (d) Show that if a is not divisible by p, then for any $c \in Z$ there exist an integer b such that $(ab)/R = c/R$.

*10. $(X,\mathfrak{T})$ and $(Y,\mathfrak{U})$ are two topological spaces and $\mathfrak{B} = \{Q \in \mathfrak{P}(X \times Y)$: $Q = A \times B$ for some $A \in \mathfrak{T}$ and $B \in \mathfrak{U}\}$. Prove that (a) the intersection of any finite number of elements of $\mathfrak{B}$ is also an element of $\mathfrak{B}$ and (b) the set $\mathfrak{V} = \{Q \in \mathfrak{P}(X \times Y)$: Q *is a union of elements of* $\mathfrak{B}\}$ is a topology for the set $X \times Y$, called the *product topology* of $\mathfrak{T}$ and $\mathfrak{U}$.

*11. Prove that if $(X,\mathfrak{T})$ is a topological space and R is an equivalence relation in the set X, then the set $\mathfrak{Q} = \{\mathfrak{S} \in \mathfrak{P}(X/R)$: $\bigcup_{a/R \in \mathfrak{S}} a/R \in \mathfrak{T}\}$ is a topology for the set X/R, called the *quotient topology* of $\mathfrak{T}$ by R.

CHAPTER 4

MAPPINGS

A. Mappings

Most readers are familiar with the graphical concept of functions. This involves in general a set A of objects called *arguments*, a set B of objects called *values* and an act of associating with each argument in A a unique value in B. In elementary calculus, an expression $y = f(x)$ is used to represent an act of associating with each argument x (a real number) a unique value y (also a real number). Within the framework of set theory, this situation can be conveniently formulated by means of relations.

DEFINITION 4.1. *A mapping from (or of, or on) a set A into (or to) a set B is a relation $f = (A,B,F)$ from A to B such that*

(i) $\mathrm{pr}_1 f = A$,

(ii) *for all $a \in A$ and $b,b' \in B$, $b = b'$ if $(a,b) \in F$ and $(a,b') \in F$.*

Mappings, *maps* and *functions* are synonymous. The symbol $f\colon A \to B$ is used exclusively to denote that f is a mapping of A into B. A, B and $\mathrm{pr}_2 f$ are respectively called the *domain*, the *range* and the *image* of the mapping f. The image of f is usually denoted by $\mathrm{Im}\, f$. For each element a of A, we denote by $f(a)$ the unique element of B such that $(a,f(a)) \in F$. $f(a)$ is called the *value of f at a* or the *image of a under f*. a is called a *pre-image of $f(a)$ under f*. We also say that f maps a into (or onto) $f(a)$.

Most functions of one variable in elementary calculus are mappings of a subset of the set R of all real numbers into the set R. The functions of two variables are then mappings of a subset of $R \times R$ into the set R.

We should note that if $f\colon A \to B$, then for every $a \in A$, there exists a unique $b \in B$ such that $b = f(a)$; but, given $b \in B$, there may not exist any $a \in A$ such that $b = f(a)$, or there may exist one or more $a \in A$ satisfying this condition.

Figure 12 is an example of $f\colon A \to B$ where b_1 is not a value of any element of A under f and b_2 is the value of two elements a_1 and a_3 of A under f.

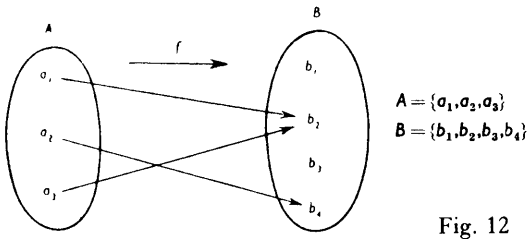


Fig. 12

The following theorem is an immediate consequence of 4.1 and 3.5:

THEOREM 4.2. *Two mappings $f: A \to B$ and $g: C \to D$ are equal if and only if $A = C$, $B = D$ and $f(a) = g(a)$ for all $a \in A$.*

In other words, two mappings are equal if and only if they have the same domain, the same range and the same value at each element of the domain. Furthermore, it follows from 4.2 that we may define a mapping by specifying its domain and range and its value at each element of the domain. However, unlike the most commonly used notation in elementary calculus, here it is essential for us to distinguish the mapping f from $f(a)$, the latter being only the image of the element a under f.

EXAMPLE 4.3. Let A and B be two non-empty sets, and b a fixed element of B. Then the relation (A,B,F), where $F = A \times \{b\}$, is a mapping from A into B whose value at every $a \in A$ is b. We call this mapping the *constant mapping from A into B with the value b.*

EXAMPLE 4.4. Let A be the set of real numbers, and F the subset $\{(a,b) \in A \times A: b = a^2\}$ of $A \times A$. Then the relation (A,A,F) is a mapping from A into A. The graph F of this mapping is a parabola.

EXAMPLE 4.5. Let A be the set of real numbers, B the set of complex numbers and $F = \{(a,b) \in A \times B: b = ai\}$ where $i^2 = -1$. Then (A,B,F) is a mapping from A into B.

EXAMPLE 4.6. If A is the set of all triangles in a Euclidean plane and B the set of real numbers, then we can define a mapping $f: A \to B$ by requiring $f(a) =$ *area of* a for each $a \in A$.

EXAMPLE 4.7. If A is a subset of a set B, then the mapping $f: A \to B$, defined by $f(a) = a$ for each $a \in A$, is called the *inclusion mapping* of A *into* B. The inclusion mapping of A into A is called the *identity mapping of* A, and is denoted by i_A.

DEFINITION 4.8. *Let $f: A \to B$ and $A' \subset A$. Then the mapping $g: A' \to B$, defined by $g(a) = f(a)$ for all $a \in A'$, is called the* restriction *of f to A', and f is called an* extension *of g to A.*

It is customary to write $g = f \mid A'$. If $A \subset B$, then the inclusion mapping of A into B is obviously the restriction to A of the identity mapping of B.

B. Compositions

Given two mappings $f = (A,B,F)$ and $g = (B,C,G)$, then as relations their composition $g \circ f = (A,C,G \circ F)$ is defined. It is easily seen that $g \circ f$ satisfies the conditions (i) and (ii) of 4.1 and is therefore a mapping. We call this mapping the *composition of f and g*. It should be noted that $g \circ f$ is defined if and only if the range of f is equal to the domain of g. Accordingly, if $f\colon A \to B$ and $g\colon B \to C$, then the composition $g \circ f\colon A \to C$ is the mapping such that

$$g \circ f(a) = g(f(a)) \text{ for all elements } a \text{ of } A.$$

It follows from 3.12(b) that the composition of mappings is *associative*; i.e., for any mappings $f\colon A \to B$, $g\colon B \to C$ and $h\colon C \to D$,

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

On the other hand, the composition of mappings is generally not commutative. By this we mean that there exist mappings f and g for which $f \circ g$ and $g \circ f$ are defined, but $f \circ g \neq g \circ f$. In the first place, in order that $f \circ g$ and $g \circ f$ are defined, it is necessary that $f\colon A \to B$ and $g\colon B \to A$. In this case, $f \circ g\colon B \to B$ and $g \circ f\colon A \to A$, and therefore $f \circ g \neq g \circ f$ if $A \neq B$. From the following example we shall see that even if $A = B$, $f \circ g$ may not be equal to $g \circ f$. Let $A = \{a,b\}$ where $a \neq b$, and let $f\colon A \to A$, $g\colon A \to A$ be defined by

$$f(a) = a, \quad f(b) = a;$$
$$g(a) = b, \quad g(b) = b.$$

Then $f \neq g$, and since $f \circ g = f$ and $g \circ f = g$, we have $g \circ f \neq f \circ g$.

The forming of the composition $g \circ f$ suggests an act of associating with each pair of mappings $f\colon A \to B$ and $g\colon B \to C$ a unique mapping $g \circ f\colon A \to C$. This act of association will define a mapping if we can show that for any sets X and Y, all mappings of X into Y form a set.

THEOREM 4.9. *Let X and Y be sets. Then the collection* $\mathrm{Map}(X,Y)$ *of all mappings $f\colon X \to Y$ constitutes a set.*

PROOF. Let $X \times Y$ be the cartesian product of X and Y, and $\mathfrak{P}(X \times Y)$ the power set of $X \times Y$. Then $Z = \{X\} \times \{Y\} \times \mathfrak{P}(X \times Y)$ is again

a set. Obviously $\mathrm{Map}(X,Y)$ is a subset of Z whose elements satisfy the conditions (i) and (ii) of 4.1. Therefore $\mathrm{Map}(X,Y)$ is a set. ∎

Now that there is a set $\mathrm{Map}(X,Y)$ for any two sets X and Y, we naturally want to know under what conditions $\mathrm{Map}(X,Y)$ is non-empty, i.e. under what conditions mappings of X into Y exist. Clearly if neither X nor Y is empty, then $\mathrm{Map}(X,Y)$ is non-empty; for the other cases where X or Y is empty, we have the theorem below.

THEOREM 4.10. *Let X and Y be sets. Then*

(a) $\mathrm{Map}(\emptyset,Y)$ *is a singleton,*

(b) $\mathrm{Map}(X,\emptyset)=\emptyset$ *if* $X\neq\emptyset$.

PROOF. (a) Obviously $(\emptyset,Y,\emptyset)$ is the only mapping of $\emptyset$ into Y; therefore $\mathrm{Map}(\emptyset,Y)$ is a singleton (no matter whether Y is empty or not!).

(b) Let x be an element of X. If f belongs to $\mathrm{Map}(X,\emptyset)$, then by the condition (i) of 4.1, $f(x)\in\emptyset$, which is impossible. Therefore $\mathrm{Map}(X,\emptyset)=\emptyset$. ∎

In particular, when $X=Y$, the set $\mathrm{Map}(X,X)$ is not empty. Among the elements of this set, the identity mapping $i_X\colon X\to X$ plays an important role in the theory of sets.

As is seen in the discussion preceding 4.9 we can define for any sets A, B and C a mapping $\Phi_{ABC}\colon \mathrm{Map}(A,B)\times\mathrm{Map}(B,C)\to\mathrm{Map}(A,C)$ such that $\Phi_{ABC}(f,g)=g\circ f$ for all $f\colon A\to B$ and $g\colon B\to C$. If, in particular, $A=B$, we have $\Phi_{BBC}(i_B,g)=g\circ i_B=g$ for all $g\colon B\to C$; if $B=C$, we have $\Phi_{ABB}(f,i_B)=i_B\circ f=f$ for all $f\colon A\to B$.

C. Direct images and inverse images

A mapping $f\colon A\to B$ associates with each subset X of A a unique subset $f[X]=\{f(a):a\in X\}$ of B, called the *direct image of X under f*. This association defines in a natural way a mapping of $\mathfrak{P}(A)$ into $\mathfrak{P}(B)$. We shall introduce no special notation for this mapping, whose behaviour with respect to the symbols $\subset$, $\cap$ and $\cup$ is as follows: for any subsets X, X' of A,

(a) *if* $X\subset X'$, *then* $f[X]\subset f[X']$,

(b) $f[X\cup X']=f[X]\cup f[X']$, and

(c) $f[X\cap X']\subset f[X]\cap f[X']$.

(a) and (c) are easy consequences of the definition of direct image. To prove (b), we have in the first place $f[X] \subset f[X \cup X']$ and $f[X'] \subset f[X \cup X']$. Hence $f[X] \cup f[X'] \subset f[X \cup X']$. On the other hand, if $b \in f[X \cup X']$, then there exists some $a \in X \cup X'$ such that $b = f(a)$; i.e., there exists some a belonging to X or to X' such that $b = f(a)$. Therefore $b \in f[X]$ or $b \in f[X']$ and consequently $b \in f[X] \cup f[X']$. This proves that $f[X \cup X'] \subset f[X] \cup f[X']$ and hence (b) is true.

Although the equation $f[X \cup X'] = f[X] \cup f[X']$ is true for all $X, X' \subset A$, the equation $f[X \cap X'] = f[X] \cap f[X']$ is not always true. For example, let A and B both be the set of all integers, X the set of all positive integers, X' the set of all negative integers, and $f\colon A \to B$ the mapping which maps all even integers into 0 and all odd integers into 1. Then $X \cap X' = \emptyset$ and so $f[X \cap X'] = \emptyset$, while $f[X] = f[X'] = \{0,1\}$; hence $f[X] \cap f[X'] \neq f[X \cap X']$. Equally unsatisfactory is the relation between direct image and complement, for, in general, there is no inclusion relation between the sets $f[A \setminus X]$ and $B \setminus f[X]$.

For any mapping $f\colon A \to B$ we can also define, in a natural way, a mapping of $\mathfrak{P}(B)$ into $\mathfrak{P}(A)$. The value of this mapping at each subset Y of B is the subset $f^{-1}[Y] = \{x \in A\colon f(x) \in Y\}$ of A, called the *inverse image of Y under f*. For a singleton $\{b\}$ of B, $f^{-1}[\{b\}]$ is usually denoted by $f^{-1}[b]$. This mapping is well-behaved with respect to the symbols $\subset$, $\cap$, $\cup$ and $\setminus$; in fact, for any subsets Y and Y' of B,

(d) *if* $Y \subset Y'$, *then* $f^{-1}[Y] \subset f^{-1}[Y']$,

(e) $f^{-1}[Y \cup Y'] = f^{-1}[Y] \cup f^{-1}[Y']$,

(f) $f^{-1}[Y \cap Y'] = f^{-1}[Y] \cap f^{-1}[Y']$,

(g) $f^{-1}[B \setminus Y] = A \setminus f^{-1}[Y]$.

(d) is an easy consequence of the definition of inverse images. The proof of (e) is similar to that of (f). Let us prove (f) and (g).

Clearly $f^{-1}[Y \cap Y'] \subset f^{-1}[Y]$ and $f^{-1}[Y \cap Y'] \subset f^{-1}[Y']$, therefore, $f^{-1}[Y \cap Y'] \subset f^{-1}[Y] \cap f^{-1}[Y']$. On the other hand, if $a \in f^{-1}[Y] \cap f^{-1}[Y']$, then $f(a) \in Y$ and $f(a) \in Y'$, and hence $f(a) \in Y \cap Y'$. This means that $a \in f^{-1}[Y \cap Y']$. Therefore $f^{-1}[Y \cap Y'] \supset f^{-1}[Y] \cap f^{-1}[Y']$. This proves (f).

For all $a \in A$, $a \in f^{-1}[B \setminus Y]$ if and only if $a \in A$ and $f(a) \notin Y$, i.e., if and only if $a \in A$ and $a \notin f^{-1}[Y]$. Therefore (g) is true.

Finally, applying these mappings one after the other on subsets of A or B, we have the following results: for all subsets X of A and all subsets Y of B,

(h) $X \subset f^{-1}[f[X]]$,

(i) $Y \supset f[f^{-1}[Y]]$.

The proof of these is quite straightforward and is suggested in the figure below.

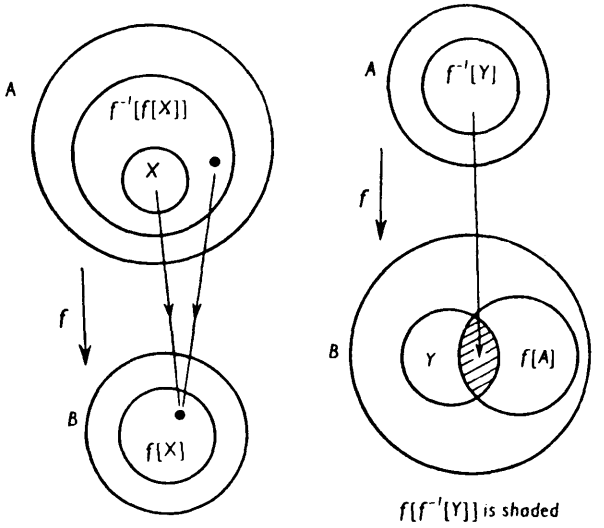


Fig. 13

D. Injective, surjective and bijective mappings

Since any mapping $f\colon A \to B$ is defined as a relation $f = (A,B,F)$ that satisfies (i) $\mathrm{pr}_1 f = A$ and (ii) $b = b'$ for all $(a,b)\in F$ and $(a,b')\in F$, it is sensible to ask under what conditions the mapping f is *invertible*; i.e., under what condition is the inverse relation $f^{-1} = (B,A,F^{-1})$ a mapping $f^{-1}\colon B \to A$? Clearly a necessary and sufficient condition for this is that (iii) $\mathrm{pr}_1 f^{-1} = B$ and (iv) $a = a'$ for all $(b,a)\in F^{-1}$ and $(b,a')\in F^{-1}$. Formulated in terms of f, (iii) and (iv) become

(v) $f[A] = B$ and

(vi) *for any elements a and a' of A, $a = a'$ if $f(a) = f(a')$.*

Some mappings may satisfy one but not the other of these conditions.

EXAMPLES 4.11. (α) The identity mapping i_A clearly satisfies both (v) and (vi). (β) If B is a proper subset of A, then the inclusion mapping $i\colon B \to A$ satisfies (vi) but not (v). (γ) If A is a set consisting of more than one element and B is a singleton, then the constant mapping $f\colon A \to B$, satisfies (v) but not (vi).

These possibilities lead us to formulate the following definition:

DEFINITION 4.12. *A mapping* $f: A \to B$ *is said to be*

(a) surjective *if and only if the condition* (v) *is satisfied,*

(b) injective *if and only if the condition* (vi) *is satisfied,*

(c) bijective *if and only if it is both surjective and injective.*

Surjective mappings are also called *onto* mappings, injective mappings are also called *one-to-one* mappings, and bijective mappings are also called *one-to-one correspondences*. From the above discussion, it can be seen that a mapping f is invertible if and only if it is bijective; in this case, the mapping f^{-1} is called the *inverse mapping* of f. From 3.12(a), we have $(f^{-1})^{-1} = f$; therefore, f^{-1} is also a bijective mapping if f is bijective.

The composition, when defined, of two surjective (respectively, injective, bijective) mappings is clearly a surjective (respectively, injective, bijective) mapping. In particular, it follows from 3.12(c) that if $f: A \to B$ and $g: B \to C$ are bijective mappings, then $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Conditions (v) and (vi) are formulated in terms of elements of A and elements of B. In the following theorems, surjective and injective mappings are characterized by conditions formulated in terms of the more fundamental concepts of mappings and compositions alone.

THEOREM 4.13. *Let* $f: A \to B$ *be a mapping. Then* f *is surjective if and only if the following condition is satisfied*:

(vii) *For any set* X, *two mappings* $\varphi: B \to X$ *and* $\psi: B \to X$ *are equal if* $\varphi \circ f = \psi \circ f$.

PROOF. Assume that f is surjective, and that $\varphi \circ f = \psi \circ f$ is satisfied for mappings $\varphi: B \to X$ and $\psi: B \to X$. If $B = \emptyset$, then $\mathrm{Map}(B, X)$ is a singleton by 4.10(a); therefore (vii) is satisfied. If $B \neq \emptyset$, then since $f: A \to B$ is surjective, for every $b \in B$, there exists $a \in A$ such that $f(a) = b$. Therefore, for every $b \in B$, $\varphi(b) = \varphi(f(a)) = \varphi \circ f(a) = \psi \circ f(a) = \psi(f(a)) = \psi(b)$; hence $\varphi = \psi$, showing that the condition (vii) is satisfied.

Conversely we assume that (vii) holds for $f: A \to B$. If $A = \emptyset$, and X is any set, then $\mathrm{Map}(A, X)$ is a singleton. Therefore it follows from (vii) that $\mathrm{Map}(B, X)$ is a singleton for any set X. Hence $B = \emptyset$, and $f: \emptyset \to \emptyset$ is surjective. If $A \neq \emptyset$, then $f[A] \neq \emptyset$. Assume that $f[A] \neq B$ and let $b_0 \in B \setminus f[A]$. Then a mapping $\psi: B \to B$ exists such that $\psi(b_0) \in f[A]$ and $\psi(b) = b$ for every $b \in B \setminus \{b_0\}$. Now $i_B \circ f = \psi \circ f$, but $i_B \neq \psi$, contradicting (vii). Therefore it follows that the assumption $f[A] \neq B$ is wrong, and hence f is surjective.

There is another, similar theorem:

THEOREM 4.14. *Let f: $A \to B$ be a mapping. Then f is injective if and only if the following condition is satisfied.*

(viii) *For any set X, two mappings φ: $X \to A$ and ψ: $X \to A$ are equal if $f \circ \varphi = f \circ \psi$.*

Conditions (vii) and (viii) express the fact that a mapping f is surjective (respectively injective) if and only if it is *right* (respectively *left*) *cancellable*. Such characterizations of surjective mappings and injective mappings are very useful in modern abstract algebra.

For the characterization of bijective mappings, there is the following theorem:

THEOREM 4.15. *Let f: $A \to B$ be a mapping. Then f is bijective if and only if there is a mapping g: $B \to A$ such that $f \circ g = i_B$ and $g \circ f = i_A$. In this case, $g = f^{-1}$ and $f = g^{-1}$.*

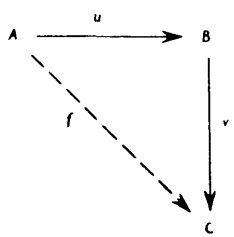
PROOF. If f is bijective, then its inverse mapping f^{-1}: $B \to A$ obviously satisfies the conditions $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$.

