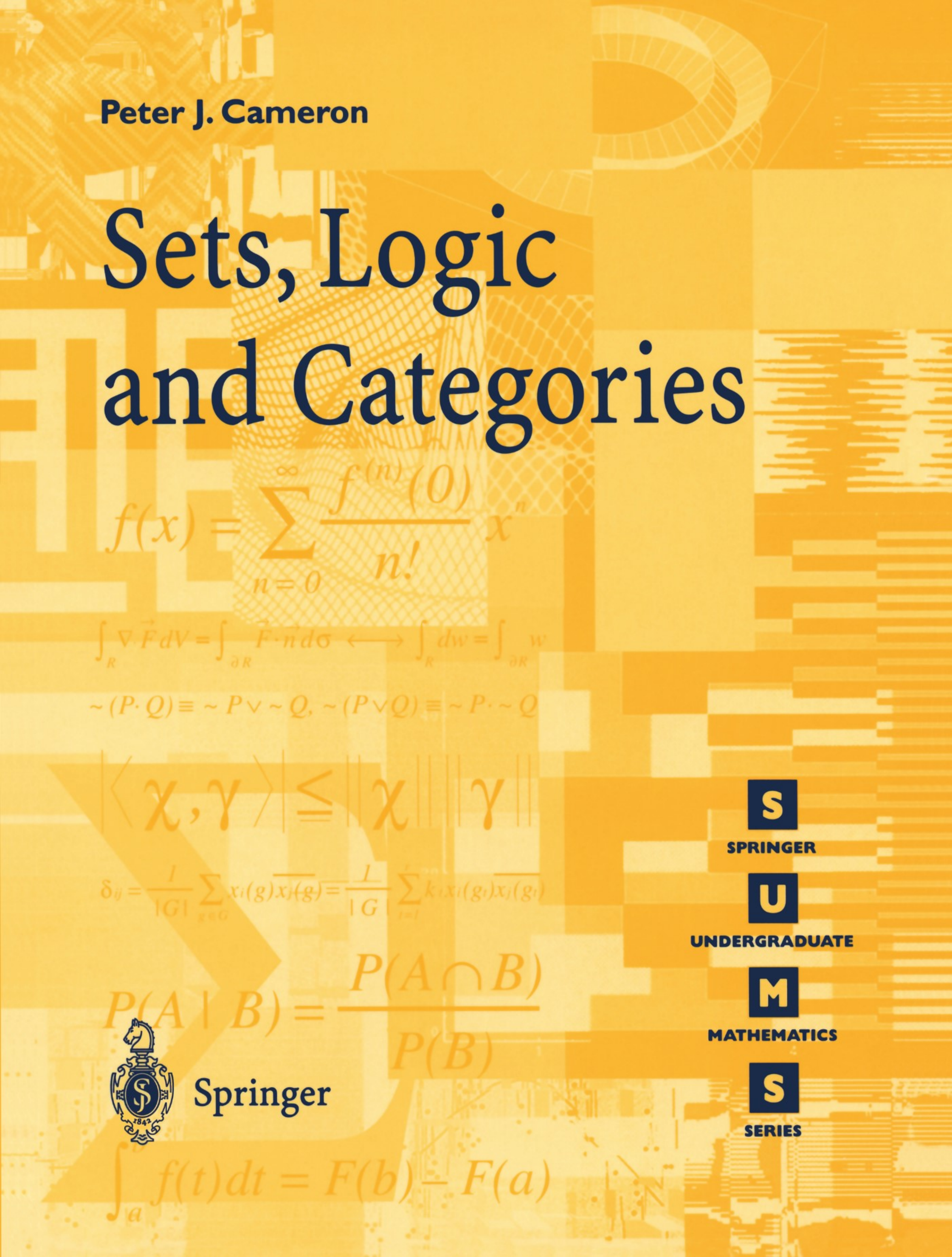
Peter J. Cameron

Sets, Logic and Categories


$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$
$\int_R \nabla \cdot \vec{F}\, dV = \int_{\partial R} \vec{F} \cdot \vec{n}\, d\sigma \longleftrightarrow \int_R dw = \int_{\partial R} w$
$\sim(P \cdot Q) \equiv \sim P \vee \sim Q, \sim(P \vee Q) \equiv \sim P \cdot \sim Q$
$|\langle \chi, \gamma \rangle| \leq \|\chi\| \|\gamma\|$
$\delta_{ij} = \frac{1}{|G|} \sum_{g \in G} x_i(g) \overline{x_j(g)} = \frac{1}{|G|} \sum_{t=1}^{r} k_t x_i(g_t) \overline{x_j(g_t)}$
$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
$\int_a^b f(t)\,dt = F(b) - F(a)$

Springer

S
SPRINGER
U
UNDERGRADUATE
M
MATHEMATICS
S
SERIES

Peter J. Cameron

Sets, Logic and Categories

Springer

1
Naïve set theory

> **set,** *n.*: aggregate, array, association, bale, batch, battery, body, bracket, bunch, bundle, cast, category, class, clump, cluster, collection, conglomerate, ensemble, family, genre, genus, group, grouping, ilk, kind, lot, mould, nature, order, pack, parcel, quantity, ring, section, sector, sort, species, style, type, variety.