Conversely, if g: $B \to A$ exists such that $g \circ f = i_A$ and $f \circ g = i_B$, then for any φ: $B \to X$ and ψ: $B \to X$ such that $\varphi \circ f = \psi \circ f$, we have $(\varphi \circ f) \circ g = (\psi \circ f) \circ g$; i.e., $\varphi \circ (f \circ g) = \psi \circ (f \circ g)$. But, by our assumption, this is $\varphi \circ i_B = \psi \circ i_B$. Therefore $\varphi = \psi$, and hence by 4.13, f is surjective. Similarly, we can prove that f is injective.

Since $f^{-1} \circ f = i_A = g \circ f$, it follows from 4.13 that $f^{-1} = g$. Similarly $f = g^{-1}$. ∎

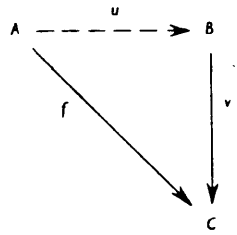
E. Factorizations

Let u: $A \to B$ and v: $B \to C$ be mappings. Then we get a mapping $f = v \circ u$: $A \to C$ the composition of u and v. In other words, we can fill into the following diagram a unique mapping f, so as to make it commutative.

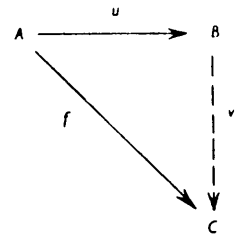


Factorization is the converse of composition. Let us look into the factorization of mappings by investigating problems of the following types:

1. Given $f: A \to C$ and $v: B \to C$, does there exist $u: A \to B$ such that $f = vou$? In other words, can we fill into the diagram some u so as to make it commutative?



2. Given $f: A \to C$ and $u: A \to B$, does there exist $v: B \to C$ so that $f = vou$? In other words, can we fill into the diagram some v so as to make it commutative?



In general these problems have no solutions. Take for instance a non-constant mapping f and a constant mapping v in which case problem 1 has no solution. Therefore we shall only consider special cases of these problems in which restrictions are imposed on u or v.

Let $v: B \to C$ be an *injective* mapping. We wish to find the condition that f has to satisfy so that problem 1 can be solved. Assuming that $u: A \to B$ exists such that $f = vou$, then u is unique by 4.14 and also

(*) $\quad \operatorname{Im} f \subset \operatorname{Im} v.$

Conversely, if the condition (*) is satisfied by f, then for each $a \in A$, $f(a) \in \operatorname{Im} v$. Therefore a unique $b \in B$ exists such that $v(b) = f(a)$. A mapping $u: A \to B$ is then determined by the following condition:

$$\text{for all } a \in A,\ u(a) = b \text{ where } v(b) = f(a).$$

Obviously $f = v \circ u$. Therefore the condition (*) is necessary and sufficient for the problem to be solvable.

Now consider problem 2, which has many applications in mathematics. We now assume that, dually, $u\colon A \to B$ is *surjective*. If $v\colon B \to C$ exists such that $f = v \circ u$, then v is unique by 4.13 and we have

(**) for all $a, a' \in A$, $f(a) = f(a')$ if $u(a) = u(a')$.

We shall show that the condition (**) is also sufficient. Indeed, for each $b \in B$, since u is surjective, there exists $a \in A$ such that $u(a) = b$. If a' is any other element of A such that $u(a') = b$, then by (**) $f(a) = f(a')$; therefore a mapping $v\colon B \to C$ is uniquely determined by the following condition:

for every $b \in B$, $v(b) = f(a)$ where a is any element of A for which $u(a) = b$.

Obviously this mapping v satisfies our requirement that $f = v \circ u$. Hence the condition (**) is necessary and sufficient for problem 2 to be solvable. We state our result as a theorem.

THEOREM 4.16. *Let $u\colon A \to B$ be a surjective mapping. Then for each mapping $f\colon A \to C$, a mapping $v\colon B \to C$ exists such that $f = v \circ u$ if and only if the condition (**) above is satisfied. In this case, the mapping v is unique.*

To give an application of this theorem, let A be a set and R an equivalence relation in A. Then the mapping $u\colon A \to A/R$ defined by

$$u(a) = a/R \quad \textit{for all } a \in A,$$

is called the *natural surjection* of A onto A/R. A mapping $f\colon A \to C$ is said to be *compatible* with the equivalence relation R if and only if (**) is satisfied. Then we have the following corollary:

COROLLARY 4.17. *Let A be a set, R an equivalence relation in A and $u\colon A \to A/R$ the natural surjection. Then for each mapping $f\colon A \to C$, a mapping $v\colon A/R \to C$ exists such that $f = v \circ u$ if and only if f is compatible with R. In this case, the mapping v is unique.*

F. Exercises

1. Let $f: A \to B$ be a bijective mapping, x an element of A and y an element of B. Prove that there exists a bijective mapping $g: A \to B$ so that $g(x) = y$.

2. Z is the set of all integers. Give examples of mappings $f: Z \to Z$ which are

   (a) injective but not surjective,

   (b) surjective but not injective.

3. For three sets A, B and C we define $A \times B \times C$ as the set

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

   Prove that the mapping $f: A \times B \times C \to A \times (B \times C)$ defined by $f(a,b,c) = (a,(b,c))$ is bijective.

4. For mappings $f: A \to C$, $g: B \to D$, the mapping $f \times g: A \times B \to C \times D$ is defined by $f \times g(a,b) = (f(a),g(b))$. Prove that $f \times g$ is

   (a) surjective if and only if f and g are surjective,

   (b) injective if and only if f and g are injective.

5. Let $f: A \to B$. Prove that for any two subsets X, X' of A and for any subset Y of B,

   (a) $f[X \setminus X'] \supset f[X] \setminus f[X']$,

   (b) $f[X \cap f^{-1}[Y]] = f[X] \cap Y$.

6. Prove (e), (h), (i) of Section **C**.

7. Prove 4.14.

8. Let A be a set and X a subset of A. Then the mapping $\gamma_X: A \to \{0,1\}$ such that for each $a \in A$,

$$\gamma_X(a) = 0 \text{ if } a \notin X$$
$$\gamma_X(a) = 1 \text{ if } a \in X$$

   is called the *characteristic function* of X. Show that the set of all characteristic functions of subsets of A is in one-to-one correspondence with the power set $\mathfrak{P}(A)$ of A.

9. Let $f: A \to B$ and $g: B \to C$ be mappings. Prove that

   (a) if $g \circ f$ is surjective, then g is surjective,

   (b) if $g \circ f$ is injective, then f is injective.

   Give counter-examples to show that the converse of each of the statements (a) and (b) is not true.

10. Let $f: A \to B$ be a mapping.

    (a) Show that for any set X, the mapping $\Phi: \mathrm{Map}(X,A) \to \mathrm{Map}(X,B)$ such that $\Phi(\varphi) = f \circ \varphi$ for all $\varphi: X \to A$ is injective if and only if f is injective.

(b) Show that for any set Y, the mapping Ψ: $\mathrm{Map}(B,Y) \to \mathrm{Map}(A,Y)$ such that $\Psi(\psi) = \psi \circ f$ for all ψ: $B \to Y$ is injective if and only if f is surjective.

11. Let f: $A \to B$ and g: $B \to C$ be mappings. Prove that for all subsets X of A and for all subsets Z of C,
    (a) $g \circ f[X] = g[f[X]]$,
    (b) $(g \circ f)^{-1}[Z] = f^{-1}[g^{-1}[Z]]$.

12. Prove that a mapping f: $A \to B$ is surjective if and only if $Y = f[f^{-1}[Y]]$ for all $Y \subset B$.

*13. Prove that, for a mapping f: $A \to B$, the following conditions are equivalent:
    (a) f is injective;
    (b) $X = f^{-1}[f[X]]$ for all $X \subset A$;
    (c) $f[X \cap X'] = f[X] \cap f[X']$ for all X, $X' \subset A$.

14. Let f be a mapping from a set A into a set C, and g a mapping from the set A into a set B. Prove that f admits a factorization $f = h \circ g$, where h: $B \to C$, if and only if, for all a, $a' \in A$, $g(a) = g(a')$ implies $f(a) = f(a')$. Also show that h is unique if g is surjective.

15. Let A and B be sets, and R and S equivalence relations in A and B respectively. Show that if f: $A \to B$ is compatible with R and S, (i.e. $f(a) S f(a')$ if aRa' for all a and a' of A) then there is a unique mapping h: $A/R \to B/S$ such that $v \circ f = h \circ u$ where u: $A \to A/R$ and v: $B \to B/S$ are the natural surjections.

*16. A mapping f of a topological space X into a topological space Y is said to be *continuous* if and only if $f^{-1}[B]$ is open for every open set B of Y. Prove that for all mappings f: $X \to Y$, the following conditions are equivalent:
    (a) f is continuous;
    (b) for every $x \in X$, and for every neighbourhood V of $f(x)$, $f^{-1}[V]$ is a neighbourhood of x;
    (c) $f^{-1}[B]$ is closed for every closed set B of Y.

*17. Let f be a mapping from a topological space $(X, \mathfrak{T})$ onto a set Y. Prove that $\mathfrak{T}' = \{B \in \mathfrak{P}(Y)\colon f^{-1}[B] \in \mathfrak{T}\}$ is a topology for Y for which f is continuous.

PART II

CHAPTER 5

FAMILIES

A. Families

In mathematics, for the sake of convenient formulation and easy reference, we very often introduce subscripts, superscripts and the like to index the objects (e.g. points, lines, indeterminates, etc.) of our discussion. The indices are usually numbers or letters. More generally, given two sets A and I, the indexing of certain elements of A by elements (*indices*) taken from I naturally involves the concept of mapping; after such a process is carried out, the indexed elements of A, together with their indices, will receive more attention than the process itself. To handle this kind of situation efficiently, we introduce the concept of *family*.

DEFINITION 5.1. *Let A be a set.* A family *of elements of A is an ordered triple (F, I, f) where F is a subset of A, I a set and $f : I \to F$ a surjective mapping.*

Hence a family (F, I, f) of elements of a set A is equal to a family (G, J, g) of elements of a set B, if and only if (a) $I = J$ and (b) $f(i) = g(i)$ for each $i \in I$. These two conditions are obviously necessary; they are also sufficient, for the equality $F = G$ follows from these conditions and the assumption that both mappings f and g are surjective.

Let (F, I, f) be a family of elements of a set A. If we denote for each $i \in I$ the element $f(i)$ of A by x_i, then we can represent the family (F, I, f) in the more convenient form $(x_i)_{i \in I}$. The set I is called the *index set* of the family $(x_i)_{i \in I}$ and the element x_i of A is called the *term of the index i* of the family or simply the *i-th term* of the family. Under this notation, a necessary and sufficient condition for the equality of two families is given in the following theorem:

THEOREM 5.2. *Let $(x_i)_{i \in I}$ be a family of elements of a set A and $(y_j)_{j \in J}$ be a family of elements of a set B. Then these two families are equal if and only if* (a) $I = J$, (b) *for each $i \in I$, $x_i = y_i$.*

Given a family $(F, I, f) = (x_i)_{i \in I}$ of elements of a set A, the set F is sometimes called *the set of all terms* of the family $(x_i)_{i \in I}$. It is clear from 5.2 that two distinct families may very well have the same set of terms. Therefore we should be careful to distinguish the set of all terms of a family $(x_i)_{i \in I}$ from the family $(x_i)_{i \in I}$ itself.

Conversely, given a set A, the ordered triple (A, A, i_A), where $i_A: A \to A$ is the identity mapping of the set A, is a family of elements of A. This family is denoted by $(x_x)_{x \in A}$ where $x_x = x$ for all $x \in A$. The set of all terms of this family is clearly the set A itself.

We define subfamilies and the empty family, corresponding to the concepts of subsets and the empty set, as follows:

DEFINITION 5.3. *Let $(x_i)_{i \in I}$ be a family of elements of a set A. A family $(y_j)_{j \in J}$ of elements of a set B is a* subfamily *of the family $(x_i)_{i \in I}$ if and only if* (a) *J is subset of I, and* (b) *$y_j = x_j$ for each $j \in J$.*

As examples of subfamilies of a family $(x_i)_{i \in I}$ of elements of a set A, the family $(x_i)_{i \in I}$ itself deserves mention and also the *empty family* $(x_i)_{i \in \emptyset}$ of elements of A. The set of all terms of the empty family $(x_i)_{i \in \emptyset}$ of elements of A is obviously the empty set $\emptyset$.

Of particular interest to us are *families of sets* and *families of subsets of a set*. These are defined as follows:

DEFINITION 5.4. *Let $(A_i)_{i \in I}$ be a family of elements of a set A. If the elements of the set A are themselves sets, then $(A_i)_{i \in I}$ is called a* family of sets. *In particular, if $A = \mathfrak{P}(B)$, i.e. if A is the power set of a set B, then $(A_i)_{i \in I}$ is called a* family of subsets *of the set B.*

B. Intersections and unions

In this section we shall investigate the possibility of generalizing the concepts of *intersection* and *union* by using the results of the last section.

Let A be a set and $(A_i)_{i \in I}$ a family of subsets of A. Then the set $\mathfrak{S}$ of all terms of this family is clearly a subset of the power set $\mathfrak{P}(A)$; the intersection and the union of members of $\mathfrak{S}$ are defined respectively as the intersection $\bigcap_{i \in I} A_i$ and the union $\bigcup_{i \in I} A_i$ of the family $(A_i)_{i \in I}$ of subsets of A. It is easily seen that

$$\bigcap_{i \in I} A_i = \{x \in A : x \in A_i \text{ for all } i \in I\}$$

and that

$$\bigcup_{i \in I} A_i = \{x \in A : x \in A_i \text{ for some } i \in I\}.$$

Clearly $\bigcap_{i \in I} A_i$ is the largest subset (in the sense of inclusion) of A included in each A_i, and $\bigcap_{i \in I} A_i$ is the smallest subset (in the sense of

inclusion) of A including each A_i. In particular, for the empty family $(A_i)_{i\in\emptyset}$ of subsets of A, we have

$$\bigcap_{i\in\emptyset} A_i = A \quad \text{and} \quad \bigcup_{i\in\emptyset} A_i = \emptyset.$$

Notice that the equality $\bigcap_{i\in\emptyset} A_i = A$ indicates the dependence of the intersection of the empty family on the set in which the intersection is formed, in this case A. To illustrate this state of affairs, let us consider an example. Let A and B be two different sets. If f is the only mapping $f : \emptyset \to \emptyset$, then the family $(\emptyset, \emptyset, f)$ is the empty family of subsets of the set A and at the same time it is the empty family of subsets of the set B. Considered as a family of subsets of the set A, the intersection of $(\emptyset, \emptyset, f)$ is A; whereas considered as a family of subsets of the set B, the intersection of $(\emptyset, \emptyset, f)$ is B.

Let us now consider the intersection of an arbitrary non-empty family $(A_i)_{i\in I}$ of sets. Since $I \neq \emptyset$, we can consider the term A_j of the family for an arbitrary $j \in I$. A_j is a set, and applying the axiom of specification, we obtain a subset $\{x \in A_j : x \in A_i \textit{ for all } i \in I\}$ of the set A_j. This set can also be written as

$$\{x : x \in A_i \textit{ for all } i \in I\}$$

which is clearly the set of all common elements of the sets A_i. Therefore the following definition is justified:

DEFINITION 5.5. *Let $(A_i)_{i\in I}$ be a family of sets. If $I \neq \emptyset$, then the* intersection *of the family $(A_i)_{i\in I}$ is the set*

$$\bigcap_{i\in I} A_i = \{x : x \in A_i \textit{ for all } i \in I\}.$$

Note that in 5.5 we have only defined the intersection of a non-empty family of sets. Now what is the reason for excluding the case where the index set I is empty? The statement

$$x \in A_i \textit{ for all } i \in I,$$

used for specifying the elements of the intersection, is equivalent to the conditional

$$i \in I \to x \in A_i.$$

In the case where $I = \emptyset$, the above conditional and hence also the original statement are obviously true statements for all objects x. Therefore $\{x : x \in A_i \textit{ for all } i \in \emptyset\}$ fails to constitute a set. For this reason we give no definition of the intersection of the empty family of sets.

To define the union of an arbitrary family $(A_i)_{i\in I}$ of sets as the set of all elements of the sets A_i, it is necessary to have a certain set U

which includes each A_i as a subset. This is similar to the situation in Section **2G.** The axiom of inclusion introduced to overcome the similar difficulty there is too weak to be applicable here. Therefore we need a stronger axiom for constructing sets:

AXIOM OF UNION. *Given any set* $\mathfrak{S}$ *whose elements are sets, there exists a set* U *which includes all the elements of* $\mathfrak{S}$ *as subsets.*

The axiom of inclusion is clearly a special case of the axiom of union, where $\mathfrak{S} = \{A, B\}$. With the axiom of union included in our system of axioms, the axiom of inclusion becomes redundant.

Let $(A_i)_{i\in I}$ be a family of sets. Then it follows that the set $\mathfrak{S}$ of all the terms of this family is a set of sets, and hence the axiom of union is applicable. Let U be a set such the $U \supset A_i$ for each $i \in I$. Using the axiom of specification, we obtain a subset $\{x \in U\colon x \in A_i \textit{ for some } i \in I\}$ of U. This is clearly the set of all elements that belong to at least one of the sets A_i. Furthermore this set, which can be shown by a method similar to that used in Section **2G** to be independent of the choice of any particular U, can be written as $\{x : x \in A_i \textit{ for some } i \in I\}$. We have the following definition:

DEFINITION 5.6. *Let* $(A_i)_{i\in I}$ *be a family of sets. The* union *of this family is the set*

$$\bigcup_{i\in I} A_i = \{x : x \in A_i \textit{ for some } i \in I\}.$$

The union of the empty family $(A_i)_{i\in\emptyset}$ is easily seen to be the empty set. Thus,

$$\bigcup_{i\in\emptyset} A_i = \emptyset.$$

Finally we have no difficulty in seeing that the intersection of a family $(A_i)_{i\in I}$ of subsets of a set A is the same (as long as $I \neq \emptyset$) as the intersection of the family $(A_i)_{i\in I}$ of sets, and the union of a family $(A_i)_{i\in I}$ of subsets of a set A is the same as the union of the family $(A_i)_{i\in I}$ of sets. Moreover, for any two sets A and B, the intersection $A \cap B$ and the union $A \cup B$ defined in Chapter 2 are respectively the same as $\bigcap_{i\in I} C_i$ and $\bigcup_{i\in I} C_i$ where $I = \{0, 1\}$ and $C_0 = A$, $C_1 = B$.

C. Cartesian products

In Section **3B,** the cartesian product $A \times B$ of any two sets A and B is defined to be the set of all ordered pairs (a, b) where $a \in A$ and $b \in B$. Although we have also introduced earlier in Section **3A** the concepts of ordered triples and ordered n-tuples, they are nevertheless

not the appropriate tools for constructing the cartesian product of an arbitrary family of sets. Let us first compare the concept of ordered pairs with that of families. For this purpose, let X be a set and $I = \{0,1\}$. (Here it is immaterial that the numbers 0 and 1 are used; in fact any set consisting of exactly two elements, for example $\{\emptyset, \{\emptyset\}\}$, will do.) Then each ordered pair (a, b) of elements of X determines a unique family $(x_i)_{i\in I}$ of elements of X in such away that $a = x_0$ and $b = x_1$. Conversely, each family $(x_i)_{i\in I}$ determines a unique ordered pair (a, b) in such away that $x_0 = a$ and $x_1 = b$. Using this correspondence, we may hope to find a set consisting of families $(x_i)_{i\in I}$ of elements of $A \cup B$, corresponding to the cartesian product $A \times B$; and, generalizing it, we give a satisfactory definition of the cartesian product of a family of sets.

Let A and B be sets, $X = A \cup B$ and $I = \{0, 1\}$. Then $(X_i)_{i\in I}$ where $X_0 = A$ and $X_1 = B$ is a family of two sets. Consider the collection P of all those families $x = (x_i)_{i\in I}$ where $x_0 \in X_0 = A$ and $x_1 \in X_1 = B$. Then P is a set (see Theorem 5.7 below) and the mapping $\Phi : P \to A \times B$, defined by $\Phi(x) = (x_0, x_1)$ for each $x \in P$, is bijective. This means that the sets P and $A \times B$ are in one-to-one correspondence. For this reason, though the sets P and $A \times B$ are not equal, their difference can be regarded as merely a matter of notation.

The generalization is now straightforward. Let $(A_i)_{i\in I}$ be a family of sets and $X = \bigcup_{i\in I} A_i$. We now have to find out whether the collection P of all families $(x_i)_{i\in I}$ of elements of X where $x_i \in A_i$ for all $i \in I$ constitutes a set, before we can define P as the cartesian product of the family $(A_i)_{i\in I}$. That this is so will be obvious once the following theorem has been established.

THEOREM 5.7. *Let X and I be two sets. Then the collection of all families (F, I, f) of elements of X with the fixed index set I constitutes a set.*

PROOF. For each subset F of X, we consider the set

$$M_F = \{F\} \times \{I\} \times \mathrm{Map}(I,F).$$

Then $(M_F)_{F\in\mathfrak{P}(X)}$ is a family of sets and hence the union $U = \bigcup_{F\in\mathfrak{P}(X)} M_F$ is a set. Obviously $(F,I,f) \in M_F$. Therefore the families of elements of X with a fixed index set I form a subset of U. ∎

Now the following definition is justified:

DEFINITION 5.8. *Let $(A_i)_{i\in I}$ be a family of sets. The* cartesian product *of this family is the set $\prod_{i\in I} A_i$ of all those families $x = (x_i)_{i\in I}$ of elements of $\bigcup_{i\in I} A_i$ such that $x_i \in A_i$ for each $i \in I$.*

Let U be a set such that A_i is a subset of U for all $i \in I$. Then the set of all families $x = (x_i)_{i \in I}$ of elements of U where $x_i \in A_i$ is, by 5.2, equal to the cartesian product defined in 5.8. Therefore we may use any set U with the above property to define the cartesian product.

Using a similar argument, we obtain the following theorem:

THEOREM 5.9. *Let $(A_i)_{i \in I}$ and $(B_i)_{i \in I}$ be two families of sets where B_i is a subset of A_i for all $i \in I$. Then $\prod_{i \in I} B_i$ is a subset of $\prod_{i \in I} A_i$.*

REMARKS. Some authors define a family of elements of a set A as a mapping $f: I \to A$ and the cartesian product of a family $(A_i)_{i \in I}$ as the set of all mappings $f: I \to \bigcup_{i \in I} A_i$ such that $f(i) \in A_i$ for all $i \in I$. In that case 5.2 does not hold and consequently only a weaker form of 5.9 is true: if $(A_i)_{i \in I}$ and $(B_i)_{i \in I}$ are families of sets and B_i is a subset of A_i for all $i \in I$, then the cartesian product of the family $(B_i)_{i \in I}$ is in one-to-one correspondence with a subset of the cartesian product of the family $(A_i)_{i \in I}$.

D. The axiom of choice

Now that we have obtained for any family of sets a set called the cartesian product of the family, it is desirable to know under what condition the cartesian product is a non-empty set. Here again our knowledge is limited to a few special cases for which the required condition can be found.

Corresponding to the two statements of 4.10, we have the following two special cases:

THEOREM 5.10. *Let $(A_i)_{i \in I}$ be a family of sets and P the cartesian product of the family. Then the following statements hold:*

(a) *If $I = \emptyset$, then P is a singleton.*

(b) *If $I \neq \emptyset$ and $A_j = \emptyset$ for some $j \in I$, then $P = \emptyset$.*

PROOF. (a) P contains only the empty family as its element.
(b) Assume that P is non-empty. Then there would be an element $x = (x_i)_{i \in I}$ in P, and for this element we would have $x_j \in A_j$. Since A_j is empty, this is a contradiction. ∎

In the case where I is non-empty and A_i is non-empty for all indices i of I, we get the following results:

(c) If the index set I is a singleton, then it is easily seen that the cartesian product and the only (non-empty) set of the family stand in one-to-one correspondence. Therefore the cartesian product is non-empty.

(d) If the index set I consists of exactly two elements, then, as proved earlier, our cartesian product in the sense of 5.8 and the set $A \times B$ stand in one-to-one correspondence. Therefore we again have a non-empty cartesian product.

(e) In a slightly more general setting, we get the following result: the cartesian product of a finite family of non-empty sets is non-empty. (For definition, see Chapter 7.)

However, for a general family of non-empty sets, we have no more information on its cartesian product. Since there are 'larger' families than those mentioned above, it would seem very strange indeed if the cartesian product of a 'larger' family of non-empty sets could be empty while that of a 'smaller' one was not. Therefore it seems to us reasonable to accept the following axiom:

AXIOM OF CHOICE. *The cartesian product of a non-empty family of non-empty sets is non-empty.*

REMARKS. The axiom of choice has been a matter of dispute among mathematicians for a long time. We do not intend to go into the history or the controversy here. However, it has been proved that the axiom of choice is consistent with the usual axioms of set theory under the assumption that these usual axioms are consistent with each other. For the interested reader we refer to the excellent exposition by A. A. FRAENKEL, *Abstract Set Theory* and the book by K. GÖDEL mentioned at the end of Section **9 E.**

For some applications of this axiom an equivalent formulation given below is useful.

THEOREM 5.11. *The axiom of choice is equivalent to the following statement: for any non-empty set A, there exists a choice function φ of A, i.e. a mapping $\varphi\colon \mathfrak{P}(A)\setminus\{\emptyset\} \to A$ exists so that $\varphi(B) \in B$ for all non-empty subsets B of A.*

PROOF. Assume that the axiom of choice holds good and let A be a non-empty set. Putting $X = \mathfrak{P}(A)\setminus\{\emptyset\}$, we have a non-empty family $(B_B)_{B\in X}$ of non-empty subsets of A where $B_B = B$. Then each element of the non-empty cartesian product $\prod_{B\in X} B_B$ yields a choice function of A.

Conversely, assume that a choice function exists for any non-empty set and let $(A_i)_{i\in I}$ be a non-empty family of non-empty sets. Then the union A of this family is non-empty. If φ is a choice function of A, then from the mapping $f\colon I \to A$, defined by $f(i) = \varphi(A_i)$ for each $i \in I$, we get an element of the cartesian product of the family. ∎

E. Projections

Consider a non-empty family $(A_i)_{i\in I}$ of non-empty sets. Then, by the axiom of choice, the cartesian product P of this family is non-empty, and therefore we have non-trivial mappings of P into A_i. The most important of these mappings are the *projections* of P.

DEFINITION 5.12. *Let $(A_i)_{i\in I}$ be a non-empty family of non-empty sets and P the cartesian product of the family. For each $j \in I$, the mapping $\mathrm{pr}_j : P \to A_j$, where $\mathrm{pr}_j(x) = x_j$ for all $x = (x_i)_{i\in I}$, is called the* projection of the index j *or simply the* j-th projection *of the cartesian product.*

We shall see that the projections are surjective mappings. Let j be a fixed index of I and x_j an element of A_j. Consider the family $(B_i)_{i\in I}$ where $B_j = \{x_j\}$ and $B_i = A_i$ for all $i \neq j$. This is clearly a non-empty family of non-empty sets; therefore its cartesian product Q is non-empty and, by 5.9, is a subset of P. Now for each $x \in Q$, we have $\mathrm{pr}_j(x) = x_j$, and therefore pr_j is surjective.

We have seen in Section **4E** that the natural surjection u of a set A onto a quotient set A/R can be used to factorize certain mappings of A into a set B. The projections here serve a similar purpose.

THEOREM 5.13. *Let $(A_i)_{i\in I}$ be a family of sets and P the cartesian product of the family. Then for any set X and for any family $(f_i)_{i\in I}$ of mappings where $f_i : X \to A_i$ for all $i \in I$, there is a unique mapping $f : X \to P$ such that $f_i = \mathrm{pr}_i \circ f$ for all $i \in I$.*

PROOF. Let f and g be mappings and $f_i = \mathrm{pr}_i \circ f$ and $f_i = \mathrm{pr}_i \circ g$ for each $i \in I$. Then for each $x \in X$ we have for all $i \in I$

$$f(x)_i = \mathrm{pr}_i(f(x)) = \mathrm{pr}_i(g(x)) = g(x)_i.$$

That means $f = g$ and hence the uniqueness of f. Now the mapping $f : X \to P$ defined by $f(x) = (f_i(x))_{i\in I}$ clearly satisfies the requirement of the theorem. ∎

F. Exercises

1. Let $(A_i)_{i\in I}$ be a family of subsets of a set A. Prove the generalized De Morgan's laws:
   (a) $A\setminus(\bigcup_{i\in I}A_i) = \bigcap_{i\in I}(A\setminus A_i)$
   (b) $A\setminus(\bigcap_{i\in I}A_i) = \bigcup_{i\in I}(A\setminus A_i)$

2. Let A and B be sets, $f: A\to B$ a mapping, $(A_i)_{i\in I}$ a family of subsets of A, and $(B_i)_{i\in I}$ a family of subsets of B. Prove that:
   (a) $f[\bigcup_{i\in I}A_i] = \bigcup_{i\in I}f[A_i]$
   (b) $f[\bigcap_{i\in J}A_i] \subset \bigcap_{i\in I}f[A_i]$
   (c) $f^{-1}[\bigcup_{i\in I}A_i] = \bigcup_{i\in I}f^{-1}[A_i]$
   (d) $f^{-1}[\bigcap_{i\in I}A_i] = \bigcap_{i\in I}f^{-1}[A_i]$

3. A family of the form $(A_{(i,j)})_{(i,j)\in I\times J}$ is sometimes called a ***double family*** and is sometimes represented by using '*double indices*' as $(A_{ij})_{i\in I;\,j\in J}$. Prove that if $(A_{ij})_{i\in I;\,j\in J}$ is a family of sets, then
   (a) $\bigcup_{\substack{i\in I\\ i\in J}}A_{ij} = \bigcup_{i\in J}(\bigcup_{i\in I}A_{ij}) = \bigcup_{i\in I}(\bigcup_{i\in J}A_{ij})$
   (b) for $I\neq\emptyset$ and $J\neq\emptyset$
   $\bigcap_{\substack{i\in I\\ i\in J}}A_{ij} = \bigcap_{i\in J}(\bigcap_{i\in I}A_{ij}) = \bigcap_{i\in I}(\bigcap_{i\in J}A_{ij})$
   If $(A_{ij})_{i\in I;\,j\in J}$ is a family of subsets, then (b) holds without restriction on the index sets.

4. Let $(A_i)_{i\in I}$ be a family of sets and $\varphi: K\to I$ a surjective mapping. Prove the following commutative laws:
   (a) $\bigcup_{k\in K}A_{\varphi(k)} = \bigcup_{i\in I}A_i$
   (b) for $K\neq\emptyset$ and $I\neq\emptyset$
   $\bigcap_{k\in K}A_{\varphi(k)} = \bigcap_{i\in I}A_i$
   In the case of a family of subsets, (b) holds without restriction.

5. Let $(A_j)_{j\in J}$ be a family of sets and $J = \bigcup_{k\in K}I_k$. Prove the following associative laws:
   (a) $\bigcup_{j\in J}A_j = \bigcup_{k\in K}(\bigcup_{i\in I_k}A_i)$
   (b) for $K\neq\emptyset$ and $I_k\neq\emptyset$ for each $k\in K$
   $\bigcap_{j\in J}A_j = \bigcap_{k\in K}(\bigcap_{i\in I_k}A_i)$
   If $(A_i)_{i\in I_k}$ is a family of subsets of the same set A for each $k\in K$, then (b) holds without restriction.

6. Formulate and prove the distributive laws of unions and intersections.

7. Let $(A_i)_{i\in I}$ be a family of sets. Prove that $\prod_{i\in I} A_i = \emptyset$ if and only if $I \neq \emptyset$, and $A_i = \emptyset$ for some $i\in I$.

8. Prove the following statement: if $(A_i)_{i\in I}$ and $(B_i)_{i\in I}$ are families of sets such that $\prod_{i\in I} B_i \subset \prod_{i\in I} A_i$, and $B_i \neq \emptyset$ for each $i\in I$, then $B_i \subset A_i$ for each $i\in I$.

9. Let $(A_i)_{i\in I}$ be a family of sets, and P the cartesian product of this family. Prove that if Y is a set and $(\varphi_i)_{i\in I}$ is a family of mappings where $\varphi_i : Y \to A_i$ for each $i\in I$, such that the following condition is satisfied:

   (CP) for any set X and any family $(f_i)_{i\in I}$ of mappings where $f_i: X \to A_i$, there exists a unique mapping $f: X \to Y$ such that $f_i = \varphi_i \circ f$ for each $i\in I$,

   then there exists a unique bijective mapping $\Phi: Y \to P$ such that $\varphi_i = \mathrm{pr}_i \circ \Phi$. (This means that the cartesian product is uniquely defined by the condition (CP) up to a one-to-one correspondence.)

CHAPTER 6

NATURAL NUMBERS

A. Definition

What is a natural number? We are all familiar with the words 'zero', 'one', 'two', etc., but do we know exactly what objects have these names? In this section, we shall try to answer these questions; in other words, we shall give a definition of natural numbers. As we already know something about some objects called sets, we shall define natural numbers by means of certain sets.

Let us take, for instance, the natural number *two*. We all have an intuitive idea of *twoness*. For example, we say that each of the unordered pairs $\{a,b\}$, and $\{c,d\}$, where $a \neq b$ and $c \neq d$, has exactly *two* elements. Thus a common property of these sets is their twoness; and their having this common property may be expressed by a bijective mapping between them. Therefore we can define, in particular, the natural number *two* as a certain standard set that conveys the idea of twoness and, in general, the natural number n as a certain standard set that conveys the idea of *n-ness*.

Accordingly, the natural number *zero* should be a set conveying the idea of *zeroness*. Since there is one and only one set $\emptyset$ which contains no element, the natural number zero, denoted by 0, should be the empty set $\emptyset$. The natural number *one*, denoted by 1, should be a singleton $\{x\}$ and we might as well take $x = \emptyset$; thus $1 = \{\emptyset\}$. A certain unordered pair of distinct objects will then be defined as 2, but we already have such objects in hand, namely 0 and 1. Therefore we may define 2 as the set $\{0,1\}$, and proceed in this way to define 3 as the set $\{0,1,2\}$ and so on.

The process described above suggests the notion of the *successor* of a set and we now define the *successor* x^+ *of a set* x to be the set $x \cup \{x\}$. Thus we want to define 3 as 2^+ and 4 as 3^+, and so on *ad infinitum*.

However, we do not know if this 'and so on *ad infinitum*' is possible; in other words, whether a set exists within which this construction of successors can be carried out indefinitely. The axioms introduced earlier are not strong enough to guarantee the existence of a set large enough for this purpose; we therefore postulate the following axiom:

AXIOM OF INFINITY. *There exists a set of sets which contains* $\emptyset$ *and the successor of each of its elements.*

Such a set will be called a *successor set.*

We shall now prove that there is a smallest successor set, i.e. a successor set that is a subset of every successor set.

THEOREM 6.1. *There exists a unique successor set that is a subset of every successor set, and the elements of this set can be obtained by repeated constructions of successors of* $\emptyset$.

PROOF. Let A be a successor set and let $\mathfrak{S}$ be the set of all subsets of A which are successor sets. Then the intersection N of the sets of $\mathfrak{S}$ is easily seen to be a successor set. If B is any suceessor set, then $A \cap B$ is also a successor set and $A \cap B \subset A$. Thus $N \subset A \cap B$, implying that $N \subset B$. If N' is also a successor set that is included in every successor set, then $N \subset N'$ and $N' \subset N$, proving that $N = N'$.

Clearly N contains $\emptyset$ and all its successive successors. On the other hand, these elements of N constitute a subset M of N and M is also a successor set. Thus $M = N$ and the last assertion is proved. ∎

We are now in a position to define natural numbers.

DEFINITION 6.2. *The set N obtained in* 6.1 *is called* the set of all natural numbers. *An object n is called a* natural number *if and only if n belongs to N.*

In conformity with the usual notation, we denote by 0 the natural number zero, which is $\emptyset$, 1 the natural number one which is 0^+ and so on.

We owe our definition of natural numbers to JOHN VON NEUMANN[1] (1903–1957) and we emphasize that natural numbers, for us, are sets. This approach may be a little uncomfortable at first, but, with some practice, it does not take long to get used to.

B. Peano's axioms

Having defined the set of all natural numbers in the last section, we shall now consider its most important properties.

THEOREM 6.3. *The set N of all natural numbers has the following properties:*

(1) 0 *is a natural number.*

(2) *For each natural number n, there is a unique natural number* n^+ *which is the successor of n.*

[1] John von Neumann (1923): Zur Einführung der Transfiniten Zahlen, *Acta Szeged* **1.**

(3) 0 *is not the successor of any natural number.*

(4) *If m and n are natural numbers and* $m \neq n$, *then* $m^+ \neq n^+$.

(5) (The principle of mathematical induction). *If S is a subset of N that satisfies the following conditions:*

(i) $0 \in S$, *and*

(ii) *if* $n \in S$, *then* $n^+ \in S$,

then $S = N$.

(1), (2), (3) and (5) are easy consequences of the definition of natural numbers. In particular, (5) is the basis of the so-called method of *proof by mathematical induction*: in proving that all the natural numbers have a certain property P, we first show that 0 has the property P and then proceed to show that any arbitrary natural number n^+ has the property P under the *induction assumption* that n has the property P.

To illustrate this important method, we shall use mathematical induction to prove the following lemma, which is needed for proving (4) above.

LEMMA 6.4. *For each natural number n,*

(a) *n is a set of natural numbers,*

(b) *every element of n is also a subset of n, and*

(c) $n \notin n$.

PROOF. Let S be the set of all those natural numbers n for which (a) is true. Obviously $0 \in S$. If $n \in S$, then n is a set of natural numbers. Therefore $n^+ = n \cup \{n\}$ is also a set of natural numbers, and $n^+ \in S$. Now $S = N$ follows from (5), proving (a).

Let T be the set of all those natural numbers n for which (b) is true. Obviously $0 \in T$. If $n \in T$, then for all $m \in n$, we have $m \subset n$. For each $x \in n^+ = n \cup \{n\}$, we have $x \in n$ or $x = n$. In both cases $x \subset n^+$. Therefore $n^+ \in T$, and $T = N$ follows from (5), proving (b).

The proof of (c) is similar. ∎

We can now prove (4).

PROOF OF (4). Let m and n be natural numbers such that $m^+ = n^+$; i.e., $m \cup \{m\} = n \cup \{n\}$. Since $m \in n \cup \{n\}$, we must have $m \in n$ or $m = n$. Similarly, we get $n \in m$ or $n = m$. Hence $m = n$ is true or $m \in n$ and $n \in m$ are true. Therefore, by 6.4, we must conclude that $m = n$. ∎

The statements (1) to (5) in 6.3 are known as *Peano's axioms*. One could define the set of all natural numbers as a set satisfying these axioms. As we have adopted the approach of John von Neumann

instead, these axioms are consequences of the definition of natural numbers.

The most important properties of natural numbers may be derived directly from 6.3. For example, we have:

(6) *If n is a natural number, then $n \neq n^+$.*

(7) *If n is a natural number and $n \neq 0$, then there exists a unique natural number m such that $n = m^+$; in this case, we write $n = m + 1$ and $m = n - 1$.*

Lastly we should notice that the set N of all natural numbers is not a natural number, for otherwise we would have $N \in N$, contradicting 6.4(c) above.

C. The usual order relation of natural numbers

Having defined the set N of all natural numbers, we now introduce a most important relation into this set, namely the *usual order relation* of N. The notation and terminology used here are the same as those of Chapter 8, where the general theory of order relations will be discussed. Since natural numbers are sets, we shall formulate the usual order relation in terms of inclusion.

DEFINITION 6.5. *The relation $\leqq$ in the set N of all natural numbers such that for all $m, n \in N$, $m \leqq n$ if and only if $m \subset n$, is called the* usual order relation *of N.*

If $m \leqq n$, we say that *m is less than or equal to n* or *n is greater than or equal to m*. If $m \leqq n$ and $m \neq n$, we say that *m is less than n* or *n is greater than m* and write $m < n$. For $m \leqq n$ ($m < n$) we sometimes also write $n \geqq m$ ($n > m$).

The usual order relation clearly has the following properties for any three natural numbers m, n and p:

$0 \leqq n$,

$m \leqq n$ and $n \leqq m$ if and only if $m = n$, and

if $m \leqq n$ and $n \leqq p$, then $m \leqq p$.

Using the terminology of order relation, we say respectively

0 is the *least* natural number,

the usual order relation is *anti-symmetric*, and

the usual order relation is *transitive*.

Every natural number is less than its successor. Furthermore, there is

THEOREM 6.6. *For any natural numbers m and n,*

(a) $m < n^+$ *if and only if* $m \leqq n$,

(b) $m^+ \leqq n$ *if and only if* $m < n$.

PROOF. (a) Clearly if $m \leqq n$, then $m < n^+$. Conversely let $m < n^+$. Then $m \neq n^+$ and $m \subset n \cup \{n\}$. From this inclusion we get $m \subset n$ (i.e. $m \leqq n$) if we prove that $n \notin m$. Assume, to the contrary, that $n \in m$. Then we would have $n \in m$ and $n \subset m$, and hence $n^+ \subset m$. This, together with $m \subset n^+$ would imply $n^+ = m$, which is contradictory to $m < n^+$. Therefore $m \subset n$ and $m \leqq n$.

(b) Clearly if $m^+ \leqq n$, then $m < n$. To prove the converse, let S be the set of all natural numbers n, such that for all $m \in N$, $m^+ \leqq n$ if $m < n$. Since 0 is the least natural number, $0 \in S$. Assume now that $n \in S$ and $m < n^+$, then $m \leqq n$ by (a). That is $m = n$ or $m < n$. In either case we have $m^+ \leqq n^+$, proving that $n^+ \in S$. Then $S = N$ follows from the principle of mathematical induction. ∎

In the definition of the usual order relation of N, $m \leqq n$ is expressed in terms of *inclusion*. Using the results of 6.6, we can now express $m < n$ in terms of *belonging*.

THEOREM 6.7. *For any natural numbers m and n,* $m < n$ *if and only if* $m \in n$.

COROLLARY 6.8. *For each natural number n,* $n = \{m \in N : m < n\}$.

This useful corollary shows that every natural number is the set of all those natural numbers less than itself. Another important property of the usual order relation is that we can always compare two natural numbers with respect to this relation. In the terminology of order relation, we say that the set N *is totally ordered* by the usual order relation.

THEOREM 6.9 (Law of trichotomy). *Given any two natural numbers m and n, exactly one of the following three statements is true:* (a) $m = n$, (b) $n < m$, (c) $m < n$.

PROOF. It is easily seen that at most one of the three statements is true. Now let S be the set of all natural numbers m such that for all $n \in N$, at least one of (a), (b), (c) is true. Since 0 is the least natural number, $0 \in S$. We now assume that $m \in S$ and compare m^+ and n. By our induction assumption there are two cases. Case (1) $n \leqq m$. This implies $n < m^+$. Case (2) $m < n$. Then from 6.6(b), $m^+ \leqq n$. In both cases we have $m^+ \in S$. Therefore $S = N$ and our theorem is proved. ∎

We now derive another important property of the set N of all natural numbers.