Set theory is the most fundamental part of mathematics. The definition of almost any kind of mathematical object (a group, a ring, a vector space, a topological space, a Hilbert space ...) begins: a ⟨*thing*⟩ consists of a set, together with some extra structure in the form of operations, relations, subsets, sets of subsets, functions to the real numbers, or whatever. Also, as we will see, these operations, relations, etc. are themselves special kinds of sets.

However, it is very difficult to say what a set is. In the beginning, a set was simply a collection, class, or aggregate of objects, put together according to any rule we could imagine, or no rule at all. Of course this doesn't give a *definition* of a set, since if we try to explain what a collection, class or aggregate is, we find ourselves going round in circles. Later, we will see how this view is modified in the axiomatic approach, and the strengths and weaknesses of this approach.

In this chapter, after looking at Russell's Paradox as a cautionary tale, we review some of the basic notation and terminology about sets: subsets, unions and intersections, power sets, relations, functions, equivalence and order, finite and countable sets. The chapter ends by returning to Russell for another cautionary tale about the *Axiom of Choice.*

1.1 Handle with care

The foundations of mathematics have never been free of controversy. Set theory was developed by Georg Cantor in the late nineteenth century. But it soon became clear that the simple viewpoint that sets are just collections of elements, perhaps gathered together according to some rule, is untenable. *Russell's Paradox* demonstrates this. In fact, the logician Gottlob Frege was the first to develop mathematics on the foundation of set theory. He learned of Russell's Paradox while his work was in press, and wrote,

> A scientist can hardly meet with anything more undesirable than to have the foundation give way just as the work is finished. In this position I was put by a letter from Mr Bertrand Russell as the work was nearly through the press.

What then was this devastating communication?

It can be put in several related forms, of which the original is still the clearest. If arbitrary collections of objects form sets, we have to face the possibility that a set may be an element of itself: for example, the set consisting of all sets has this property. Most sets are not members of themselves, however. We may feel that it is slightly dangerous to have sets belonging to themselves, and restrict attention to 'good' sets which do not. Now Russell asked:

> *Let S be the set of all sets which are not members of themselves. Is S a member of itself?*

Either S is a member of itself, or it is not. Now, if S is a member of itself, then it is one of those sets which are not members of themselves, and so it is not a member of itself, which is a contradiction. On the other hand, if S is not a member of itself, then it satisfies the criterion of membership of S, which is also contradictory.[1]

Of other forms of this self-referential paradox, here are three.

One of the most famous is the *Liar Paradox*, invented by the Cretan philosopher and seer Epimenides. It is referred to by a number of classical authors, including the writer of the Pauline letter to Titus in the Christian Bible:

[1] Russell might have been pre-empted by the mediaeval theologian Thierry of Chartres, who postulated the 'Form of all other forms', which he interpreted as God. (Forms, in the Platonic sense, have something in common with the modern concept of sets.) His views were regarded as pantheistic and potentially heretical, but not as containing the seeds of a logical contradiction. See David Knowles, *The Evolution of Mediaeval Thought* [30].

> One of themselves [the Cretans], even a prophet of their own, said, the Cretans are alway liars, evil beasts, slow bellies.
>
> *Titus* 1, 12

If Epimenides said, 'All Cretans are liars', is his statement true or false?

An English adjective is called *autological* if it describes itself, and *heterological* if it does not. So, for example, 'short' and 'pentasyllabic' are autological, while 'long' and 'trisyllabic' are heterological. Indeed, most adjectives are heterological: consider 'hard', 'soft', 'red', 'blue'. (What about adjectives like 'fluffy' or 'gnarled'?)

Is 'heterological' heterological?

For the last paradox, we consider games between two players. We are not too precisely concerned with their structure; all that we require is that there are rules which determine whose move it is at any particular stage and what moves are allowed. When no move is possible, the game is over, though we allow the possibility that it ends in a draw.

A game in the above sense is called *well-founded* if any play of the game ends after finitely many moves (though there is not required to be a fixed upper bound for the number of moves). A trivial example is the following: the first player chooses a positive integer, and then the players take turns choosing positive integers, each smaller than the last one chosen. The game ends when no further choice is possible (that is, when the last number chosen is 1).

The *Hypergame* is played as follows. On the first move, the first player chooses any well-founded game. Then the players play that game, but with their roles swapped (so that the second player moves first in the chosen game). For example, if the first player chooses chess, then the second player takes the white pieces.[2]

Is the Hypergame well-founded? Obviously it is, since the games which can be chosen at the first move are required to be well-founded, and then the play of the Hypergame lasts only one move longer than the play of the chosen game.