THEOREM 6.10. *The set N of all natural numbers satisfies the* well-ordering condition: *every non-empty subset of N contains a least element, i.e. one which is less than or equal to every element of the subset in question.*

PROOF. Let M be a non-empty subset of N. We shall consider the set $L = \{x \in N : x \leqq y \text{ for all } y \in M\}$. L is non-empty since $0 \in L$. Moreover $L \neq N$ since M has at least one element and the successor of that element cannot belong to L. It follows from the principle of mathematical induction that there exists $s \in L$ such that $s^+ \notin L$. We shall show that s is the least element of M. Since $s \in L$, $s \leqq y$ for all $y \in M$. Supposing $s \notin M$, then $s < y$ for all $y \in M$ and from 6.6(b) $s^+ \leqq y$ for all $y \in M$, implying that $s^+ \in L$. Therefore we must conclude that $s \in M$. ∎

Using the language of the theory of order relations, we say that the set N is *well-ordered* by the usual order relation. Using the above result, we can reformulate the principle of mathematical induction in the following statement, which is also known as the *second principle of mathematical induction:*

THEOREM 6.11. *Let S be a subset of N. If for all $n \in N$ the condition*

(*) *if $m \in S$ for all $m \in N$ which are less than n, then $n \in S$*

is satisfied, then $S = N$.

PROOF. Assume that S is a proper subset of N. Then the subset $N \setminus S$ of N is non-empty and therefore has, by the last theorem, a least element, say n. Since any natural number less than n belongs to S, n also belongs to S by (*), contradicting the definition of n. ∎

We notice that $0 \in S$ is implicitly required in 6.11, since for 0 the hypothesis of (*) is trivially true.

It is clear that any subset M of N also satisfies the well-ordering condition of 6.10. Therefore if we replace each N by M in the statement 6.11 we get a theorem for M.

D. Recursion theorems

Besides the method of proof by mathematical induction, which we have used repeatedly in the preceding sections, there is also a method of *definition (or construction) by mathematical induction* which is most useful in some proofs of existence. This is an application of the *recursion theorem* 6.12 which enables us to define mappings f of N into some set A so that each value $f(n^+)$ is recursively determined by the values $f(m)$ where $m \leqq n$.

We now introduce some special terms used in the formulation of the recursion theorem. We say that a family of elements of a set A is a *natural sequence* if the index set is either a natural number or the set N of all natural numbers; in the former case the natural sequence is said to be *finite* and in the latter *infinite*. It readily follows from the results of Section **5B** and **5C** that the totality of all natural sequences of elements of a set is a set.

THEOREM 6.12. (The recursion theorem). *Let A be a non-empty set, $\mathfrak{S}$ the set of all natural sequences of elements of A and $\varphi: \mathfrak{S} \to A$ a mapping. Then there exists a unique mapping $f: N \to A$, such that for each $n \in N$,*

$$f(n) = \varphi(f^n)$$

where f^n is the finite natural sequence $(f(m))_{m<n}$ of elements of A.

PROOF. The present theorem follows from the fact that the set N is well-ordered and is a special case of the recursion theorem 8.18 of Section **8C.** We can use the arguments in the proof of 8.18 to prove 6.12, however we shall give a simpler proof here.

The uniqueness of such a mapping f is readily verified by a simple application of the principle of mathematical induction.

The existence of such a mapping f is obviously equivalent to the existence of an infinite natural sequence

$$(f_i)_{i \in N} = (f_0, f_1, f_2, \ldots)$$

such that

$$f_i = \varphi((f_j)_{j<i}) = \varphi((f_0, \ldots, f_{i-1})),$$

for we only have to identify $f(i) = f_i$ $(i \in N)$. If we denote, as usual, by $\emptyset$ the empty natural sequence, then, to obtain f_0, we need only set

$$f_0 = \varphi(\emptyset).$$

Consequently, for f_1 we have no other choice than $f_1 = \varphi((f_0))$, and hence we get the first two terms f_0, f_1 of the required infinite sequence. Assume now that for an arbitrary $n \in N$, there is a sequence $(f_0, f_1, \ldots, f_{n-1}, f_n)$ such that $f_i = \varphi((f_0, \ldots, f_{i-1}))$ for any $i \leqq n$. Then we put

$$f_{n+1} = \varphi((f_0, f_1, \ldots, f_n))$$

and obtain a sequence

$$(f_0, \ldots, f_n, f_{n+1})$$

such that

$$f_i = \varphi((f_0, \ldots, f_{i-1})).$$

Therefore by the principle of mathematical induction we have an infinite sequence $(f_0, f_1, \ldots)$ with the required property. ∎

Each application of the recursion theorem is called a *definition (or construction) by mathematical induction*. Under certain circumstances, we may find the mapping φ difficult to define, in which case we may use the second recursion theorem:

THEOREM 6.13. (The second recursion theorem). *Let A be a non-empty set, a an element of A and $\psi: A \to A$ a mapping. Then there exists a unique mapping $f: N \to A$ such that*

(i) $f(0) = a$, *and*

(ii) $f(n^+) = \psi(f(n))$ *for all* $n \in N$.

PROOF. The uniqueness of f follows from a direct application of the principle of mathematical induction.

For the existence of f, we construct a mapping $\varphi: \mathfrak{S} \to A$, where $\mathfrak{S}$ is the set of all natural sequences of elements of A, in the following way:

(a) $\varphi(s) = a$ if s is the empty sequence or an infinite sequence;

(b) $\varphi(s) = \psi(s_n)$ if $s = (s_m)_{m<n+1}$.

The condition (b) means that $\varphi(s)$ is the value of ψ at the last term of the sequence s. Then by 6.12 a unique mapping $f: N \to A$ exists such that for each $n \in N$, $f(n) = \varphi(f^n)$. Obviously $f(0) = a$, since f^0 is the empty sequence; for each $n \in N$, $f(n^+) = \varphi(f^{n+1}) = \psi(f(n))$. ∎

E. The arithmetic of natural numbers

To formulate the usual arithmetic of natural numbers within the framework of our theory we shall introduce two mappings from $N \times N$ into N which have the familiar properties of *addition* and *multiplication*.

We shall first show that addition may be defined and is unique.

THEOREM 6.14. *There exists one and only one mapping $\alpha: N \times N \to N$ that satisfies the following conditions:*

(i) $\alpha(x,0) = x$ *for all* $x \in N$,

(ii) $\alpha(x,y^+) = \alpha(x,y)^+$ *for all* $x,y \in N$.

PROOF. We first prove the existence of such a mapping. Let us consider the mapping $\psi: N \to N$ where for each $y \in N$, $\psi(y) = y^+$. Applying the recursion theorem, we obtain for each $x \in N$ a unique mapping $\alpha_x: N \to N$ where $\alpha_x(0) = x$ and $\alpha_x(y^+) = \psi(\alpha_x(y)) = \alpha_x(y)^+$ for each $y \in N$. We now define $\alpha: N \times N \to N$ as follows:

$$\alpha(x,y) = \alpha_x(y) \quad \text{for each } (x,y) \in N \times N.$$

Then for all $x,y \in N$, we get $\alpha(x,0) = \alpha_x(0) = x$ and $\alpha(x,y)^+ = \alpha_x(y^+) = \alpha_x(y)^+ = \alpha(x,y)^+$. Thus α satisfies (i) and (ii) and we have established the existence of α.

To prove that α is the only mapping satisfying both (i) and (ii) we assume that $\alpha^*: N \times N \to N$ is another mapping satisfying the same conditions. Then applying the principle of mathematical induction, we can easily see that for each $x \in N$, $\alpha(x,y) = \alpha^*(x,y)$ for all $y \in N$. Hence $\alpha(x,y) = \alpha^*(x,y)$ for all $(x,y) \in N \times N$, and the uniqueness of α is established. ∎

We can now define the addition of natural numbers.

DEFINITION 6.15. *The mapping α of* 6.14 *is called the* addition *of natural numbers. $\alpha(x,y)$ is called the* sum *of the natural numbers x and y, and is denoted by $x + y$.*

Thus we can say that there is one and only one way to define the addition of natural numbers so that for all natural numbers x and y,

(1) $x + 0 = x$, and

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

and so that these imply that

(3) $x + 1 = x^+$.

The following familiar properties of the addition of natural numbers can be proved easily: for any natural numbers x, y and z

(4) $(x + y) + z = x + (y + z)$ *(the associative law of addition)*

(5) $x + y = y + x$ *(the commutative law of addition)*

(6) $x + y = 0$ *if and only if* $x = 0$ *and* $y = 0$

(7) $x + y = x + z$ *if and only if* $y = z$

(8) $x = x + y$ *if and only if* $y = 0$

The proofs of these properties are based on repeated applications of the principle of mathematical induction, and with the exception of (5) are left as exercises.

Proof of (5). We first prove that for all $y \in N$,

(9) $0 + y = y$.

Obviously, it is true when $y = 0$. Now assume that $0 + y = y$, then using (2) we have $0 + y^+ = (0 + y)^+ = y^+$. Therefore (9) is true for all natural numbers y.

Next we shall prove that for all natural numbers x and y,

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

Let x be fixed arbitrarily. Then the above equation holds when $y = 0$. Assume now that it is true for a natural number y. Then using (2) again, we have $x^+ + y^+ = (x^+ + y)^+ = (x + y)^{++} = (x + y^+)^+$. Thus (10) is true for all $x,y \in N$.

Now let S be the set of all those natural numbers x such that $x + y = y + x$ for all $y \in N$. By (1) and (9) we have $0 \in S$. Assume that $x \in S$. Then by (10) and (2), $x^+ + y = (x + y)^+ = (y + x)^+ = y + x^+$, implying that $x^+ \in S$. It follows that $S = N$ and this completes the proof. ∎

Using the arithmetic of N, we may express the usual order relation in terms of addition.

THEOREM 6.16. *For any two natural numbers x and y, $x \leqq y$ if and only if $y = x + u$ for some natural number u.*

PROOF. Let us consider the set $S = \{x \in N$: *if* $x \leqq y$, *then* $y = x + u$ *for some* $u \in N\}$. Clearly $0 \in S$. Now assume that $x \in S$. If $x^+ \leqq y$, then $x < y$, and hence $y = x + t^+$ for some $t \in N$ by the induction assumption and the property (8). But then $y = x + t^+ = (x + t)^+ = x^+ + t$. Hence $x^+ \in S$ and $S = N$. Therefore for all $x,y \in N$, if $x \leqq y$ then $y = x + u$ for some $u \in N$.

To prove the converse, it is sufficient to prove that for all $x,u \in N$, $x \subset x + u$. Indeed, for any fixed x, $x \subset x + 0$. Assume that $x \subset x + u$. Since $x + u \subset (x + u)^+$ and $(x + u)^+ = x + u^+$, therefore $x \subset x + u^+$. Hence, by induction, $x \subset x + u$ for all $x,u \in N$. ∎

The above theorem may be reformulated as another *law of trichotomy*: *for any two natural numbers x and y, exactly one of the following three statements is true*: (a) $x = y$, (b) $x = y + v$ *for some non-zero natural number v, or* (c) $y = x + u$ *for some non-zero natural number u.*

To define the multiplication of natural numbers, we adopt a procedure analogous to that for defining the addition.

THEOREM 6.17. *There is one and only one mapping $\mu: N \times N \to N$ which satisfies the following conditions*:

(i) $\mu(x,0) = 0$ *for all* $x \in N$,

(ii) $\mu(x,y^+) = \mu(x,y) + x$ *for all* $x,y \in N$.

PROOF. For each $x \in N$, we define a mapping $\psi_x: N \to N$ in the following way: $\psi_x(y) = y + x$ for each $y \in N$. Then we have, for all $x \in N$, a unique mapping $\mu_x: N \to N$ satisfying the condition that

$$\mu_x(0) = 0 \quad \text{and} \quad \mu_x(y^+) = \mu_x(y) + x \quad \text{for all } y \in N.$$

Now retracing every step in the proof of 6.14, we obtain a complete proof of the present theorem. ∎

Definition 6.18. *The mapping* μ *of* 6.17 *is called the* multiplication *of natural numbers.* $\mu(x,y)$ *is called the* product *of the natural numbers* x *and* y, *and is denoted by* $x \cdot y$ *or simply* xy.

Thus we may say that there is one and only one way to define multiplication of natural numbers so that for all natural numbers x and y,

(11) $x0 = 0$, and

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

and hence also

(13) $x1 = x$.

The following familiar properties of the multiplication of natural numbers are easily proved. For any natural numbers x, y and z

(14) $xy = yx$ *(the commutative law of multiplication)*

(15) $(xy)z = x(yz)$ *(the associative law of multiplication)*

(16) $x(y + z) = xy + xz$ *(the distributive law)*

(17) $xy = 0$ *if and only if* $x = 0$ *or* $y = 0$

(18) $xy = xz$ *and* $x \neq 0$ *only if* $y = z$

Even (*odd*) natural numbers are defined as natural numbers of the form $2n$ ($2n + 1$). Thus each natural number is either odd or even; and the successor of an even (odd) natural number is odd (even).

F. Integers and rational numbers

The set N of all natural numbers is inadequate for a number of reasons. There is neither a natural number x for which $2 + x = 1$ nor a natural number y for which $2y = 1$. This means that subtraction and division by non-zero natural numbers cannot always be carried out within the set N. Thus in order to extend the scope of arithmetic, we shall construct two new sets, the set of all integers and the set of all rational numbers, the first of which will admit unrestricted subtraction and second of which will admit unrestricted division, except by zero, as well as unrestricted subtraction. The set of all rational numbers will include in a certain sense as a subset the set of all integers, which in turn will include N as a subset, so we can say that we have extended N and that N is *embedded* in the set of all integers and the set of all rational numbers.

The Integers

Let us consider the relation R in the set $N \times N$ such that for all (m,n), $(p,q) \in N \times N$, $(m,n)R(p,q)$ if and only if $m + q = n + p$. R is reflexive, symmetric and transitive, and is thus an equivalence relation

in $N \times N$. We shall call the quotient set $(N \times N)/R$ *the set of all integers*, and denote it by Z. An object is called an *integer* if and only if it belongs to Z. Thus every integer is an equivalence class $(m,n)/R$.

The integers satisfy the following conditions:

(a) $(m,n)/R = (m', n')/R$ if and only if $m + n' = m' + n$,

(b) *if* $(m,n)/R = (m', n')/R$ and $(p,q)/R = (p', q')/R$, then
$(m+p, n + q)/R = (m' + p', n' + q')/R$, and
$(mp + nq, mq + np)/R = (m'p' + n'q', m'q' + n'p')/R$.

In view of (a) and (b), we can define addition and multiplication of integers in the following way:

$$(m,n)/R + (p,q)/R = (m + p, n + q)/R,$$
$$(m,n)/R \cdot (p,q)/R = (mp + nq, mq + np)/R.$$

Addition and multiplication are both associative and commutative, and multiplication is distributive over addition. In other words, for all $a, b, c \in Z$, $(a + b) + c = a + (b + c)$, $(ab)c = a(bc)$, $a + b = b + a$, $ab = ba$, and $a(b + c) = ab + ac$.

For any two integers a and b, there exists a unique integer c so that $a = b + c$. Indeed, if $a = (m,n)/R$ and $b = (p,q)/R$, then $c = (m+q, n+p)/R$ is the unique solution of $b + c = a$. This integer c will be denoted by $a - b$, and simply by $-b$ if $a = (0,0)/R$.

Now it follows from the law of trichotomy that every integer $(m,n)/R$ may be expressed uniquely in exactly one of the following forms:

(i) $(0,0)/R$ or

(ii) $(u,0)/R$ where $u \in N$ and $u \neq 0$, or

(iii) $(o,v)/R$ where $v \in N$ and $v \neq 0$.

(i) is called *the zero integer*, (ii) *a positive integer* and (iii) *a negative integer*. Thus Z consists of the zero integer, the positive integers and the negative integers.

The mapping $\xi : N \to Z$ where for each $n \in N$, $\xi(n) = (n,0)/R$, is easily seen to be injective, and satisfies the conditions $\xi(m + n) = \xi(m) + \xi(n)$ and $\xi(mn) = \xi(m)\xi(n)$ for all $m,n \in N$. Thus we can identify every non-negative integer $(n,0)/R$ with the natural number n and consider N as a subset of Z; in other words, we just write n for $(n,0)/R$. It follows that the negative integer $(0,n)/R$ may be written as $-n$.

The Rational Numbers

We have seen that subtraction can always be carried out in the set Z of all integers. However, this set is still inadequate as far as division is

concerned, and the equation $2x = 1$ still admits no solution in Z. Thus we want to extend Z further.

Let $Z' = Z \setminus \{0\}$. Then the relation S in $Z \times Z'$ defined by

$$(a,b)S(c,d) \quad \textit{if and only if} \quad ad = bc$$

is an equivalence relation. We call the quotient set $Q = (Z \times Z')/S$ *the set of all rational numbers.* An object is a *rational number* if and only if it belongs to Q. We shall denote the rational number $(a,b)/S$ by $\frac{a}{b}$ or a/b.

Rational numbers satisfy the following conditions:

(c) $\frac{a}{b} = \frac{a'}{b'}$ if and only if $ab' = a'b$

(d) if $\frac{a}{b} = \frac{a'}{b'}$ and $\frac{c}{d} = \frac{c'}{d'}$, then $\frac{ad + bc}{bd} = \frac{a'd' + b'c'}{b'd'}$ and $\frac{ac}{bd} = \frac{a'c'}{b'd'}$

This allows us to define addition and multiplication in Q as follows:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \quad \text{and} \quad \frac{a}{b}\frac{c}{d} = \frac{ac}{bd}$$

Both addition and multiplication are associative and commutative and multiplication is distributive over addition.

The mapping $\zeta: Z \to Q$ where $\zeta(a) = \frac{a}{1}$ is an injective mapping such that

$$\zeta(a + b) = \zeta(a) + \zeta(b),$$
$$\zeta(ab) = \zeta(a)\zeta(b),$$

and in view of this, we *embed* Z in Q by identifying the rational number $\frac{a}{1}$ with the integer a and consider Z as a subset of Q.

Subtraction and division by non-zero elements can always be carried out in Q. Indeed, for any two rational numbers $\frac{a}{b}$ and $\frac{c}{d}$, the equation $\frac{a}{b} = \frac{c}{d} + x$ admits the unique solution $x = \frac{a}{b} + \frac{-c}{d}$, and if furthermore $\frac{c}{d} \neq 0$, the equation $\frac{a}{b} = \frac{c}{d}y$ admits the unique solution $y = \frac{ad}{bc}$.

G. Exercises

1. Let x be a set whose successor x^+ is a natural number. Prove that x is also a natural number.

2. If n is a natural number such that $0 \neq n$ and $1 \neq n$, show that $1 \in n$.

3. Prove that for any two natural numbers m and n, $m^+ < n^+$ if and only if $m < n$.

4. Prove that for any natural number n, there is no natural number m such that $n < m < n^+$.

5. Let M be a non-empty subset of N such that every element of M is less than or equal to a fixed element n of N. Prove that there exists an element m of M such that $x \leqq m$ for every element x of M.

6. Prove the second principle of mathematical induction from the principle of mathematical induction, and hence deduce that the set of all natural numbers satisfies the well-ordering condition.

7. Let A be a non-empty set and a an element of A. For each $n \in N$, let ψ_n be a mapping from A into A. Prove that there exists a unique mapping $f: N \to A$ such that $f(0) = a$ and $f(n^+) = \psi_n(f(n))$ for each $n \in N$.

8. Prove that for any three natural numbers m, n, p, $m + p < n + p$ if and only if $m < n$.

9. If m, n, p are any natural numbers, and $p \neq 0$, prove that $mp < np$ if and only if $m < n$.

10. Let m and n be any natural numbers and $n \neq 0$. Prove that there exist unique natural numbers q and r such that $m = qn + r$ and $0 \leqq r < n$.

CHAPTER 7

FINITE AND INFINITE SETS

A. Equipotent sets

In this chapter we shall be mainly concerned with the problem of making comparisons between sets. So far our comparisons of any two sets have concerned whether or not one is a subset of the other; in other words, whether or not there exists an identity mapping of one set onto a subset of the other set. This also means that our main tool for comparison has so far been identity mappings. We now propose to compare sets on a broader base and in fact by means of bijective mappings. As a result of this comparison, sets may be grouped into *equipotence classes* in such a way that any two sets of the same equipotence class stand in one-to-one correspondence to each other. Using the natural numbers as standard sets for comparison, we may then introduce the concept of *finite sets*. In this section we study the main properties of bijective mappings.

DEFINITION 7.1. *Two sets are said to be* equipotent *to each other (or simply equipotent) if and only if there exists a bijective mapping between them.*

We indicate that two sets A and B are equipotent by writing $A \sim B$. Using this notation, we have

(a) $A \sim A$ for all sets A,

(b) if $A \sim B$ and $B \sim C$, then $A \sim C$, and

(c) if $A \sim B$, then $B \sim A$.

For the sake of convenience, we may therefore say that the equipotence is reflexive, transitive and symmetric. For the same reason, we may speak of *equipotence classes of sets*. Such a class, however, is not a set.

If a set A is equipotent to a subset of a set B, we write $A \lesssim B$. Then the symbol $\lesssim$ obviously has the following properties:

(d) $A \lesssim A$ for all sets A; and

(e) if $A \lesssim B$ and $B \lesssim C$, then $A \lesssim C$.

This symbol is therefore reflexive and transitive.

Our new symbols $\sim$ and $\lesssim$ correspond to the well-known symbols $=$ and $\subset$. We may now ask ourselves if it is true that for any two sets A and B

(f) if $A \lesssim B$ and $B \lesssim A$, then $A \sim B$; and

(g) $A \lesssim B$ or $B \lesssim A$.

The symbols $=$ and $\subset$ obviously do have the property corresponding to (f) but not the property corresponding to (g). We shall prove (f) presently in 7.2. The statement (g) expresses the fact that any two sets A and B can be compared by $\lesssim$. This statement is true and will be proved in Chapter 8.

THEOREM 7.2. (Schröder-Bernstein Theorem) *If each of the sets A and B is equipotent to a subset of the other, then A and B are equipotent.*

PROOF. By hypothesis, there exist injective mappings $f: A \to B$ and $g: B \to A$. Suppose there is a subset S of A so that $g[B \setminus T] = A \setminus S$, where $T = f[S]$, then f induces a bijective mapping $f': S \to T$ where $f'(a) = f(a)$ for all $a \in S$, and similarly g induces a bijective mapping $g': B \setminus T \to A \setminus S$. In this case a bijective mapping $h: A \to B$ can be defined by

$$h(a) = f'(a) \quad \text{for all } a \in S$$

$$h(a) = g'^{-1}(a) \quad \text{for all } a \in A \setminus S.$$

Thus our proof will be complete if we can establish the existence of such a set S.

For each subset X of A let us denote by X^* the subset $A \setminus g[B \setminus f[X]]$ of A. Under this notation, our set S is a subset of A such that $S = S^*$. It is easy to see that

(*) if $X_1 \subset X_2 \subset A$, then $X_1^* \subset X_2^*$

Now let

$$\mathfrak{S} = \{X \in \mathfrak{P}(A): X^* \subset X\}.$$

Then we shall see that $S = \bigcap_{X \in \mathfrak{S}} X$ is our required set. For each $X \in \mathfrak{S}$, we have $S \subset X$ and $X^* \subset X$, and it follows from (*) that $S^* \subset X^* \subset X$. This implies that $S^* \subset S$. Using (*) again, we have $(S^*)^* \subset S^*$, showing that $S^* \in \mathfrak{S}$. By the definition of S we have $S \subset S^*$. Hence $S = S^*$. ∎

If two sets are equipotent then it is easy to see that their power sets are equipotent. We now want to know whether any set is equipotent to its own power set. The answer is given in the following theorem, known as Cantor's Theorem:

THEOREM 7.3. (Cantor's Theorem)[1] *Every set is equipotent to a proper subset of its power set, but is not equipotent to the power set itself.*

PROOF. We first prove that any set A is not equipotent to its power set $\mathfrak{P}(A)$. It is sufficient to show that no surjective mapping $f: A \to \mathfrak{P}(A)$ exists. Assume that f is surjective and let $B = \{x \in A: x \notin f(x)\}$. Since $B \in \mathfrak{P}(A)$ and f is surjective, there exists $a \in A$ so that $f(a) = B$. We then have two mutually exclusive cases $a \in B$ and $a \notin B$. If $a \in B$ (respectively $a \notin B$), then $a \notin f(a)$ (respectively $a \in f(a)$). But $f(a) = B$. Therefore both cases lead to contradictions. Hence we must conclude that f is not surjective. Therefore A and $\mathfrak{P}(A)$ are not equipotent.

On the other hand, the mapping $g: A \to \mathfrak{P}(A)$ where $g(x) = \{x\}$ for every $x \in A$ is injective, which proves that A is equipotent to the subset $g[A]$ of $\mathfrak{P}(A)$. Since g cannot be bijective, $g[A]$ must be a proper subset of $\mathfrak{P}(A)$. ∎

B. Finite sets

We have seen that equipotence can be used as a tool for comparing sets. Let us now apply this to the natural numbers and sets of natural numbers.

THEOREM 7.4. *Natural numbers have the following properties:*

(F_1) *Any two equipotent natural numbers are equal.*

(F_2) *Every subset A of a natural number n is equipotent to a natural number m where $m \leqq n$; furthermore $m = n$ if and only if $A = n$.*

PROOF. (F_1) Let S be the set of all natural numbers n such that any natural number m equipotent to n is equal to n. Obviously 0 belongs to S. Assume that n belongs to S and n^+ is a natural number equipotent to n^+. If $f: m^+ \to n^+$ is a bijective mapping, we can easily find a bijective mapping $g: m^+ \to n^+$ so that $g(m) = n$ (see Chapter 4, Exercise 1). This means that g induces a bijective mapping between the sets m and n. From the induction assumption, we have $m = n$; hence $m^+ = n^+$. Therefore $n^+ \in S$.

(F_2) Let T be the set of all natural numbers n such that (F_2) is true. Since (F_2) is trivially true for 0, 0 belongs to T. Assume that n belongs to T and A is a subset of n^+. Then we have two cases:

[1] GEORG CANTOR (1845–1918) initiated the study of set theory in the last quarter of the last century. The basic idea of the proof as well as the methods used in the proofs 7.10 and 7.12 are already found in G. Cantor (1892): Über eine elementare Frage der Mannigfältigkeitslehre, *Jahresbericht der Deutschen Mathematiker Vereinigung* **1**, pp. 75–78.

Case 1: n does not belong to A. Here it follows that A is also a subset of n. By our induction assumption, A is equipotent to a natural number m, which is therefore unique by (F_1), so that $m \leqq n$; hence $m < n^+$. This proves that the first statement of (F_2) holds for n^+. The second statement also holds for n^+ since $A \neq n^+$ and $m < n^+$.

Case 2: n belongs to A. Let $A' \coloneqq A \setminus \{n\}$. Then A' is a subset of n, and hence A' is equipotent to a unique natural number m such that $m \leqq n$. This means that A is equipotent to m^+ and $m^+ \leqq n^+$. Now $A = n^+$ if and only if $A' = n$, therefore if and only if $m^+ = n^+$. Hence (F_2) holds for n^+.

In both cases n^+ belongs to T. This completes the proof of the theorem. ∎

It follows from (F_1) that each set can be equipotent to at most one natural number. Natural numbers therefore can be used as standard sets for comparison by equipotence. In other words, they can serve as norms for measuring sets. On the other hand, natural numbers have been constructed to convey the intuitive idea of 'finiteness'. Therefore we have two kinds of sets.

DEFINITION 7.5. *A set is* finite *if and only if it is equipotent to a natural number; otherwise it is* infinite.

Since each finite set A is equipotent to a unique natural number $n(A)$, we may now conveniently say that A *has* $n(A)$ *elements*; and since each natural number is finite, we may say that natural numbers are *finite numbers*. Thus a finite set is a set consisting of a finite number of elements.

With respect to equipotence, finite sets and natural numbers have corresponding properties. Rewriting 7.4, we obtain some elementary properties of finite sets.

COROLLARY 7.6. *The following statements hold:*

(a) *Every finite set has a unique number of elements.*

(b) *Two finite sets are equipotent if and only if they have the same number of elements.*

(c) *If a set A is finite, then every subset of A is finite.*

(d) *A subset of a finite set A has no more elements than A.*

(e) *If a set is finite, then it is not equipotent to any of its proper subsets.*

By (a) and (b) finite sets can be grouped into equipotence classes according to the number of elements they have.

The converse of (c) is trivially true, since A is a subset of itself. We shall see later in Section **D** that the converse of (c) is also true, which therefore means that the property

A is not equipotent to any of its proper subsets

characterizes finite sets.

Finally, let us decide whether the set N of all natural numbers is a finite set. Let A be the set of all even natural numbers. Then A is clearly a proper subset of N and the mapping $f:N \to A$ defined by $f(n) = 2n$ for each $n \in N$, is bijective. This means that N is equipotent to one of its proper subsets and hence it must be infinite.

THEOREM 7.7. *The set N of all natural numbers is an infinite set.*

We recall that N is defined as the smallest successor set; therefore successor sets are infinite sets. For this reason, we have called the axiom that postulates the existence of a successor set the *axiom of infinity*.

C. Countable sets

The concept of natural numbers is also closely connected with the ordinary process of counting. More precisely, if A is a non-empty finite set and n^+ is the number of elements of A, then by means of a bijective mapping $f: n^+ \to A$ we can count the elements of A beginning with the element $f(0)$ of A, then $f(1)$ and so on, finishing with $f(n)$. During the entire process, we always count the element $f(m^+)$ where $m < n$ immediately after having counted the element $f(m)$ of A. In the same way, we can count the even natural numbers by means of the bijective mapping f in the proof of 7.7. With N as a standard set, we have another equipotence class of sets.

DEFINITION 7.8. *A set is said to be* countably infinite *if and only if it is equipotent to N. A* countable *set is a set which is either finite or countably infinite. A set which is not countable is* uncountable.

It follows from this definition that N is countable.

Corresponding to 7.6(c) we have the following theorem about subsets of a countable set:

THEOREM 7.9. *Every subset of a countable set is countable.*

PROOF. Let X be a countable set and A a subset of X. If X is finite, then every subset of X is finite, and hence countable. Therefore we need

only consider the case where X is a countably infinite set and A is an infinite subset of X. For the present purpose, we may assume, without loss of generality, that X is the set N of natural numbers.

For each $a \in A$ consider the subsets $S_a = \{x \in A: x < a\}$ and $T_a = \{x \in A: a < x\}$ of A. Since $S_a = a \cap A$ and a is a finite set, S_a is a finite set; since A is an infinite set, T_a is non-empty. If we denote by $n(S_a)$ the number of elements of the finite set S_a, we obtain a mapping $f: A \to N$ where $f(a) = n(S_a)$ for all $a \in A$. It remains to be proved that f is bijective. If a and b are elements of A such that $a < b$, then S_a is a proper subset of S_b. Hence $n(S_a) < n(S_b)$, proving f to be injective. Clearly $0 = f(a_0)$ where a_0 is the least element of A. If $n = f(a)$ for some element a of A, then $n^+ = f(b)$ where b is the least element of the non-empty subset T_a of A. Hence f is surjective, by induction. ∎

From the above theorem, we see that countability is a property of sets that is inherited by subsets. Therefore we say that the countability is *hereditary*.

For proving the cartesian product $N \times N$ to be countable, we shall use Cantor's *method of diagonal enumeration*. This can be described briefly. First we arrange the elements of $N \times N$ into an infinite array

$$\begin{array}{llllll}
(0,0) & (0,1) & (0,2) & (0,3) & \ldots \\
(1,0) & (1,1) & (1,2) & \ldots & \\
(2,0) & (2,1) & \ldots & & \\
(3,0) & \ldots & & & \\
\ldots & & & &
\end{array}$$

Then we begin our enumeration by counting step by step

(o) the first element from the left on the top row—i.e. (0,0);

(i) the second element from the left on the top row and then each element down the diagonal—i.e. (0,1), (1,0);

(ii) the third element from the left on the top row and then each element down the diagonal—i.e. (0,2), (1,1), (2,0),

and so forth. Now it is a simple exercise in arithmetic to define the mapping $f: N \times N \to N$ corresponding to the above counting process.

THEOREM 7.10. *The set $N \times N$ is countably infinite.*

PROOF. Let us define a mapping $f: N \times N \to N$ as follows. For every $(x,y) \in N \times N$, let $f(x,y) = p + x$, where $2p = (x+y)(x+y+1)$. If (x,y), $(x',y') \in N \times N$ and $(x,y) \neq (x',y')$, there are two cases:

Case 1. $x' + y' > x + y$. Then $x' + y' \geqq x + y + 1$, implying that $f(x',y') > f(x,y)$. Similarly for $x + y > x' + y'$.

Case 2. $x + y = x' + y'$. If $x < x'$, then $f(x,y) < f(x',y')$. Similarly for $x > x'$.

In both cases, $f(x,y) \neq f(x',y')$. Therefore f is injective, and $N \times N$ is equipotent to a subset of N.

Since the subset $\{(x,y) \in N \times N : y = 0\}$ of $N \times N$ is equipotent to N, $N \times N$ is infinite. Hence $N \times N$ is countably infinite. ∎

Using the above theorems, we can prove that the set of all integers and the set of all rational numbers are countable.

We shall now give some examples of uncountable sets.

EXAMPLE 7.11. Applying 7.3, we see that the set of all sets of natural numbers, i.e. the power set of N, is uncountable.

EXAMPLE 7.12. An infinite decimal is a natural sequence $(a_i)_{i \in N}$ of natural numbers where $a_i \leqq 9$ for $i \neq 0$, and may be written as

$$a_0 \cdot a_1 a_2 \ldots$$

Every infinite decimal represents a non-negative real number, and vice versa. Two distinct infinite decimals $s = (a_i)_{i \in N}$ and $r = (b_i)_{i \in N}$ represent the same real number if and only if there exists $n \in N$ such that

(i) $a_n = b_n{}^+$,

(ii) $a_i = b_i$ if $i < n$, and

(iii) $a_i = 0$ and $b_i = 9$ if $i > n$.

That is,

$$r = b_0 \cdot b_1 b_2 \ldots b_n \; 9\; 9\; 9 \ldots \quad \text{and}$$

$$s = b_0 \cdot b_1 b_2 \ldots b_n{}^+ 0\; 0\; 0 \ldots$$

We shall now show that the set of all non-negative real numbers is uncountable. Let us first assume the contrary, in which case every non-negative real number is represented by an element of a set $S = \{s^k : k \in N\}$ of infinite decimals. We shall now find a non-negative real number that is not represented by any s^k of S. Let $s^k = (a_i^k)_{i \in N}$. There exists an infinite decimal $t = (a_i)_{i \in N}$ such that $a_0 \neq a_0^0$, and $a_i \neq a_i^i$, 0, 9 if $i \neq 0$.

For each $k \in N$, $t \neq s^k$ since $a_k \neq a_k^k$. On the other hand, since $a_i \neq 0, 9$ for all $i \neq 0$, t cannot represent a non-negative real number which is represented by any s^k of S. This means that the non-negative real number represented by t does not belong to the set of all non-negative real numbers, which is clearly impossible. Hence the set of all non-negative real numbers is uncountable, and this implies that the set of all real numbers is uncountable.

This proof that the set of all real numbers is uncountable is due to G. CANTOR.

D. Infinite sets

In Section **B**, we have seen that finite sets can be classified according to the number of elements they have. Because there are not as yet sufficient standard infinite sets, we can decide only whether an infinite set is countable or not. A more detailed classification of infinite sets will be given in Chapter 9; here we shall study only some basic properties of infinite sets.

Clearly any infinite set A has finite subsets (e.g. $\emptyset$, singletons, etc.) and infinite subsets (e.g. A itself, $A \setminus \{x\}$ where $x \in A$, etc.). The next theorem shows that among the infinite subsets there is always at least one that is countable.

THEOREM 7.13. *Every infinite set has a countably infinite subset.*

PROOF. Let A be an infinite set. To prove that A has a countably infinite subset, it is sufficient to prove that there exists an injective mapping $f: N \to A$. We shall now define f by induction.

Let a be a choice function of the non-empty set A, i.e. $a(B) \in B$ for all non-empty subsets B of A. Let $\mathfrak{S}$ be the set of all natural sequences of elements of A. For each $s \in \mathfrak{S}$ we denote by A_s the set of all terms of the sequence s. Then we define a mapping $\varphi: \mathfrak{S} \to A$ such that for each $s \in S$

$$\varphi(s) = \begin{cases} a(A \setminus A_s) & \text{if } s \text{ is finite} \\ a(A_s) & \text{if } s \text{ is infinite} \end{cases}$$

By the recursion theorem, we obtain a mapping $f: N \to A$ such that $f(n) = \varphi(f^n)$ where f^n is the finite natural sequence $(f(m))_{m<n}$. Since $f(n) \notin \{f(m): m < n\}$, the mapping f is injective. ∎

Finally we can prove the converse of 7.6(e).

THEOREM 7.14. *A set is infinite if and only if it is equipotent to a proper subset of itself.*

PROOF. If a set is equipotent to one of its proper subsets, then by 7.6(e) it is infinite.

Now let A be an infinite set. By 7.13, A has a countably infinite subset $B = \{x_i : i \in N\}$, where $x_i \neq x_j$ if $i \neq j$. A is the union of the disjoint subsets B and $A \setminus B$. Let $B' = B \setminus \{x_0\}$ and $A' = A \setminus \{x_0\}$. Then A' is a proper subset of A and is also the union of the disjoint subsets B' and $A \setminus B$. We can now construct a mapping $f: A \to A'$ as follows:

$$f(x_i) = x_{i+1} \text{ for each } x_i \in B,$$

$$f(x) = x \text{ for each } x \in A \setminus B.$$

It is easily seen that f is bijective. Hence A is equipotent to its proper subset A'. ∎

COROLLARY 7.15. *A set is finite if and only if it is not equipotent to any proper subset of itself.*

7.14 and 7.15 provide alternative definitions respectively for infinite sets and finite sets without reference to natural numbers.

E. Exercises

1. Let A and B be finite sets. Prove that
$$n(A \cap B) + n(A \cup B) = n(A) + n(B).$$

2. Let A, B and C be finite sets. Prove that
$$n(A) + n(B) + n(C) + n(A \cap B \cap C)$$
$$= n(A \cup B \cup C) + n(A \cap B) + n(B \cap C) + n(C \cap A).$$

3. Prove that for any finite sets A and B, $n(A \times B) = n(A)\, n(B)$.

4. Let A and B be finite sets and $n(A) = m$, $n(B) = n$. Prove that $n!/(n-m)!$ is the number of injective mappings from A into B.

5. A family is said to be finite if and only if its index set is finite. Prove that the union and the cartesian product of a finite family of finite sets are both finite sets.

6. Prove that a non-empty set A is countable if and only if there exists a surjective mapping from N onto A.

7. A family is said to be countable if and only if its index set is countable. Prove that the union of a countable family of countable sets is countable.

8. For each $n \in N$, let $T_n = \{x \in N : x \geqq n\}$. Prove that
   (a) T_n is infinite for all $n \in N$,
   (b) a non-empty subset M of N is equal to T_n for some $n \in N$ if and only if $m^+ \in M$ for each $m \in M$.

9. Prove that the set of all finite subsets of N is countable.

*10. Let A be an infinite set and B a countable subset of A such that $A \setminus B$ is infinite. Prove that A and $A \setminus B$ are equipotent.

*11. Prove that a set A is infinite if and only if it satisfies the following condition: for every mapping $f : A \to A$, there exists a non-empty proper subset B of A such that $f[B] \subset B$.

(A hint: for the necessity, assume that A is infinite and $f : A \to A$ does not satisfy the condition, and consider the set of all terms of the natural sequence $(x_i)_{i \in N}$ of elements of A where $x_0 \in A$ and $x_{i+1} = f(x_i)$ for all $i \in N$).

*12. Let A be a set and $\mathfrak{F}(A)$ the set of all finite subsets of A.
   (a) Prove that $\mathfrak{F}(A)$ is the smallest (in the sense of inclusion) of all subsets $\mathfrak{G}$ of $\mathfrak{P}(A)$ satisfying the following conditions:
      (1) $\emptyset \in \mathfrak{G}$,
      (2) if $X \in \mathfrak{G}$ and $x \in A$, then $X \cup \{x\} \in \mathfrak{G}$.
   (b) Make use of (a) to prove that the union of two finite subsets X and Y of A is finite.
   (c) Deduce from (a) and (b) that if A is finite, then $\mathfrak{P}(A)$ is finite.

*13. A property P of a set is *hereditary* if and only if each subset of a set having P also has P; it is *productive* (respectively, *countably productive*,

finitely productive) if and only if the cartesian product of a non-empty family (respectively, a non-empty countable family, a non-empty finite family) of sets having P also has P; it is *divisible* if and only if every quotient set of a set with P also has P. The following table is filled out by + or −, depending on whether the property at the head of the column is or is not of the kind listed on the left. Prove that the following table is correct:

	Finite	Countable	Infinite
Hereditary	+	+	−
Divisible	+	+	−
Productive	−	−	+
Countably productive	−	−	+
Finitely productive	+	+	+

***14.** Let A be an infinite set. Prove that every subset $\mathfrak{S}$ of $\mathfrak{P}(A)$ that satisfies the conditions

(i) B is a finite set for every $B \in \mathfrak{S}$, and (ii) $A = \bigcup_{B \in \mathfrak{S}} B$

is equipotent to A.

CHAPTER 8

ORDERED SETS

A. Order relations

The concept of order in elementary mathematics and in daily life is so familiar to everybody that a motivation seems hardly to be necessary here. In fact we have discussed at some length the usual order relation of natural numbers. In this chapter we shall develop the general theory of order relations within the framework of set theory. The familiar results of the usual order relation of natural numbers may now serve as examples to illustrate the more abstract concepts of this chapter.

DEFINITION 8.1. *Let A be a set. Then a relation R defined in A is said to be an* order relation *in A if and only if the following conditions are satisfied:*

(i) *aRb and bRa if and only if $a = b$; and*

(ii) *if aRb and bRc, then aRc.*

The condition (i) means that R is *reflexive* and *anti-symmetric*; (ii) means that R is *transitive*. It is clear that if R is an order relation in A, then the inverse relation R^{-1} is also an order relation in A.