But, since the Hypergame is well-founded, the first player may choose it on the first move. Then the second player starts playing the Hypergame by choosing a well-founded game. This player may also choose the Hypergame. Then the two players may continue choosing the Hypergame for ever, which contradicts the fact established above, that the Hypergame is well-founded.

[2] The rule that a game is drawn if the same position occurs for the third time guarantees that chess is well-founded. Littlewood [34] gives an upper bound of about $10^{10^{70.5}}$ for the number of possible chess games, comparable to the odds against a snowball surviving in Hell for a week.

Because of these paradoxes, the foundational subject of Set Theory must be set up with very great care. On the other hand, because of its importance, we must make the effort.

The matter is further complicated by *Gödel's Second Incompleteness Theorem*, an important result proved by Kurt Gödel in 1930. According to this theorem, we can never be sure that our set theory does not contain a contradiction lurking in its structure. Any formal theory which is strong enough to describe the natural numbers with their usual arithmetic, according to Gödel, cannot prove its own consistency. There are only two options: we can give a *relative consistency* proof, by proving the consistency of one subject within another. (This is what was done for non-Euclidean geometry, where models were constructed within Euclidean geometry.) Alternatively, we may simply work with the axiomatic theory and hope that any contradictions will come to light after a while. It is commonly accepted now that set theory does provide a secure foundation for mathematics.

In this chapter, we regard a set in the traditional way, but admit that not every collection of elements that can be imagined forms a set. Indeed, this is the positive conclusion we draw from Russell's Paradox:

Theorem 1.1

There is no set S such that $x \in S$ if and only if $x \notin x$.

Proof

If such a set S exists, then $S \in S$ if and only if $S \notin S$, which is a contradiction. So no such S can exist. □

We develop the notation and terminology of set theory, allowing constructions of sets only if they are 'limited' in some way so as to avoid this paradox. Ultimately this procedure is not satisfactory; we need hard and fast rules. This will be motivated and developed in the chapter on axiomatic set theory. We will go on to a few more special topics: countable sets, and the construction of the number systems.

1.2 Basic definitions

We write $x \in y$ for 'x is a member of y'. This *membership relation* is the basic relationship; when we come to the axioms, it will be the undefined relation which is their subject.

The relation of equality can be defined in terms of membership, by the *Principle of Extension*, first stated by Leibniz:

> Two sets are equal if they have the same members.

That is, $x = y$ if and only if, for all elements z, we have $(z \in x) \Leftrightarrow (z \in y)$.

The *empty set* is a set with no members. We say 'the empty set' rather than 'an empty set' because the Principle of Extension guarantees that there can't be more than one:

Theorem 1.2

There is only one empty set.

Proof

Let e_1 and e_2 be empty sets. Then for all elements z, the statements $z \in e_1$ and $z \in e_2$ are both false; so these statements are logically equivalent, and $e_1 = e_2$ by the Principle of Extension. □

We write $\varnothing$ for the empty set.

One important consequence of what we have accepted so far is the doctrine that

$$\boxed{\textit{Everything is a set.}}$$

For, if a is an object that is not a set, then a has no members, and so a is equal to the empty set by the Principle of Extension, contradicting the assumption that it is not a set.

Given this doctrine, it is incumbent on us to show that it is reasonable, by constructing the everyday objects of mathematics (the number 27, the group of rotations of a regular dodecahedron, the Hilbert space $L^2[0,1]$, and so on) as sets. Of course, it would be much too tedious to do all this. But we will say enough in Section 1.8 to illustrate how such a programme could be carried out, and we will construct the natural numbers (as special cases of the *ordinal numbers*) as particular sets in Chapter 2.

It should be mentioned that there are other foundations of set theory which allow so-called *urelements*, basic elements which are not sets. In such theories, the Principle of Extension has to be modified so that it applies only to sets.

The set x is a *subset* of the set y (in symbols, $x \subseteq y$) if every member of x is a member of y, that is, if $(z \in x) \Rightarrow (z \in y)$ holds for all elements z.

Comparing this with the Principle of Extension, we see that $x = y$ if and only if both $x \subseteq y$ and $y \subseteq x$ hold.

We write $x \subset y$ to mean that x is a *proper subset* of y, that is, $x \subseteq y$ but $x \neq y$.

The *power set* of a set x (in symbols, $\mathcal{P}\,x$) is the set of all subsets of x. That is,

$$\mathcal{P}\,x = \{y : (\forall z)(z \in y) \Rightarrow (z \in x)\}.$$

As the name suggests, the power set of x is a set. However, if x is a large and complicated infinite set, the notion of a subset of x may not be entirely straightforward, and so there is a certain amount of vagueness in the notion of the power set of x. This vagueness will allow the possibility of several different versions of axiomatic set theory.

If $x_1, x_2, \ldots, x_n$ are finitely many sets, then we can gather them all into a set, which we denote $\{x_1, x_2, \ldots, x_n\}$. This is a set which has just these n elements as members.

Let x be a set; recall that the members of x are also sets. The *union* of x, written $\bigcup x$, is the set consisting of all members of the members of x: that is,

$$\bigcup x = \{z : z \in y \text{ for some } y \in x\}.$$

This notation is a bit unfamiliar. You will be more familiar with the union of two sets x_1 and x_2. This consists of all elements lying in either x_1 or x_2:

$$x_1 \cup x_2 = \{z : z \in x_1 \text{ or } z \in x_2\}.$$

It is just a special case of our general notion of union:

$$x_1 \cup x_2 = \bigcup\{x_1, x_2\}.$$

Similarly, the *intersection* of x, written $\bigcap x$, is the set consisting of all those elements which lie in every member of x: that is,

$$\bigcap x = \{z : z \in y \text{ for all } y \in x\}.$$

Thus, $\bigcap\{x_1, x_2\}$ is what is usually written as $x_1 \cap x_2$. We say that the sets x_1 and x_2 are *disjoint* if $\bigcap\{x_1, x_2\} = \varnothing$.

There is a problem with the empty set here (not the only place in mathematics where this occurs!). The union $\bigcup \varnothing$ is just the empty set, since there are no members of the empty set and hence no members of its members. But $\bigcap \varnothing$ should consist of everything in the universe: for, given any z, the condition that it belongs to every $y \in \varnothing$ is vacuously true (there is no $y \in \varnothing$ to provide a restriction). Because of Russell's Paradox, we do not want to allow this to be a set. Accordingly, we decree that $\bigcap x$ is defined only if $x \neq \varnothing$. (This is rather like the ban on dividing by zero in a field.) When we come to develop axiomatic set theory (in Chapter 6), we will have an axiom guaranteeing that unions exist, but not for intersections. The fact that intersections of non-empty sets exist will follow from the other axioms.

By contrast, there is no problem in considering the intersection of a set x which has the empty set as one of its members: this intersection is just the empty set!

We define the *difference* $X \setminus Y$ of two sets X and Y to consist of all elements which are in X but not in Y:

$$X \setminus Y = \{x \in X : x \notin Y\}.$$

1.3 Cartesian products, relations and functions

We next want to define the *ordered pair* (x_1, x_2). This should be an object constructed out of x_1 and x_2, with the property that

$$(x_1, x_2) = (y_1, y_2) \text{ if and only if } x_1 = x_2 \text{ and } y_1 = y_2.$$

The set $\{x_1, x_2\}$ does not have this property, since its elements do not come in any particular order: we have $\{x_1, x_2\} = \{x_2, x_1\}$. Some cleverness is needed to find the right construction. (But, once it is found, the details are not needed; only the property described above is ever used.)

We define

$$(x_1, x_2) = \{\{x_1\}, \{x_1, x_2\}\}.$$

Theorem 1.3

$$(x_1, x_2) = (y_1, y_2) \text{ if and only if } x_1 = y_1 \text{ and } x_2 = y_2.$$

Proof

The implication from right to left is clear. So suppose that $(x_1, x_2) = (y_1, y_2)$, that is, that

$$\{\{x_1\}, \{x_1, x_2\}\} = \{\{y_1\}, \{y_1, y_2\}\}.$$

The set on the left has just two members (which might happen to be equal), namely $\{x_1\}$ and $\{x_1, x_2\}$. Similarly for the set on the right. Thus the Principle of Extension shows that *either*

$$\{x_1\} = \{y_1\}, \quad \{x_1, x_2\} = \{y_1, y_2\} \quad \textit{(Case A)},$$

or

$$\{x_1\} = \{y_1, y_2\}, \quad \{x_1, x_2\} = \{y_1\} \quad \textit{(Case B)}.$$

We consider these cases in turn.

In *Case A*, we have (again by Extension) $x_1 = y_1$, and either $x_1 = y_1$, $x_2 = y_2$ *(Case A1)*, or $x_2 = y_1$, $x_1 = y_2$ *(Case A2)*. In the first case, we have reached the conclusion we want. In the second, we have

$$y_2 = x_1 = y_1 = x_2,$$

so again we are done. Similarly, in *Case B*, we have $y_1 = x_1 = y_2$ and $x_1 = y_1 = x_2$, so again $x_1 = y_1$ and $x_2 = y_2$ follow. □

Now we can define ordered n-tuples for $n > 2$ inductively by the rule

$$(x_1, \ldots, x_{n-1}, x_n) = ((x_1, \ldots, x_{n-1}), x_n).$$

They have the property that

$$(x_1, \ldots, x_n) = (y_1, \ldots, y_n) \text{ if and only if } x_1 = y_1, \ldots, x_n = y_n.$$

The *cartesian product* $X_1 \times X_2$ of two sets X_1 and X_2 is defined to be the set of all ordered pairs (x_1, x_2), where $x_1 \in X_1$ and $x_2 \in X_2$. We abbreviate $X \times X$ to X^2.