DEFINITION 8.2. *An* ordered set *is an ordered pair (A,R) where A is a set and R is an order relation in A.*

When dealing with one ordered set (A,R) at a time, we find the following abbreviations and notations convenient:

(a) We replace the symbols R and R^{-1} respectively by the familiar inequality signs $\leqq$ and $\geqq$;

(b) we write A for the ordered set $(A, \leqq)$ if no confusion is possible, and in this case an element or a subset of A is understood to be respectively an element or a subset of the set A;

(c) we say that two elements x and y of A are ***$\leqq$-comparable*** if $x \leqq y$ or $y \leqq x$;

(d) for $x \leqq y$, we say, as the case may be, that x is less than or equal to y, x is smaller than or equal to y, x precedes y, or that x is inferior to y;

(e) for $y \geqq x$, we say, as the case may be, that y is greater than or equal to x, y is larger than or equal to x, y follows x, or that y is superior to x; and

(f) for $x \leqq y$ and $x \neq y$, we write $x < y$, and say that x is less than y, x is smaller than y, x strictly precedes y, or that x is strictly inferior to y; similarly for $y > x$.

When several order relations are under consideration, signs such as $\leqq$, $\leqslant$, $\preceq$, $\preceq$, etc. can be used.

EXAMPLE 8.3. The usual order relation of natural numbers is clearly an order relation in the set N of all natural numbers.

EXAMPLE 8.4. Let A be a set. For any two subsets X and Y of A, we put

$$X \leqq Y \text{ if and only if } X \subset Y.$$

It can immediately be seen that $\leqq$ is an order relation in the power set $\mathfrak{P}(A)$ of A, and we say that $\mathfrak{P}(A)$ is *ordered by inclusion.*

EXAMPLE 8.5. Let A and B be sets and consider the set E of all mappings $f: A' \to B$, where A' is any subset of A. We can define an order relation in E as follows: for any two mappings $f: A' \to B$ and $g: A'' \to B$ of E,

$$f \leqq g \text{ if and only if } A' \subset A'' \text{ and } g \mid A' = f.$$

We also say that E is *ordered by extension.*

EXAMPLE 8.6. Let $(A, \leqq)$ be an ordered set and B a subset of the set A. We define an order relation S in B as follows: for any two elements x and y of B

$$xSy \text{ if and only if } x \leqq y.$$

This order relation S in B is said to be *induced* by the order relation $\leqq$ in A, and the ordered set B is called a subset of the ordered set A. If no confusion is possible, we shall again denote S by $\leqq$.

From the above examples, we see that in the ordered set of 8.3 any two elements are $\leqq$-comparable, whereas this is not the case for the ordered sets of 8.4 and 8.5 when A consists of more than one element.

DEFINITION 8.7. *An ordered set* $(A, \leqq)$ *is said to be* totally ordered, *or a* chain, *if and only if any two elements of* A *are* $\leqq$*-comparable.*

It follows immediately from the definition that the familiar law of trichotomy holds in any totally ordered set A: for any two elements x and y of A, it is either (i) $x < y$ or (ii) $x = y$ or (iii) $y > x$.

In an ordered set, certain elements having special properties with respect to the order relation deserve our attention; these are defined below.

DEFINITION 8.8. *Let $(A, \leqq)$ be an ordered set, x an element of the set A, and B a subset of the set A.*

(a) *x is a* maximal (minimal) element *of A if and only if for each $a \in A$, $x \leqq a$ ($a \leqq x$) implies $x = a$.*

(b) *x is a* greatest (least) element *of A if and only if $a \leqq x$ ($x \leqq a$) for all $a \in A$.*

(c) *x is an* upper (*a* lower) bound *of B if and only if $b \leqq x$ ($x \leqq b$) for all $b \in B$.*

(d) *x is a* supremum (*an* infimum) *of B if and only if x is a least (greatest) element of the set of all upper (lower) bounds of B with respect to the induced order relation.*

It is clear that if a greatest element, a least element, a supremum or an infimum exists, then it is unique. This is not the case with maximal elements, minimal elements, upper bounds and lower bounds as shown in the following example:

EXAMPLE 8.9. Let X be a set consisting of more than two elements, and $\mathfrak{A}$ the set of all non-empty proper subsets of X ordered by inclusion. Then every singleton in X is a minimal element of $\mathfrak{A}$ and the complement of each singleton in X is maximal element of $\mathfrak{A}$. If $\mathfrak{B}$ is a subset of $\mathfrak{A}$, then any subset of X that includes the union of the sets of $\mathfrak{B}$ as a subset is an upper bound of $\mathfrak{B}$, and any subset of X that is included in the intersection of the sets of B as a subset is a lower bound of $\mathfrak{B}$.

B. Mappings of ordered sets

In Section **4E** we have studied, in connection with the factorization of mappings, the compatibility of a mapping with an equivalence relation. This state of affairs is as follows. Given two sets A and B and an equivalence relation R in A, a mapping $f: A \to B$ is said to be compatible with R if and only if

$(*)$ $\quad xRy \Rightarrow f(x) = f(y)$.

Denoting the diagonal of B by D_B (see Definition 3.13), we can write $(*)$ into the equivalent statement

$(**)$ $\quad xRy \Rightarrow f(x) D_B f(y)$.

The condition $(**)$ expresses that the mapping $f: A \to B$ takes R-related elements of A into D_B-related elements of B. In other words, f is compatible with R in A and D_B in B.

This leads us to consider the more general situation in which R is any relation in the set A, S any relation in the set B, and a mapping $f: A \to B$ is compatible with R and S, i.e. satisfying the condition

(***) $xRy \Rightarrow f(x)Sf(y)$.

In particular when R and S are order relations, corresponding to the monotonic functions of classical analysis, we have the *increasing mappings*.

DEFINITION 8.10. *Let $(A, \leqq)$ and $(B, \leqslant)$ be ordered sets. A mapping $f: A \to B$ is an* increasing (*respectively a* decreasing) *mapping of the ordered set A into the ordered set B if and only if, for all x and y of A, $x \leqq y$ implies $f(x) \leqslant f(y)$ (respectively $f(x) \geqslant f(y)$). f is a* strictly increasing (*respectively* strictly decreasing) *mapping if and only if, for all x and y of A, $x < y$ implies $f(x) < f(y)$ (respectively $f(x) > f(y)$).*

Compositions of increasing mappings are easily seen to be increasing mappings.

EXAMPLE 8.11. Let A be an ordered set and $\mathfrak{P}(A)$ the power set of A ordered by inclusion. For each element a of A, let $S_a = \{x \in A : x \leqq a\}$. Then the mapping $f: A \to \mathfrak{P}(A)$ defined by $f(a) = S_a$ is a strictly increasing mapping of the ordered set A into the ordered set $\mathfrak{P}(A)$. On the other hand, the mapping $g: \mathfrak{P}(A) \to \mathfrak{P}(A)$ defined by $f(X) = A \setminus X$ for each $X \in \mathfrak{P}(A)$, is a strictly decreasing mapping of the ordered set $\mathfrak{P}(A)$ onto itself.

Corresponding to the concept of equipotence of sets, we introduce the concept of isomorphism between ordered sets.

DEFINITION 8.12. *Let A and B be ordered sets. Then a mapping $f: A \to B$ is called an* isomorphism *or a* similarity *of the ordered set A onto the ordered set B if and only if* (i) *f is bijective and* (ii) *f and f^{-1} are both increasing. Two ordered sets are said to be* isomorphic *or* similar *if and only if there exists an isomorphism between them.*

When order relations are the primary object of our study, the nature of the elements of the ordered sets in question becomes unimportant and hence the difference between isomorphic ordered sets is regarded merely as a matter of notation.

It is interesting to notice that the mapping $f: A \to \mathfrak{P}(A)$ in 8.11 defines an isomorphism of the ordered set A onto a subset of the ordered set $\mathfrak{P}(A)$. This means that for any ordered set we can always find an isomorphic model that is a set of sets ordered by inclusion.

EXAMPLE 8.13. This example emphasizes the difference between equipotence of sets and isomorphisms of ordered sets. Let $A = (N, \leqq)$ be the set of all natural numbers ordered by the usual order relation and $B = (N, \geqq)$ be the same set but ordered by the inverse relation of the usual order relation. Then A and B are distinct ordered sets, though A and B have the same elements. Here we shall show that the ordered sets A and B are essentially different; i.e. they are not isomorphic. By definition, an increasing mapping of the ordered set A into the ordered set B is a mapping $f: N \to N$ that satisfies the following condition:

$$\textit{if } x \leqq y \textit{ then } f(x) \geqq f(y).$$

Our assertion is proved if we can show that no bijective mapping $f: N \to N$ can satisfy the above condition. Let $f: N \to N$ be a bijective mapping. Then $f(0)$ is a natural number, and by our assumption on f, there exists a natural number y whose image $f(y)$ under f is the successor of $f(0)$. Now $0 < y$ but $f(0) < f(y)$. Therefore f is not an increasing mapping of the ordered set A into the ordered set B.

C. Well-ordered sets

Among the ordered sets whose general properties we have studied in the last two sections there is a class of ordered sets which distinguish themselves by the particularly simple arrangement of their elements. These are the *well-ordered sets*, first studied by G. CANTOR. In such a set, an outstanding simplicity exists that resembles the ordinary process of counting. Such a process begins with a first object, and after any object has been counted the object to be counted next follows it immediately. It is these features of the ordinary process of counting that motivate the formulation of the well-ordering condition and its equivalences. For the set N of all natural numbers we have seen in 6.10 that the well-ordering condition is satisfied, and its relation to the principle of mathematical induction is indicated in the proofs of the two Theorems 6.10 and 6.11. In a more general setting we have the following theorem:

THEOREM 8.14. *Let $(A, \leqq)$ be a totally ordered set. Then the following three statements are equivalent:*

I. The well-ordering condition. *Every non-empty subset of A contains a least element.*

II. The descending chain condition. *Every natural sequence $(a_i)_{i \in N}$ of elements of A such that $a_i \geqq a_{i+1}$ for all $i \in N$ is stationary; i.e. for some $n \in N$, $a_m = a_n$ for all $m \geqq n$.*

III. The principle of transfinite induction. *If a subset S of A satisfies the following condition* (TI), *then* $S = A$.

(TI) *An element a of A belongs to S if every element b of A strictly preceding a belongs to S.*

PROOF. I $\Rightarrow$ III. Let S be a subset of A satisfying the condition (TI). If we assume that $A \setminus S \neq \emptyset$, then we have a least element a of $A \setminus S$. For this element a the hypothesis of (TI) is satisfied. Hence a belongs to S, contradicting the fact that a belongs to $A \setminus S$.

III $\Rightarrow$ II. Let S be the subset of A consisting of all elements a of A such that every natural sequence $(a_i)_{i \in N}$ where $a_0 = a$ and $a_i \geqq a_{i+1}$ for all $i \in N$ is stationary. Then S satisfies the condition (TI) of III, and hence $S = A$.

II $\Rightarrow$ I. Assume that B is a non-empty subset of A that contains no least element. Then for each $b \in B$, the set $s(b) = \{x \in B : x < b\}$ is not empty. If g is a choice function of the non-empty set B, we define a mapping $\psi : B \to B$ by putting $\psi(b) = g(s(b))$ for all $b \in B$. Applying the recursion theorem, we get a natural sequence $(b_i)_{i \in N}$ of elements of B such that

$$b_0 > b_1 > b_2 > \ldots ,$$

contradicting II. ∎

We notice that if A contains a least element a_0, then $a_0 \in S$ is implicitly required by the condition (TI) of III, since the hypothesis of (TI) is trivially true.

DEFINITION 8.15. *An ordered set in which the well-ordering condition is satisfied is said to be* well-ordered.

The set N of all natural numbers and all subsets (including the empty subset) of N are well-ordered sets. Furthermore any subset of a well-ordered set is well-ordered by the induced order relation; thus we may also say that the well-orderedness is hereditary.

THEOREM 8.16. *A well-ordered set is a totally ordered set. Therefore in every well-ordered set both the descending chain condition and the principle of transfinite induction hold.*

PROOF. Let A be a well-ordered set and x, y any two elements of A. Then by the well-ordering condition the subset $\{x, y\}$ contains a least element. If this is x, then $x \leqq y$; otherwise $y \leqq x$. The second part of the theorem follows from 8.14. ∎

The condition (TI) of the principle of transfinite induction requires $x \in S$ under the *induction assumption* that $s(x) \subset S$ where $s(x) = \{y : y < x\}$. The principle of transfinite induction will be used frequently in the theory of well-ordered sets. Therefore it is desirable to have a name for subsets of the forms $s(x)$ as above.

DEFINITION 8.17. *Let $(A, \leqq)$ be a totally ordered set. A subset T of A is called a* segment *of A if and only if with each element x of T, T contains all elements of A preceding x.*

Trivially A itself is a segment of A, and for each element $a \in A$, the subset $s(a) = \{x \in A : x < a\}$ is a proper segment of A, i.e. the segment $s(a)$ is a proper subset of A, called the *segment of A preceding a*. For well-ordered sets, we have a one-to-one correspondence of the elements of A with the proper segments of A. Obviously for $a \neq b$, we have $s(a) \neq s(b)$. If T is a proper segment of A, then $A \setminus T$ is non-empty and contains therefore a least element, say a. Clearly $s(a) = T$.

Employing the above notation, we may reformulate the induction condition of the principle of transfinite induction as follows:

(TI) *For each $a \in A$, $a \in S$ if $s(a) \subset S$.*

The validity of the principle of transfinite induction for well-ordered sets enables us to carry out *proofs by transfinite induction* and *definitions by transfinite induction*. A proof by transfinite induction is a direct application of the principle when it is required to show that each element of a well-ordered set A has a certain property P. It is thus the demonstration of this: the least element a_0 of A has the property P and for each element a of A, a has the property P under the induction assumption that all members of the segment $s(a)$ have the property P. To understand the method of definition by transfinite induction some preparation is necessary.

Let A be a well-ordered set and X be any set. Now consider mappings whose domains are segments of A and whose range is the set X; these mappings are called *A-sequences of elements of X*. Clearly the totality of all A-sequences of elements of X constitutes a set. If $g : T \to X$ is an A-sequence, for each $a \in T$ we denote by g^a the A-sequence $g \mid s(a)$. Moreover if $b < a$, then we denote by g^{ab} the A-sequence $(g^a)^b$; clearly $g^{ab} = g^b$. Now we can formulate the recursion theorem.

THEOREM 8.18. (Theorem of transfinite recursion) *Let A be a well-ordered set, X a set and $\mathfrak{S}$ the set of all A-sequences of elements of X. Then for any mapping $\varphi : \mathfrak{S} \to X$, there is exactly one mapping $f : A \to X$ such that $f(a) = \varphi(f^a)$ for all elements a of A.*

PROOF. Uniqueness. Assume that f and g are two mappings of A into X satisfying the above condition. If a_0 is the least element of A, then $s(a_0) = \emptyset$; hence $f \mid s(a_0) = g \mid s(a_0)$ and so $f(a_0) = g(a_0)$. If for an element a of A we have $f \mid s(a) = g \mid s(a)$, then $f^a = g^a$ and hence $f(a) = \varphi(f^a) = \varphi(g^a) = g(a)$. It follows from the principle of transfinite induction that $f = g$.

Existence. Assume that a mapping f exists and satisfies the requirement that

$$f(a) = \varphi(f^a)$$

and consider the family $(f^a)_{a \in A}$ of restrictions of f. This family has the following properties:

(α_1) $f^{ab} = f^b$ for all a and b of A such that $b < a$;

(α_2) $f^a(b) = \varphi(f^{ab})$ for all $b \in s(a)$.

Conversely assume that $(g^a)_{a \in A}$ is a family of A-sequences $g^a : s(a) \to X$ with corresponding properties:

(β_1) $g^{ab} = g^b$ for all a and b of A such that $b < a$;

(β_2) $g^a(b) = \varphi(g^{ab})$ for all $b \in s(a)$,

where the expression g^{ab} denotes the restriction $g^a \mid s(b)$. In this case, we define a mapping $f : A \to X$ where

$$f(a) = \varphi(g^a) \quad \text{for all } a \in A.$$

It follows from (β_1) and (β_2) that $f^a = g^a$ for every $a \in A$. Hence

$$f(a) = \varphi(f^a)$$

and f is the mapping that we are looking for. Therefore we only have to show the existence of such a family $(g^a)_{a \in A}$. Let S be the subset of A consisting of all elements a for which g^a exists and satisfies (β_1) and (β_2). We shall now prove by transfinite induction that $S = A$. For the least element a_0 of A, we have $s(a_0) = \emptyset$ and hence g^{a_0} exists. Assume that $s(a) \subset S$ for an element a of A. By setting

$$g^a(b) = \varphi(g^b) \quad \text{for all } b \in s(a)$$

we get a mapping $g^a : s(a) \to X$. For the property (β_1) we need only verify that

$$g^{ab} = g^b \quad \text{for every } b < a.$$

For all $c \in s(b)$, we get $g^{ab}(c) = g^a(c) = \varphi(g^c)$. But by induction assumption $g^c = g^{bc}$, therefore $\varphi(g^c) = \varphi(g^{bc}) = g^b(c)$. Hence $g^{ab} = g^b$. (β_2) follows now from (β_1). This completes the proof. ∎

Any application of the theorem of transfinite recursion is called a *definition by transfinite induction.* As an example, we shall prove the important comparability theorem of well-ordered sets.

THEOREM 8.19. *Given any two well-ordered sets A and B, either* (i) *A is isomorphic to a proper segment of B or* (ii) *A is isomorphic to B, or* (iii) *B is isomorphic to a proper segment of A.*

PROOF. We can assume that both A and B are non-empty, for otherwise the theorem is trivial. We first prove that at most one of (i), (ii) or (iii) is true. This follows from the fact that no proper segment of a well-ordered set C is isomorphic to C. In fact, if T is a segment of C and $g: C \to T$ is an isomorphism, then applying the principle of transfinite induction, we easily prove that g is the identity mapping of C.

We now prove that at least one of (i), (ii) or (iii) is true. Let $\mathfrak{S}$ be the set of all A-sequences of elements of B and $\varphi: \mathfrak{S} \to B$ be the mapping such that for $g \in \mathfrak{S}$

$\varphi(g)$ is the least element of B if $B = \operatorname{Im} g$;

$\varphi(g)$ is the least element of $B \setminus \operatorname{Im} g$ if $B \neq \operatorname{Im} g$.

Then by the theorem of transfinite recursion 8.18 we have a unique mapping $f: A \to B$ such that $f(a) = \varphi(f^a)$ for each $a \in A$. For the least element a_0 of A, the above condition assures that $f(a_0)$ is the least element b_0 of B. We are interested now in the existence of further elements of A whose image under f is b_0 and consider the set $A' = \{a \in A: a \neq a_0 \text{ and } f(a) = b_0\}$. For each element a of A', we must have $\operatorname{Im} f^a = B$. Now two mutually exclusive cases arise:

Case 1. $A' \neq \emptyset$. Let a be the least element of A'. Now we shall prove that $f^a: s(a) \to B$ is an isomorphism, i.e. (iii) of the theorem holds. From the above discussion, f^a is surjective and it remains to be shown that f^a is strictly increasing, and hence injective. Let x, z be two elements of the segment $s(a)$ and $x < z$. Then $f^a(x) \leqq f^a(z)$ follows from the definition of f and the assumption that $f(z) \neq b_0$. Furthermore $f^a(x) \in \operatorname{Im} f^z$ and $f^a(z) \notin \operatorname{Im} f^z$. Therefore $f^a(x) \neq f^a(z)$.

Case 2. $A' = \emptyset$. Using the same argument as before, we see that $f: A \to B$ is a strictly increasing mapping, and hence injective. Therefore either (i) or (ii) of the theorem holds once we prove that $\operatorname{Im} f$ is a segment of B. Let $f(a) = b$ be an element of $\operatorname{Im} f$. For each $y \in B$ such that $y < b$, the subset $\{x \in A: y \leqq f(x)\}$ of A is non-empty, and hence has a least element, say x. Then $y \leqq f(x)$. On the other hand f is strictly increasing, therefore y is an upper bound for $\operatorname{Im} f^x$ and it follows from the definition of f that $f(x) \leqq y$. Therefore $f(x) = y$ and hence $v \in \operatorname{Im} f$ showing that $\operatorname{Im} f$ is a segment of B. ∎

In the theory of equipotence of sets in Section **7A**, we have asked if it is true that $A \lesssim B$ or $B \lesssim A$ for any two sets A and B. Since isomorphisms are bijective mappings, the comparability theorem 8.19 gives a partial answer to the question; that is, if A and B are well-ordered, then they are comparable by equipotence. In the next section, we shall discuss the possibility of well-ordering sets and arrive at a complete affirmative answer.

D. The well-ordering principle and its equivalences

The importance of the theory of well-ordered sets is enormously enhanced by the possibility of transferring their properties to any set, ordered or not. This is based on the *well-ordering principle*, which contends that for any set there is an order relation with respect to which the set in question is well-ordered; or briefly, any set can be well-ordered. It turns out that this principle is closely connected with, in fact equivalent to, the axiom of choice, which was postulated earlier to ensure the the non-emptiness of the cartesian product of any non-empty family of non-empty sets. The well-ordering principle was first announced by CANTOR, but he was unable to give a satisfactory proof of it, and the matter became very controversial in his time. During the subsequent years E. ZERMELO[1] and then others gave a number of equivalent formulations of the principle and later it was shown that this principle is consistent with the usual axioms of set theory, provided that they are consistent among themselves. Some of these equivalent formulations, which have a great many applications in mathematics, will be the main subject of our discussion in this section.