The term originates in the work of René Descartes. He realized that, by taking two perpendicular axes and setting up coordinates, the points of the Euclidean plane can be labelled in a unique way by ordered pairs of real numbers. We can then take the further step of saying that a point *is* a pair of real numbers, so that the set of points of the Euclidean plane is the cartesian product $\mathbb{R} \times \mathbb{R}$. Then we can say, for example, that a line or curve *is* the set of solutions of some equation in two variables, and we are well on the way to turning geometry into a branch of algebra. (Of course, this is a much less grandiose project than turning all of mathematics into set theory, but it is a step on this road!)

In a similar way, $X_1 \times \cdots \times X_n$ is the set of all ordered n-tuples $(x_1, \ldots, x_n)$, where $x_1 \in X_1$, …, $x_n \in X_n$; and we abbreviate $X \times \cdots \times X$ (n factors) to X^n.

Note that, unlike the case of union and intersection, we have not defined a concept of the cartesian product $\prod x$ of an arbitrary set x, so that $\prod\{X_1, X_2\} = X_1 \times X_2$. This can be done (more or less), but we need some more notation to set it up.

The concept of a *function* has changed over the centuries. Until the nineteenth century, a function was given by a formula, though this formula could involve such analytic processes as sums of infinite series or transcendental functions as well as arithmetic operations. In Forsyth's *Theory of Functions of a Complex Variable*, written as late as 1893 and quoted by Littlewood [34], it is explained thus:

> ... if the value of X depends on that of x and on no other variable magnitude, it is customary to regard X as a function of x; and there is usually an implication that X is derived from x by some series of operations.

A crisis was provoked by various discoveries at the end of the nineteenth century, among them a real function which is everywhere continuous but nowhere differentiable, and a curve that passes through every point of the unit square. A more up-to-date version might involve specifying that the function values can be calculated by a computer, at least in principle. The modern definition however does not require this; it is as general as possible, subject only to the requirement that 'everything is a set': we ask only that each 'input' determines a unique 'output'.

A *function* $f : X \to Y$ from the set X to the set Y is a subset of $X \times Y$ with the property that *every element of X is the first component of a unique ordered pair in f;* in other words, for any $x \in X$, there is a unique $y \in Y$ such that $(x, y) \in f$. We write $y = f(x)$ as an alternative to $(x, y) \in f$. This only defines the compound expression $y = f(x)$; but, of course, there is a natural way of giving a meaning to $f(x)$, as the unique element y of Y for which this equation holds.

The requirement that f is a function can be expressed formally as follows:

- $(\forall x \in X)(\exists y \in Y)((x, y) \in f)$;
- $(x, y_1), (x, y_2) \in f \Rightarrow (y_1 = y_2)$.

The function $f : X \to Y$ is said to be *injective* or *one-to-one* if distinct elements of X have distinct images – that is,

- $(x_1, y), (x_2, y) \in f \Rightarrow (x_1 = x_2)$;

it is *surjective* or *onto* if every element of Y is the image of some element of x – that is,

- $(\forall y \in Y)(\exists x \in X)((x, y) \in f)$;

and it is *bijective* if both of these conditions hold. A bijective function has the effect of 'matching up' the elements of X with those of Y, so that there are the 'same number' of elements in X and Y. (This remark will become significant when we come to investigate exactly what numbers to use to measure the size of a set.)

A bijective function $f : X \to Y$ has an *inverse* $f^{-1} : Y \to X$, defined by

$$f^{-1} = \{(y, x) : (x, y) \in f\}.$$

Two functions $f : X \to Y$ and $g : Y \to Z$ can be *composed* to give a function $f \circ g : X \to Z$ defined by

$$f \circ g = \{(x, z) : (\exists y \in Y)((x, y) \in f \text{ and } (y, z) \in g)\}.$$

(Note that $(f \circ g)(x) = g(f(x))$: this unfortunate reversal arises because of our notation for functions, writing the function on the left of its argument.)

A particular bijection is the *identity* function i_X on a set X, which is defined to be

$$i_X = \{(x, x) : x \in X\}.$$

Note that, if $f : X \to Y$ is any function, then $i_X \circ f = f \circ i_Y = f$; and, if f is a bijection, then $f \circ f^{-1} = i_X$, and $f^{-1} \circ f = i_Y$.

Sometimes we use the notation Y^X for the set of all functions from X to Y.

A *relation* between X and Y is just a subset of $X \times Y$. If $X = Y$, we speak of a *binary relation* on X. The reason is much as before: we want to describe relations like 'less than' or 'divides', but without making any demands on the relation other than that it is a set. We know everything about a relation when we know which ordered pairs satisfy it. Note that a function is a special kind of relation.

If R is a relation, we often write $x \; R \; y$ instead of $(x, y) \in R$. This is motivated by the common relations such as 'less than' or 'divides', which are written $x < y$ and $x \mid y$ respectively. However, even though we regard $<$ as a set of ordered pairs, we draw the line at writing expressions like $(x, y) \in <$, however correct they may be logically!

Any set has the property that all its members are sets. However, we need a more general concept, that of a *family of sets*, which is more like an ordered tuple, and has the property that its components can be repeated more than once.

A *family of sets*, indexed by the *index set* I, is a function $F : I \to W$ for some W. The sets in the family are the sets $X_i = F(i)$ for $i \in I$, and we often write the family as $(X_i : i \in I)$. (Note the round brackets, suggesting an ordered tuple!)

The *union* and *intersection* of a family of sets are defined by

$$\begin{aligned} \bigcup_{i \in I} X_i &= \bigcup \{X_i : i \in I\}, \\ \bigcap_{i \in I} X_i &= \bigcap \{X_i : i \in I\}. \end{aligned}$$

(This means, for example, that $\bigcup_{i\in I} X_i$ consists of all sets which are members of X_i for some $i \in I$.) The union is always defined, and the intersection is defined provided that $I \neq \varnothing$.

We can also define the cartesian product of a family of sets. First, given a family $(X_i : i \in I)$, defined by a function $F : I \to W$ (so that $X_i = F(i)$), we define a *choice function* for the family to be a function $f : I \to \bigcup W$ having the property that $f(i) \in F(i)$ for all $i \in I$. (The function f 'chooses' a representative element from each set in the family.) The choice function is itself a family; we often write it as $(x_i : i \in I)$, with $x_i = f(i)$. Now the *cartesian product* $\prod_{i\in I} X_i$ is defined to be the set of all choice functions for the family $(X_i : i \in I)$.

For example, suppose that $I = \{1, 2\}$, and let $(X_i : i \in I)$ be a family of sets indexed by I. A choice function picks out an element $x_1 \in X_1$ and an element $x_2 \in X_2$, and so can be represented as an ordered pair (x_1, x_2), an element of $X_1 \times X_2$. So $\prod_{i\in I} X_i$ is essentially the same as $X_1 \times X_2$ in this case.

If one of the sets in the family is empty, then the cartesian product of the family is empty: no choice function can exist. Is the converse true? It turns out that the converse is the celebrated *Axiom of Choice*, about which we will have more to say later. We note here that it can neither be proved nor disproved by the reasonable assumptions about sets that we have made so far.

For future record, we display it here, not as an 'axiom' (yet), far less as a 'theorem', but as a mysterious principle we will invoke from time to time.

Axiom of Choice: Any family of non-empty sets has a choice function.

1.4 Equivalence and order

Some special kinds of relations will be very important to us. To define these, we first list some properties which a relation may (or may not) possess.

The relation R on X is said to be

- *reflexive* if $(x, x) \in R$ for all $x \in X$;
- *irreflexive* if $(x, x) \notin R$ for all $x \in X$;
- *symmetric* if $(x, y) \in R$ implies $(y, x) \in R$;
- *antisymmetric* if $(x, y) \in R$ and $(y, x) \in R$ imply $x = y$;
- *transitive* if $(x, y) \in R$ and $(y, z) \in R$ imply $(x, z) \in R$.

Note that 'irreflexive' is not the same as 'not reflexive', and 'antisymmetric' is not the same as 'not symmetric'. Moreover, we do not define 'antisymmetric' to mean that $(x, y) \in R$ implies $(y, x) \notin R$: for this would mean that an antisymmetric relation would automatically be irreflexive, and we don't want to prejudge this.

An *equivalence relation* on X is a relation which is reflexive, symmetric and transitive. As is well known, the job that an equivalence relation performs is to partition a set. We now make the appropriate definitions to formulate this property.

If R is any relation on X, we define the *R-class* of an element x to be the set $\{y \in X : (x, y) \in R\}$ consisting of everything related to x.

A *partition* of X is a set of non-empty sets which cover X without overlap. More precisely, it is a set P with the properties

- for all $p \in P$, $p \neq \varnothing$;
- for all $p, q \in P$, $p \cap q = \varnothing$;
- $\bigcup P = X$.

(Note how simply the third condition can be stated using our notation for union.)

Theorem 1.4 (Equivalence Relation Theorem)

(a) If R is an equivalence relation on X, then the set

$$P = \{R(x) : x \in X\}$$

of R-classes is a partition of X.

(b) If P is a partition of X, then

$$R = \{(x, y) : x, y \in p \text{ for some } p \in P\}$$

is an equivalence relation on X.

(c) The constructions in (a) and (b) are mutually inverse.