THEOREM 8.20. *Each of the following statements is equivalent to the axiom of choice.*

The well-ordering principle. *Each set can be well-ordered.*

KURATOWSKI's lemma. *Every chain of an ordered set is included in a maximal (with respect to inclusion) chain.*

ZORN's lemma. *If an ordered set is* inductively ordered *(i.e. there is an upper bound for every chain), then it contains a maximal element.*

The proof of this theorem is given in four parts.

The axiom of choice $\Rightarrow$ *the well-ordering principle.*

[1] E.g. E. Zermelo: Beweis, dass jede Menge wohlgeordnet werden kann, *Mathematische Annalen* **59**(1904) and Neuer Beweis für die Möglichkeit einer Wohlordnung, *Mathematische Annalen* **65**(1909).

PROOF. The principle is trivial for the empty set. Let A be a non-empty set, φ a choice function of the set A. We shall consider non-empty subsets D of A satisfying the following conditions:

(i) D admits an order relation $\leqq$ so that $(D, \leqq)$ is well-ordered, and

(ii) for each element d of D, we have $\varphi(A \setminus s(d)) = d$.

These are called the φ-sets of the set A. We notice that the condition (ii) above expresses the requirement that the choice function φ of A always selects from $A \setminus T$ the element immediately following the proper segment T of D. Moreover it is clear that the singleton $\{\varphi(A)\}$ of A is trivially a φ-set which is included in every φ-set. The proof of this part is now given in three stages.

(a) Let D and E be φ-sets, $\leqq$ and $\leqslant$ respectively their order relations for which (i) and (ii) hold. We shall show that one of these two sets is a segment of the other. Consider now subsets F of $D \cap E$ such that

(iii) F is a segment of $(D, \leqq)$, and

(iv) F is a segment of $(E, \leqslant)$.

Then the union G of all such subsets is again a subset of $D \cap E$, satisfying (iii) and (iv). Now G cannot be a proper segment of $(D, \leqq)$ and at the same time a proper segment of $(E, \leqslant)$, for otherwise the subset $G' = G \cup \{\varphi(A \setminus G)\}$ of $D \cap E$ would satisfy (iii) and (iv) and include G as a proper subset. Therefore $F = D$ or $F = E$. In the former case, $\leqq$ is the order relation induced by $\leqslant$ and D is a segment of E; in the latter case, E is a segment of D. Thus we have shown that of any two φ-sets, one is a segment of the other.

(b) Let U be the union of all φ-sets of A. We shall show here that U is a φ-set. Since U contains at least the element $\varphi(A)$ of A, it is non-empty. We shall now define an order relation $\leqq$ in U to make it into a φ-set. By (a) any two elements x and y of U belong to some φ-set D of A. Therefore we can define $x \leqq y$ or $y \leqq x$ in U according to $x \leqq y$ or $y \leqq x$ in D. Then by (a) again, the order relation $\leqq$ in U is well-defined; moreover each φ-set of A is a segment of U. Since the descending chain condition holds in each φ-set of A, it necessarily holds in U, and hence U is well-ordered. Finally for each $u \in U$, u defines the same segment in U as in any φ-set containing u. Therefore (ii) holds in U and U is a φ-set of A.

(c) The well-ordering principle is established once we show that $U = A$. Assuming that $A \setminus U$ is non-empty, we would have a φ-set $U' = U \cup \{\varphi(A \setminus U)\}$ which includes U as a proper subset. This is

clearly impossible; therefore $A = U$ and A is well-ordered by $\leqq$ defined in (b). ∎

The well-ordering principle ⇒ Kuratowski's lemma.

PROOF. Let C be a chain in an ordered set $(A, \leqq)$ and D be the set of all elements of $A \setminus C$ which are $\leqq$-comparable with each element of C. If D is empty, then C is a maximal chain of A and nothing remains to be shown. In the case where D is non-empty, Kuratowski's lemma is established if we can construct a maximal chain E in $(D, \leqq)$.

Since we assume the well-ordering principle, we may use the method of construction by transfinite induction to find E. Let $\leqslant$ be an order relation which well-orders D. In the well-ordered set $(D, \leqslant)$, we consider the set $\mathfrak{S}$ of all D-sequences of elements of D (here segments and least elements are always those with respect to $\leqslant$). If $g : T \to D$ is a D-sequence, then we say that g is of the *first kind* if $T = s(d)$ is a proper segment of D and d is $\leqq$-comparable with each element of Im g, and that g is of the *second kind* otherwise. We denote by d_0 the least element of D, and define a mapping $\varphi : \mathfrak{S} \to D$ where for all $g \in \mathfrak{S}$

$$\varphi(g) = \begin{cases} d & \text{if } g : s(d) \to D \text{ is of the first kind} \\ d_0 & \text{if } g \text{ is of the second kind.} \end{cases}$$

By transfinite recursion, there exists a unique mapping $f : D \to D$ such that for all $d \in D$

$$f(d) = \varphi(f^d).$$

We observe that $f(d) = d$ or $f(d) = d_0$ according to whether f^d is of the first or the second kind respectively. Now it is not difficult to verify that the subset $E = \text{Im } f$ of D has the following properties:

(a) $d_0 \in E$;

(b) for each $d \in D$, $d \in E$ iff d is $\leqq$-comparable with each element of $s(d) \cap E$ (where $s(d)$ is the segment constructed in the well-ordered set $(D, \leqslant)$).

It follows from (a) and (b) that the set E is a maximal chain in the ordered set $(D, \leqq)$.

Therefore $C \cup E$ is a maximal chain in the ordered set $(A, \leqq)$ that includes the chain C as a subset. ∎

Kuratowski's lemma ⇒ Zorn's lemma.

PROOF. Note that the empty set with the trivial order relation is not an inductively ordered set. Now let $(A, \leqq)$ be a non-empty, inductively

ordered set and x an arbitrary element of A. Then $\{x\}$ is a chain in A and is therefore included in a maximal chain C of A. If m is an upper bound of C and $m \leqq y$ for some $y \in A$, then $D = C \cup \{y\}$ is a chain of A which includes both $\{x\}$ and C as subsets. Therefore $D = C$ and hence $y \in C$. Now $m = y$ follows from the assumption $m \leqq y$ and the assumption that m is an upper bound of C. Therefore m is a maximal element of A. ∎

Zorn's lemma ⇒ the axiom of choice.

PROOF. Let $(A_i)_{i \in I}$ be a non-empty family of non-empty sets, and $\mathscr{F}$ be the set of ordered pairs of the form (J,x) where $J \subset I$ and $x \in \prod_{i \in J} A_i$. Then $\mathscr{F}$ is clearly non-empty, since we can always find such ordered pairs (J,x) where J is a singleton in I. We introduce an order relation $\leqq$ in $\mathscr{F}$ in the following manner. For any (J,x) and (K,y) in $\mathscr{F}$,

$$(J,x) \leqq (K,y) \text{ if and only if } J \subset K \text{ and } x_j = y_j \text{ for all } j \in J.$$

Our next step is to show that $\mathscr{F}$ is inductively ordered. If $\mathscr{C}$ is a chain in $\mathscr{F}$, we denote by U the union of the first coordinates of all elements of $\mathscr{C}$. Then $z \in \prod_{i \in U} A_i$ can be found in the following way: for each $u \in U$, $u \in J$ for some $(J,x) \in \mathscr{C}$, and we put $z_u = x_u$. Since $\mathscr{C}$ is a chain, z is well-defined and (U,z) is an upper bound of $\mathscr{C}$. By our assumption, $\mathscr{F}$ has a maximal element (M,x). Our final step is then to show that $M = I$. Assume that there exists an index j of I which does not belong to M. Since A_j is non-empty, we have $x_j \in A_j$. Then (M', x') where $M' = M \cup \{j\}$ and

$$x'_i = x_i \text{ for } i \in M \text{ and } x'_j = x_j$$

is clearly an element of $\mathscr{F}$ greater than (M,x), contradicting the maximality of (M,x). ∎

Since we have included the axiom of choice in our system of axioms, the well-ordering principle is a theorem in our theory. Therefore given any two sets A and B, we may assume that they are well-ordered sets. By the comparability theorem 8.19, one of them is isomorphic to a segment of the other and hence one of them is equipotent to a subset of the other. Applying the Schröder-Bernstein Theorem 7.2, we get the law of trichotomy of equipotence.

COROLLARY 8.21. *Given any two sets A and B, it is either $A \prec B$ or $A \sim B$ or $B \prec A$.*

E. Exercises

1. Let R be the set of all real numbers, Q the set of all rational numbers, $A = \{x \in Q : x < \sqrt{3}\}$ and $B = A \cup \{2\}$. Which of the sets R, Q and B contains a supremum of A?

2. Prove that each finite ordered set contains a maximal element.

3. Let A and B be ordered sets. Show that the relation in $A \times B$ such that for all (a,b), (a', b') of $A \times B$,

$$(a,b) \leqq (a', b') \text{ if and only if } a \leqq a' \text{ and } b \leqq b',$$

is an order relation in $A \times B$. This is called the *product order relation.*

4. Generalize the product order relation of the previous exercise to obtain an order relation in $\prod_{i \in I} A_i$ where $(A_i)_{i \in I}$ is a family of ordered sets.

5. Let A and B be ordered sets. Prove that in $\mathrm{Map}(A,B)$, the relation such that for any $f, g \in \mathrm{Map}(A,B)$,

$$f \leqq g \text{ if and only if } f(x) \leqq g(x) \text{ for all } x \in A$$

is an order relation of $\mathrm{Map}(A,B)$.

6. Let A and B be ordered sets, and $B^A = \prod_{a \in A} B_a$ where $B_a = B$ for all $a \in A$. Prove that the mapping $\varphi : B^A \to \mathrm{Map}(A,B)$ such that for all $x = (x_a)_{a \in A}$ of B^A

$$\varphi(x) = f \text{ where } f(a) = x_a \text{ for all } a \in A,$$

is a bijective mapping. Furthermore, prove that if B^A is ordered by the product order relation and $\mathrm{Map}(A,B)$ is ordered by the order relation defined in the previous exercise, then φ is an isomorphism.

*7. Let A and B be ordered sets and $\mathrm{In}(A,B)$ the set of all increasing mappings of A into B. When $\mathrm{In}(A,B)$ is ordered by the induced order relation of $\mathrm{Map}(A,B)$ as defined in Exercise 5, show that

(a) $\mathrm{In}(A, B \times C)$ is isomorphic to $\mathrm{In}(A,B) \times \mathrm{In}(A,C)$,

(b) $\mathrm{In}(A \times B, C)$ is isomorphic to $\mathrm{In}(A, \mathrm{In}(B,C))$,

where the cartesian products are ordered by the product order relations.

8. Let A be an inductively ordered set. Show that for each element x of A there exists a maximal element m of A such that $x \leqq m$.

9. Let A be an ordered set, and $\mathfrak{I}$ the set of all subsets X of A such that any two distinct elements of X are incomparable. Show that the following statements are true:

(a) the relation in $\mathfrak{I}$ such that for any X, Y of $\mathfrak{I}$,

$$X \leqq Y \text{ if and only if for each } x \in X \text{ there exists } y \in Y \text{ such that } x \leqq y,$$

is an order relation in $\mathfrak{I}$;

(b) A is totally ordered if and only if $\mathfrak{I}$ is totally ordered, in which case A and $\mathfrak{I}$ are isomorphic; and

(c) if A is inductively ordered, then $\mathfrak{I}$ contains a greatest element.

10. Let $A = \{a,b,c\}$ be a set consisting of exactly three elements. Find an order relation in A which well-orders A. How many such order relations are there in A?

11. Prove that all finite totally ordered sets are well-ordered sets.

12. Let $(A, \leqq)$ be a well-ordered set. Prove that if $(A, \geqq)$ is also a well-ordered set, then A is finite.

13. Prove that if every countable subset of a totally ordered set A is well-ordered, then A is well-ordered.

14. Let A be a well-ordered set and B a subset of A. Show that for each isomorphism $\varphi: A \to B$, $a \leqq \varphi(a)$ for all $a \in A$.

15. Let A and B be well-ordered sets. Prove that if $\varphi: A \to B$ and $\psi: A \to B$ are isomorphisms, then $\varphi = \psi$.

16. Show that a well-ordered set is not isomorphic to any of its proper segments.

*17. A set $\mathfrak{N}$ of sets is called a *nest* if and only if $\mathfrak{N}$ is totally ordered by inclusion. When $\mathfrak{N} \subset \mathfrak{A}$ and $\mathfrak{N}$ is a nest, then we say that $\mathfrak{N}$ is a nest in $\mathfrak{A}$. Prove that each of the following statements, in which maximal and minimal elements are always those with respect to the order relation by inclusion, is equivalent to the axiom of choice:

HAUSDORFF'S MAXIMAL PRINCIPLE. If $\mathfrak{A}$ is a set of sets and $\mathfrak{N}$ a nest in $\mathfrak{A}$, then there is a maximal nest $\mathfrak{M}$ in $\mathfrak{A}$ so that $\mathfrak{M} \supset \mathfrak{N}$.

MAXIMAL PRINCIPLE. Let $\mathfrak{A}$ be a set of sets. If for every nest $\mathfrak{N}$ in $\mathfrak{A}$, there is an element of $\mathfrak{A}$ which includes every element of $\mathfrak{N}$, then $\mathfrak{A}$ contains a maximal element.

MINIMAL PRINCIPLE. Let $\mathfrak{A}$ be a set of sets. If for every nest $\mathfrak{N}$ in $\mathfrak{A}$, there is an element of $\mathfrak{A}$ that is included in every element of $\mathfrak{N}$, then $\mathfrak{A}$ contains a minimal element.

TUKEY'S LEMMA. Let $\mathfrak{A}$ be a set of sets. If $\mathfrak{A}$ is of finite character (i.e. a set A belongs to $\mathfrak{A}$ if and only if every finite subset of A belongs to $\mathfrak{A}$), then $\mathfrak{A}$ contains a maximal element.

ZERMELO'S POSTULATE. If $\mathfrak{A}$ is a set of disjoint non-empty sets, then there is a set C so that $A \cap C$ is a singleton for each element A of $\mathfrak{A}$.

CHAPTER 9

ORDINAL NUMBERS AND CARDINAL NUMBERS

A. Ordinal numbers

In Chapter 7 we have seen that each finite set A is equipotent to a unique natural number n. On the other hand the natural number n is a well-ordered set with respect to the usual order relation and no matter how the set A is well-ordered, A and n are isomorphic well-ordered sets. For this reason we may regard natural numbers as standard finite well-ordered sets in the sense that

(NA) *every finite well-ordered set is isomorphic to a unique natural number.*

This means, therefore that (i) every finite well-ordered set is isomorphic to a standard one, and (ii) any two finite well-ordered sets are isomorphic if and only if they are isomorphic to one and the same standard one.

In the last chapter we saw that any set can be well-ordered. Thus it is desirable to construct standard well-ordered sets, called *ordinal numbers*, with a property analogous to (NA) above. There are a number of equivalent definitions of ordinal numbers (see Exercise 2); here we shall adopt the one given by JOHN VON NEUMANN.

DEFINITION 9.1. *An* ordinal number *is a well-ordered set* α *such that for each* $\xi \in \alpha$, $s(\xi) = \xi$ *where* $s(\xi)$ *is the segment of elements of* α *preceding* ξ.

EXAMPLE 9.2. Each natural number together with the usual order relation is a finite ordinal number and *vice versa.*

EXAMPLE 9.3. In conformity with the usual notation used for ordinal numbers, we denote by ω the well-ordered set N of all natural numbers; ω is clearly an ordinal number. Consider the set $\omega^+ = \omega \cup \{\omega\}$ and define an order relation for ω^+ as follows. For any two elements α and β of ω, $\alpha \leqq \beta$ in ω^+ if and only if $\alpha \leqq \beta$ in ω, and $\alpha < \omega$. Then ω^+ is an ordinal number, for if $\xi \in \omega^+$, then it is either $\xi \in \omega$ or $\xi = \omega$. In both cases we have $\xi = s(\xi)$. Similarly we have ordinal numbers ω^{++}, ω^{+++} and so forth.

EXAMPLE 9.4. In general, if α is an ordinal number, then the successor set α^+, ordered in the obvious manner, is an ordinal number that contains α and hence includes it as a proper segment.

B. General properties of ordinal numbers

According to the discussion at the beginning of preceding section, the ordinal numbers are to serve as standard well-ordered sets to be compared with well-ordered sets in such a way that

(OR) *every well-ordered set is isomorphic to a unique ordinal number.*

Now we shall see in this section that the ordinal numbers defined in 9.1 are the right kind of standard well-ordered sets. This will be shown in 9.11 but first some preparations are necessary.

THEOREM 9.5. *For any ordinal number α the following statements hold:*

(a) *if $\alpha \neq 0$, then 0 is the least element of α;*

(b) *if $\xi \in \alpha$, then ξ is an ordinal number;*

(c) *α is the set of its proper segments; and*

(d) *if $\xi \in \alpha$ and $\zeta \in \xi$, then $\zeta \in \alpha$.*

PROOF. (a) The least element of α determines the segment $\emptyset$; therefore $0 = \emptyset$ is the least element of α.

(b) For each $\zeta \in \xi$ the segment of elements of α preceding ζ is equal to the segment of elements of ξ preceding ζ. Therefore $s(\zeta) = \zeta$ in ξ, and ξ is an ordinal number.

(c) and (d) are immediate consequences of 9.1. ∎

The property 9.5(d) expresses the fact that ordinal numbers are *transitive sets*.

The next theorem is to make sure that we do not have too many standard sets.

THEOREM 9.6. *Isomorphic ordinal numbers are equal.*

PROOF. Let α and β be ordinal numbers and $f: \alpha \to \beta$ an isomorphism. The theorem is established if we can prove that f is the identity mapping i_α of the set α. Let $\mathcal{F}$ be the set of all segments S of α such that $f|S$ is equal to the inclusion mapping i_S of S into β. Since 0 is the least element of both α and β and f is increasing, we necessarily have $f(0) = 0$, and hence the segment $\{0\}$ of α belongs to $\mathcal{F}$. We order the non-empty set $\mathcal{F}$ by inclusion, and then $\mathcal{F}$ is easily seen to be inductively ordered. By Zorn's lemma, $\mathcal{F}$ has a maximal element, say M. Then it only remains to be proved that $M = \alpha$. Assume that $M \neq \alpha$ and let ξ be the least element of $\alpha \setminus M$, ζ the least element of $\beta \setminus M$. Then $f(\xi) = \zeta$, since f is increasing. On the other hand, $\xi = \zeta$, since they determine the same segment M in α and β respectively. Therefore $M' = M \cup \{\xi\}$ belongs to $\mathcal{F}$ and is larger than M, contradicting the definition of M. ∎

From Theorem 9.6 above and the comparability of well-ordered sets we obtain the following corollary:

COROLLARY 9.7. *For any two ordinal numbers α and β, one and only one of the following statements holds:*

(i) *α is a proper segment of β, i.e. $\alpha \in \beta$;*

(ii) *$\alpha = \beta$;*

(iii) *β is a proper segment of α, i.e. $\beta \in \alpha$.*

Therefore ordinal numbers may be compared with each other thus:

$\alpha \leqq \beta$ if and only if α is segment of β, i.e. $\alpha \in \beta$ or $\alpha = \beta$.

Obviously the two conditions of 8.1 are satisfied and we shall speak, for convenience, of the above symbol $\leqq$ as the *canonical order relation of ordinal numbers*, even though all the ordinal numbers do not form a set, as will be shown in 9.11. Rewriting 9.7, we have the familiar *law of trichotomy: for any two ordinal numbers α and β it is either* (i) $\alpha < \beta$ *or* (ii) $\alpha = \beta$ *or* (iii) $\beta < \alpha$.

The elementary properties of ordinal numbers given in 9.5 (a), (c) and (d) may be reformulated in terms of the canonical order relation of ordinal numbers as follows:

(a) 0 is the least ordinal number;

(c) α is the set of all ordinal numbers β such that $\beta < \alpha$;

(d) the canonical order relation of ordinal numbers is transitive.

It follows from these properties that the order relation in any ordinal number (in particular the usual order relation of natural numbers) is induced by the canonical order relation of ordinal numbers.

For sets of ordinal numbers, we have the following theorem and its corollary, corresponding to familiar properties of sets of natural numbers:

THEOREM 9.8. *Let A be a set of ordinal numbers. Then A is well-ordered by the canonical order relation of ordinal numbers.*

PROOF. Let $(\alpha_i)_{i=0,1,2,\ldots}$ be a descending chain of elements of A. Then $(\alpha_i)_{i=1,2,3,\ldots}$ is a descending chain of elements of the well-ordered set α_0 and therefore it must be stationary. Hence the first descending chain must also be stationary. ∎

COROLLARY 9.9. *A set A of ordinal numbers is an ordinal number if and only if, for each $\alpha \in A$, A contains also all ordinal numbers $\beta < \alpha$.*

THEOREM 9.10. *Let A be a set of ordinal numbers. Then there is a least ordinal number σ such that $\alpha \leqq \sigma$ for all $\alpha \in A$.*

PROOF. Each element of A is a set of ordinal numbers that satisfies the condition of 9.9. Therefore the union σ of all the elements of A is a set of ordinal numbers and as such σ satisfies the condition of 9.9. This proves that σ is an ordinal number. For each $a \in A$, we have $a \subset \sigma$. Therefore $a \leqq \sigma$. If β is an ordinal number with the property that $a \leqq \beta$ for all $a \in A$, then $a \subset \beta$ for all $a \in A$. This means that $\sigma \subset \beta$ and $\sigma \leqq \beta$. ∎

The ordinal number σ may be called the *supremum* of the set A. We notice that σ need not belong to A. Take for instance $A = \omega$, then $\sigma = \omega$ and $\sigma \notin \omega$.