The last statement means that, if we start with an equivalence relation, construct a partition as in (a), and then construct the equivalence relation as in (b), then it is equal to the original equivalence relation; and similarly if we start with a partition. This theorem is a fundamental mathematical statement, and a proof will not be provided here since it is standard. See Smith [42], Proposition 1.4; Johnson [26], Theorem 4.1; Wallace [46], Theorem 1.8.

If R is an equivalence relation on X, we use the notation X/R for the set of equivalence classes (the partition of X corresponding to R in the Equivalence

Relation Theorem). The notation is chosen to resemble that for a factor group or factor ring; deliberately, as we now see.

Suppose that X and Y are sets, and f a function from X to Y. The *image* of f, written $\mathrm{Im}(f)$, is the set

$$\mathrm{Im}(f) = \{y \in Y : (\exists x \in X)(y = f(x))\}$$

of elements which can be represented as $f(x)$ for some $x \in X$. The *kernel* of f, written $\mathrm{KER}(f)$, is the relation on X defined by

$$\mathrm{KER}(f) = \{(x_1, x_2) \in X \times X : f(x_1) = f(x_2)\}.$$

Theorem 1.5 (First Isomorphism Theorem)

Let f be a function from X to Y. Then

(a) $\mathrm{Im}(f)$ is a subset of Y;

(b) $\mathrm{KER}(f)$ is an equivalence relation on X;

(c) f induces a bijection from $X/\mathrm{KER}(f)$ to $\mathrm{Im}(f)$.

The proof is an exercise.

There are two kinds of order relations, modelled by the relations ‘less than’ or ‘less than or equal’ on the integers or real numbers. They are defined as follows. A *(strict) (partial) order* on X is a relation on X which is irreflexive, antisymmetric and transitive. A *non-strict (partial) order* on X is a relation which is reflexive, antisymmetric and transitive. Note the use of brackets: it is intended that the term ‘strict’ is the default, so that if we don’t specify whether an order is strict or not it is assumed to be strict; and the term ‘partial’ is the default, so that (when we shortly meet the term ‘total order’) it is assumed that an order is partial unless it is stated that it is total.

So, on one of the standard number systems, the relation $<$ is a strict order, while $\leq$ is a non-strict order.

An example of a non-strict order is the relation $\subseteq$ or ‘is a subset of’, on any set x. (The end of this sentence is a little surprising: you may have expected the statement ‘on the power set $\mathcal{P}x$ of any set x’. But remember that everything is a set!)

Often we denote strict and non-strict orders by $<$ and $\leq$ respectively. There is a close relationship between the two types:

Theorem 1.6 (Order Theorem)

(a) If R is a strict order on X, then

$$S = R \cup \{(x,x) : x \in X\}$$

is a non-strict order on X.

(b) If S is a non-strict order on X, then

$$R = \{(x,y) \in S : x \neq y\}$$

is a strict order on X.

(c) The constructions in (a) and (b) are mutually inverse.

Again we omit the straightforward proof. We adopt the convention that if a strict and a non-strict order are related as in the Order Theorem, and the strict order is called $<$, then the non-strict order is called $\leq$, and *vice versa*. Also, we write $x > y$ as an equivalent of $y < x$, and $x \geq y$ as an equivalent of $y \leq x$. As a final piece of terminology, if $<$ is an order on X, we call the ordered pair $(X,<)$ an *ordered set.*

A *total order* is a partial order which satisfies the condition of *trichotomy*: for all $x, y \in X$, one of $(x,y) \in R$, $x = y$, and $(y,x) \in R$. Note that, if R is a strict order, then at most one of these three conditions can hold for any x and y, so R is total if exactly one always holds. Also, if we write $x < y$ for $(x,y) \in R$, the condition of trichotomy can be written: for all $x, y \in X$, one of $x < y$, $x = y$, $x > y$ holds.

Let $(X,<)$ be an ordered set. We call the element $x \in X$ a *least element* of X if, for all $y \in X$, we have $x \leq y$. We call x a *minimal element* of X if, for all $y \in X$, $y \leq x$ implies $y = x$. (So 'least' means 'smaller than everything else'; and 'minimal' means 'nothing else is smaller'.)

Theorem 1.7

Let $(X,<)$ be an ordered set.

(a) If a least element of X exists, then it is minimal, and moreover it is the unique minimal element.

(b) If X is totally ordered, then any minimal element is a least element.

Proof

(a) Suppose that x is least. We must show first that it is minimal. So suppose that $y \leq x$. Then also $x \leq y$, since x is least; so $x = y$ by antisymmetry.

Now let y be any minimal element. Then $x \leq y$, since x is least; and so $x = y$, since y is minimal. So x is the unique minimal element.

(b) Suppose that X is totally ordered, and that x is minimal. If x is not least, then there exists $y \in X$ such that $x \leq y$ does not hold. Hence $y < x$ by trichotomy. But this is impossible, since $y \leq x$ implies $y = x$, as x is minimal. □

Dually, an element x of the ordered set $(X, <)$ is *greatest* if $x \geq y$ for all $y \in X$; and x is *maximal* if $y \geq x$ implies $y = x$. The analogues of the above theorem hold for these concepts.