Using the above theorem we shall show that all the ordinal numbers do not form a set. This is an immediate consequence of the following theorem:

THEOREM 9.11. *For any set A of ordinal numbers, there exists an ordinal number which is greater than each ordinal number of the set and consequently does not belong to A.*

PROOF. Obviously the successor σ^+ of the supremum σ of A is an ordinal number satisfying the requirement. ∎

Finally we shall show that the ordinal numbers have the property (OR) mentioned at the beginning of this section. To do so, we clearly need a very large reserve of ordinal numbers. Thus it involves the construction of sets to such an extent that a further axiom is necessary.

AXIOM OF SUBSTITUTION. *Let A be a set and $S(a,b)$ a statement function. If for each $a \in A$ all the objects b for which '$S(a,b)$' is true form a set, then there exist a set X and a mapping $F: A \to X$ whose value $F(a)$ at each $a \in A$ is exactly the set of all objects b for which '$S(a,b)$' is true.*

The above axiom permits us to construct the set $B = \operatorname{Im} F$ consisting of all $F(a)$. In other words, it permits us to substitute each $a \in A$ by $F(a)$, thus obtaining a set B.

THEOREM 9.12. *Every well-ordered set is isomorphic to a unique ordinal number.*

PROOF. The uniqueness is a direct consequence of 9.6. Let A be a well-ordered set. The proof of the existence of a unique ordinal number isomorphic to A can be made in two steps. First, we shall show by transfinite induction that for each element a of A, the segment $s(a)$ is isomorphic to a unique ordinal number. For the least element of A the statement above is trivial. Assume that for each $x < a$, $s(x)$ is isomorphic to a unique ordinal number. Under this induction assumption we can apply the axiom of substitution to the set $s(a)$ and the statement $S(x,\beta)$:

β is an ordinal number isomorphic to a proper segment of $s(x)$. Then we get a mapping F: $s(a) \to X$ (where X may be assumed to be a set of ordinal numbers without loss of generality) so that $F(x)$ is the ordinal number β isomorphic to $s(x)$. The mapping F is necessarily a strictly increasing mapping. Moreover $\alpha = \operatorname{Im} F$ is an ordinal number, since it satisfies the condition of 9.9. Therefore $s(a)$ is isomorphic to the ordinal number α; hence the proof by induction is complete.

Finally we can apply the axiom of substitution once again to the set A and the statement $T(a,\alpha)$: α is an ordinal number isomorphic to a proper segment of $s(a)$. Then we get a mapping $G: A \to Y$ so that $G(a)$ is the ordinal number α isomorphic to the segment $s(a)$ of A. Using a similar argument, we see that A is isomorphic to the ordinal number $\gamma = \operatorname{Im} G$. ∎

Our main result 9.12 now justifies the following definition:

DEFINITION 9.13. *Let A be a well-ordered set. Then the* ordinal number $\operatorname{Ord}(A)$ *of A is the unique ordinal number isomorphic to A.*

To summarize: (i) every well-ordered set has an ordinal number and (ii) any two well-ordered sets are isomorphic if and only if they have one and the same ordinal number. Furthermore, a process of counting beyond the natural numbers can be carried out in an obvious manner by means of the ordinal numbers.

C. The arithmetic of ordinal numbers

1. Addition

The addition of ordinal numbers is defined as an extension of the addition of natural numbers. Consider the equation $2 + 3 = 5$. To obtain the well-ordered set 5, we may first replace the well-ordered set $3 = \{0, 1, 2\}$ by an isomorphic ordered set disjoint from $2 = \{0, 1\}$, say $3' = \{2, 3, 4\}$, and then form the union $2 \cup 3' = \{0, 1, 2, 3, 4\}$ which is our sum 5.

More generally, let α and β be ordinal numbers. Consider the set $\alpha' = \{(\xi,0): \xi \in \alpha\}$ and $\beta' = \{(\zeta,1): \zeta \in \beta\}$. They are obviously disjoint, and their union $C = \alpha' \cup \beta'$ is then well-ordered by the order relation defined as follows:

$$(\xi,0) < (\zeta,1) \text{ for all } \xi \in \alpha \text{ and } \zeta \in \beta$$

$$(\xi,0) < (\zeta,0) \text{ if and only if } \xi < \zeta$$

$$(\xi,1) < (\zeta,1) \text{ if and only if } \xi < \zeta$$

If γ is the ordinal number of the well-ordered set C, then we define $\alpha + \beta$ as γ.

We notice here that our addition of ordinal numbers may not be commutative when the summands are not finite. For example we have

$$0 + \alpha = \alpha + 0 = \alpha$$

whereas

$$1 + \omega = \omega, \quad \omega + 1 = \omega^+ \quad \text{and } \omega \neq \omega^+.$$

Unlike the addition of natural numbers developed in Chapter 6, our present addition can handle any number of summands at one time. From the above example, we rather expect that our sum will depend on the order of the summands in which they appear in the sum.

For convenience of formulation, we shall call a family a sequence if the index set of the family is an ordinal number. Let $(\alpha_\xi)_{\xi\in\lambda}$ be a sequence of ordinal numbers. We consider the disjoint sets $A_\xi = \{(\beta,\xi):\beta\in\alpha_\xi\}$ and their union U (called the *disjoint union* of the sets α_ξ). The set U is then well-ordered by the relation defined as follows:

$$(\beta,\xi) < (\gamma,\zeta) \quad \textit{if and only if} \quad \text{(a) } \xi < \zeta \textit{ or (b) } \beta < \gamma \textit{ when } \xi = \zeta.$$

The sum $\sigma = \sum_{\xi\in\lambda} \alpha_\xi$ of the sequence is defined as $\mathrm{Ord}(U)$.

2. *Multiplication*

The product $\alpha\beta$ of two ordinal numbers α and β may be defined as the sum of the sequence $(\alpha_\xi)_{\xi\in\beta}$ where $\alpha_\xi = \alpha$ for all $\xi\in\beta$. By iteration the products of a finite sequence of ordinal numbers can then be defined. However, we may also adopt a different approach. Let $(\alpha_i)_{i\in n}$ be a finite sequence of ordinal numbers and P the cartesian product of the sets α_i. For any two different elements f and g of P we say

$$f < g \textit{ if and only if there is } i\in n \textit{ such that } f_i < g_i \textit{ and } f_j = g_j \textit{ for all } j > i.$$

Then P is well-ordered by the above relation and the product $\prod_{i\in n} \alpha_i = \alpha_0 \ldots \alpha_{n-1}$ is defined as $\mathrm{Ord}(P)$.

It is not difficult to see that the two approaches always give the same results. For this we have only to verify that the disjoint union of the family $(\alpha_\xi)_{\xi\in\beta}$ (where $\alpha_\xi = \alpha$ for all $\xi\in\beta$) and the cartesian product of the family $(\gamma_i)_{i=0,1}$ (where $\gamma_0 = \alpha$ and $\gamma_1 = \beta$) with their order relations are isomorphic well-ordered sets.

Here again the multiplication is not commutative. For example, we have

$$0\alpha = \alpha 0 = 0$$

and

$$1\alpha = \alpha 1 = \alpha;$$

whereas

$$2\omega = \omega \quad \text{and} \quad \omega 2 \neq \omega.$$

Here $\varphi: 2\omega \to \omega$ defined by $\varphi(i,n) = 2n + i$ is an isomorphism, hence $2\omega = \omega$; whereas ω is isomorphic to the proper segment $\omega \times \{0\}$ of $\omega 2$ preceding $(0,1)$.

3. Exponentiation

Now the exponentiation can be defined in the usual way as *repeated multiplication*. Given an ordinal number α and a finite ordinal number n, the power α^n is defined as the product $\prod_{i \in n} \alpha_i$ where $\alpha_i = \alpha$ for all $i \in n$.

D. Cardinal numbers

Ordinal numbers were introduced as standard sets for comparing well-ordered sets by means of isomorphisms. If order relation is not our primary concern, then our main tool for the comparison of sets, as in Chapter 6, is equipotence. Our standard sets in this case are called *cardinal numbers* and should be sets of such a kind that

(CA) *every set is equipotent to a unique cardinal number.*

From the results of Chapter 7, we expect that natural numbers are the only finite cardinal numbers as well as the only finite ordinal numbers. In this section we shall find sets of the character (CA). We know that (i) each such set can be well-ordered and hence isomorphic to an ordinal number and (ii) isomorphic well-ordered sets are equipotent sets, and therefore we may hope to find our standard sets among the ordinal numbers. Furthermore, if we can be sure that all ordinal numbers equipotent to a fixed one form a set, then this set is well-ordered by the canonical order relation of ordinal numbers, and in this case we may hope that the least element of this set will serve our purpose.

Let α be an ordinal number. Then by 7.3 the power set $\mathfrak{P}(\alpha)$ is a set that has the following properties:

(a) α is equipotent to a proper subset of $\mathfrak{P}(\alpha)$, and

(b) α is not equipotent to $\mathfrak{P}(\alpha)$.

By the well-ordering principle, we may well-order the set $\mathfrak{P}(\alpha)$. Let π be the ordinal number of the well-ordered set $\mathfrak{P}(\alpha)$ and consider the set $B = \{\beta \in \pi : \beta \sim \alpha\}$ where $\beta \sim \alpha$ means that the sets β and α are equipotent. If γ is an ordinal number such that $\gamma \sim \alpha$, then we have $\gamma < \pi$ and hence $\gamma \in \pi$. For otherwise it would imply that $\mathfrak{P}(\alpha)$ is equipotent to a subset of α, which is impossible. Therefore the set B is the set of all ordinal numbers that are equipotent to α. The following definition is now justified:

DEFINITION 9.14. *A* cardinal number *is an ordinal number* α *such that* $\alpha \leqq \beta$ *for all ordinal numbers* β *which are equipotent to* α.

In other words, α is the least element of the set of all ordinal numbers that are equipotent to it.

Clearly each natural number is a finite cardinal number and *vice versa*. Transfinite cardinal numbers are usually denoted by the first letter $\aleph$ (aleph) of the Hebrew alphabet.

The set N of all natural numbers is the least infinite ordinal number and as such it is denoted by ω. Therefore it is also a cardinal number; as such it is usually denoted by $\aleph_0$.

We must now show that the cardinal numbers defined in 9.14 are the standard sets we are looking for.

THEOREM 9.15. *Each set* A *is equipotent to a unique cardinal number. This is called the* cardinal number of the set A *and is denoted by* $\mathrm{Card}(A)$.

PROOF. Let A be a set. Then carrying out a similar construction as before, we obtain a non-empty set B of all ordinal numbers equipotent to A. The least (with respect to the canonical order relation of ordinal numbers) element α of B is obviously the unique cardinal number equipotent to A. ∎

It follows from the above theorem that (i) *every set has a cardinal number* and (ii) *two sets are equipotent if and only if they have the same cardinal number*.

The canonical order relation of ordinal numbers induces in a unique way a canonical order relation of cardinal numbers; and with respect to this order relation, the usual law of trichotomy holds. We can show similarly that corresponding to 9.11 all the cardinal numbers do not form a set.

Finally the addition and multiplication of cardinal numbers can be defined in an obvious manner. Let $\mathscr{F} = (\alpha_i)_{i \in I}$ be a family of cardinal numbers, P the cartesian product of the family $\mathscr{F}$ and U the union of the disjoint sets $A_i = \{(\xi, i) : \xi \in \alpha_i\}$. Then the *sum* of the family $\mathscr{F}$ is defined as the cardinal number of U and the *product* of the family $\mathscr{F}$ is defined as the cardinal number of the cartesian product P. Now the cardinal numbers are ordinal numbers, and therefore we have two additions and two multiplications defined on them. Note that for cardinal numbers neither the two additions nor the two multiplications coincide. For instance, we have $\omega + 1 = \omega^+ (\neq \omega)$ and $\omega\omega = \omega^2 (\neq \omega)$, whereas $\aleph_0 + 1 = \aleph_0$ and $\aleph_0 \aleph_0 = \aleph_0$. However, the following equalities hold for any ordinal numbers α and β:

$$\mathrm{Card}(\alpha + \beta) = \mathrm{Card}(\alpha) + \mathrm{Card}(\beta)$$

$$\mathrm{Card}(\alpha\beta) = \mathrm{Card}(\alpha)\mathrm{Card}(\beta).$$

E. Cantor's continuum hypothesis

As a cardinal number, the set N of all natural numbers is denoted by $\aleph_0$. This, then, is the least transfinite ordinal number and also the least transfinite cardinal number. Analogously, the least uncountable ordinal number Ω (see Exercise 1) is a cardinal number; and as such it is usually denoted by $\aleph_1$. Comparing these two cardinal numbers by the canonical order relation of cardinal numbers, we see that $\aleph_1$ is the least cardinal number strictly following $\aleph_0$.

In Chapter 7, we have seen that the set N of all natural numbers is equipotent to a subset of its power set $\mathfrak{P}(N)$ but is not equipotent to $\mathfrak{P}(N)$ itself. For the cardinal numbers of these two sets, we have $\aleph_0 < \mathrm{Card}(\mathfrak{P}(N))$. But the cardinal number of the set $\mathfrak{P}(N)$ can be obtained by exponentiation; namely, $\mathrm{Card}(\mathfrak{P}(N)) = 2^{\aleph_0}$, since

$$\mathfrak{P}(N) \sim \mathrm{Map}(N,2) \sim \prod_{i \in N} a_i$$

where $a_i = 2$ for all $i \in N$. Note that $2^{\aleph_0}$ is also the cardinal number of the set R of all real numbers.

It follows then that we have $\aleph_1 \leqq 2^{\aleph_0}$. In his famous *continuum hypothesis*, CANTOR conjectured that $\aleph_1 = 2^{\aleph_0}$. In 1938, K. GÖDEL proved the consistency of the continuum hypothesis with the usual axioms of set theory, i.e., that the equation $\aleph_1 = 2^{\aleph_0}$ is non-contradictory.[1] In 1964, PAUL J. COHEN[2] and PETR VOPĚNKA[3] proved, quite independently of each other, that the continuum hypothesis is independent of the usual axioms of set theory; i.e. that the inequality $\aleph_1 \neq 2^{\aleph_0}$ is non-contradictory.

[1] K. Gödel (1938): The consistency of the axiom of choice and the generalized continuum-hypothesis with the axioms of set theory, *Annals of Mathematics Studies* **3.** Princeton.

[2] Paul J. Cohen (1964): The independence of continuum hypothesis I and II, *Proceedings of the National Academy of the United States of America* **50** and **51.**

[3] Petr Vopěnka (1964): The independence of continuum hypothesis, *Commentationes Mathematicae Universitatis Carolinae* **5,** suppl. I. Prague.

F. Exercises

1. Denote by Ω the least uncountable ordinal number, and by Ω^+ the successor of Ω. Show that
   (a) Ω is the greatest element of Ω^+,
   (b) for each $\alpha \in \Omega$, the segment $s(\alpha)$ is a countable set, and
   (c) if A is a countable subset of Ω^+ and $\Omega \notin A$, then the supremum of A is less than Ω.

*2. A set $\mathfrak{A}$ is said to be transitive if and only if the following condition is satisfied:

   [T] if $C \in B$ and $B \in \mathfrak{A}$, then $C \in \mathfrak{A}$.

   Prove that J. VON NEUMANN's Definition (9.1) of ordinal numbers is equivalent to each of the following definitions of ordinal numbers:

   E. ZERMELO's Definition. An ordinal number is a set $\mathfrak{A}$ of sets such that
   (a) $\mathfrak{A} = \emptyset$ or $\emptyset \in \mathfrak{A}$,
   (b) for each $X \in \mathfrak{A}$, it is either $X^+ = \mathfrak{A}$ or $X^+ \in \mathfrak{A}$,
   (c) for each subset $\mathfrak{B}$ of $\mathfrak{A}$ it is either $\bigcup_{X \in \mathfrak{B}} X = \mathfrak{A}$ or $\bigcup_{X \in \mathfrak{B}} X \in \mathfrak{A}$.

   P. BERNAYS' Definition. An ordinal number is a set $\mathfrak{A}$ of sets such that
   (d) $\mathfrak{A}$ is transitive,
   (e) every transitive proper subset of $\mathfrak{A}$ is an element of $\mathfrak{A}$.

3. Let α and β be ordinal numbers. Show that $\alpha < \beta$ if and only if there exists an ordinal number $\gamma \neq 0$ such that $\alpha + \gamma = \beta$.

4. Let α and β be any ordinal numbers. Prove that
   (a) $\alpha = \sum_{\xi \in \alpha} \alpha_\xi$ where $\alpha_\xi = 1$ for all $\xi \in \alpha$.
   (b) $\alpha\beta = \sum_{\xi \in \beta} \alpha_\xi$ where $\alpha_\xi = \alpha$ for all $\xi \in \beta$.

5. Let α, β and γ be ordinal numbers. Prove that
   (a) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$,
   (b) $(\alpha\beta)\gamma = \alpha(\beta\gamma)$,
   (c) $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.

6. Prove that $(1 + 1)\omega < 1\omega + 1\omega$. This means that the distributive law $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$ does not hold in the arithmetic of ordinal numbers.

7. Let α be any ordinal number, and m and n any natural numbers. Prove that
   (a) $\alpha^{m+n} = \alpha^m \cdot \alpha^n$,
   (b) $\alpha^{mn} = (\alpha^m)^n$.

8. Prove that $2^2 \cdot (\omega^+)^2 < (2 \cdot \omega^+)^2$. This means that the exponential rule $(\alpha \cdot \beta)^n = \alpha^n \cdot \beta^n$ does not hold in the arithmetic of ordinal numbers.

9. Let α and β be ordinal numbers both greater than 1. Prove that
   (a) $\alpha + \beta \leqq \alpha\beta$,
   (b) if β is finite, then $\alpha\beta \leqq \alpha^{\beta}$.
10. Let α, β and γ be cardinal numbers. Prove that in the arithmetic of cardinal numbers
   (a) $\alpha + \beta = \beta + \alpha$,
   (b) $\alpha\beta = \beta\alpha$,
   (c) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$,
   (d) $(\alpha\beta)\gamma = \alpha(\beta\gamma)$,
   (e) $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$,
   (f) $\alpha^{\beta+\gamma} = \alpha^{\beta}\alpha^{\gamma}$,
   (g) $\alpha^{\beta\gamma} = (\alpha^{\beta})^{\gamma}$,
   (h) $(\alpha\beta)^{\gamma} = \alpha^{\gamma}\beta^{\gamma}$.
11. Let α and β be ordinal numbers. Prove that
   (a) $\mathrm{Card}(\alpha + \beta) = \mathrm{Card}(\alpha) + \mathrm{Card}(\beta)$,
   (b) $\mathrm{Card}(\alpha\beta) = \mathrm{Card}(\alpha) \cdot \mathrm{Card}(\beta)$.
12. Prove that if one of the non-zero cardinal numbers α and β is infinite, then $\alpha\beta = \alpha + \beta$.

*13. Prove that for any infinite cardinal number α and any natural number $n \geqq 1$, $\alpha^n = \alpha$.

14. Let α, β, γ and δ be cardinal numbers. Prove that if $\alpha \leqq \beta$ and $\gamma \leqq \delta$, then $\alpha\gamma \leqq \beta\delta$.

*15. Prove that if two families $(\alpha_i)_{i\in I}$ and $(\beta_i)_{i\in I}$ of cardinal numbers are such that $\alpha_i < \beta_i$ for all $i \in I$, then ZERMELO's inequality $\sum_{i\in I}\alpha_i < \prod_{i\in I}\beta_i$ holds.

SPECIAL SYMBOLS AND ABBREVIATIONS

The numbers and letters refer to the chapters and sections respectively

$X \wedge Y$	: conjunction	1B
$X \vee Y$	: disjunction	1C
$\sim X$	: negation	1D
$X \rightarrow Y$	: conditional	1E
$X \leftrightarrow Y$	: biconditional	1G
X eq Y	: equivalence	1H
T	: valid formula	1I
F	: negation of a valid formula	1I
'X'	: name	1J
'X' $\Rightarrow$ 'Y' (or $X \Rightarrow Y$): 'X' implies 'Y'		1K
'X' $\Leftrightarrow$ 'Y' (or $X \Leftrightarrow Y$): 'X' is equivalent to 'Y'		1K
$\forall$	: universal quantifier	1L
$\exists$	: existential quantifier	1L
$x \in A$	: x belongs to A	2A
$x \notin A$	: x does not belong to A	2A
$B \subset A$ (or $A \supset B$): A includes B		2C
$B \not\subset A$ (or $A \not\supset B$): A does not include B		2C
$\{x \in A: P(x)\}$: the set of all those x in A for which '$P(x)$' is true ...		2C
$\emptyset$	: the empty set	2C
$\{x: P(x)\}$: set of all x for which '$P(x)$' is true		2C
$\{a, b, c, \ldots, d\}$: set whose elements are $a, b, c, \ldots, d$ ...		2E
$\{a, b\}$	: unordered pair of a and b	2E
$\{a\}$	: singleton of a	2E
$A \cap B$	: intersection of A and B	2F
$A \cup B$	: union of A and B	2G
$E \setminus A$	: complement of A in E	2H
$\mathfrak{P}(A)$	: power set of A	2I
$\bigcup_{B \in \mathfrak{S}} B$	: union of all B in $\mathfrak{S}$	2J

LIST OF AXIOMS

The numbers and letters refer to the chapters and sections respectively

INDEX

The numbers and letters refer to the chapters and sections respectively