Finally in this section, we define the concept of *isomorphism*. The definition is given only for sets with a single relation, but can be reworked to apply much more generally.

Let R be a relation on a set X, and S a relation on a set Y. We say that the structures (X, R) and (Y, S) are *isomorphic* if there is a bijection $f : X \to Y$ with the property that for all $x_1, x_2 \in X$, if $(x_1, x_2) \in R$, then $(f(x_1), f(x_2)) \in S$, and conversely. So the sets X and Y can be matched up so that the relation is satisfied by corresponding pairs in the two sets. We write $(X, R) \cong (Y, S)$ to denote that the structures (X, R) and (Y, S) are isomorphic. If the relations are understood, we sometimes abuse the notation and simply say that X and Y are isomorphic (and write $X \cong Y$).

Isomorphism is an 'equivalence relation' on the class of relational structures: that is,

- (X, R) is isomorphic to itself (by means of the identity function);
- if $(X, R) \cong (Y, S)$ (by a bijection f), then $(Y, S) \cong (X, R)$ (by the inverse bijection);
- if $(X, R) \cong (Y, S)$ (by the bijection f) and $(Y, S) \cong (Z, T)$ (by the bijection g), then $(X, R) \cong (Z, T)$ (by the composition of f and g).

1.5 Bijections

One of our goals is to develop a system of numbers to measure the sizes of sets, finite or infinite. The natural numbers play this role for finite sets. Though we haven't defined the natural numbers yet, it will turn out that the number n is a 'standard' set with n elements, and an arbitrary set has n elements if and only if it can be put into one-to-one correspondence with the 'standard' set n. In fact, the notion that two sets 'have the same number of elements' is really

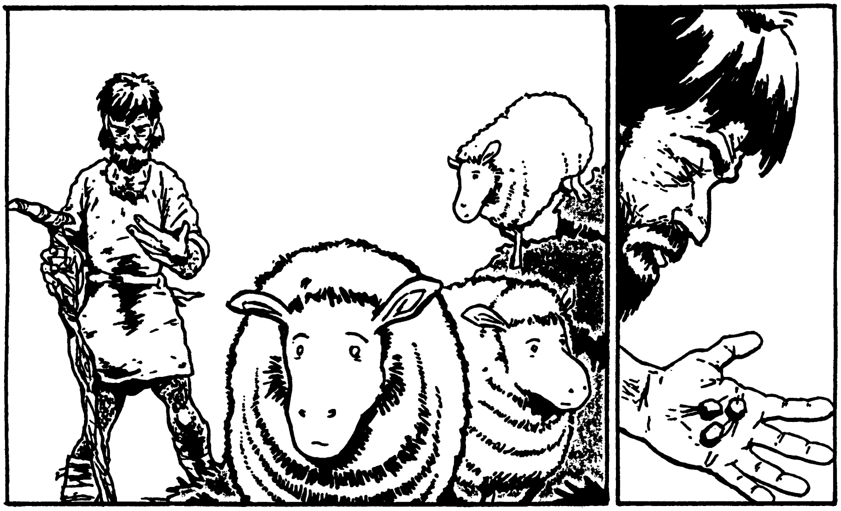


Fig. 1.1. Counting sheep

more basic than any statement about what this number is, since it only requires that there is a bijection between the sets. (See Figure 1.1.) Georges Ifrah [23] tells the story of an American archaeological team working in the palace of Nuzi, near Kirkuk in modern Iraq. They found a clay envelope inscribed with a list of 48 sheep and goats; when opened, the envelope contained 48 clay balls. Presumably the clay envelope was made by a literate accountant, while the clay balls were to enable the shepherd to check that the flock was complete. The shepherd would simply have checked that there was a bijection between the clay balls and the sheep, and would not have needed to be able to count to 48. The significance of the find was brought home to the archaeologists when their uneducated servant, sent to the market to buy chickens, was unable to say how many chickens he had purchased, but produced a collection of pebbles, one for each chicken.

So we take the bold approach: we say that two sets X and Y *have the same cardinality* if there is a bijection between them – we do not define yet what the cardinality of a set is. We write $|X| = |Y|$ if X and Y have the same cardinality, but, again, we do not yet assign any meaning to the symbol $|X|$ in isolation. (This will be done later!)

More generally, we say that the set X *has smaller cardinality than* the set Y (in symbols, $|X| \leq |Y|$) if there is an injection (a one-to-one mapping) from X to Y. If this holds, and if X and Y do not have the same cardinality, then we say that X *has strictly smaller cardinality than* Y, and write $|X| < |Y|$.

Surprisingly, many assertions which might seem quite obvious or natural cannot be proved without the Axiom of Choice. These include the statements

- any two sets X and Y are *comparable* (in the sense that either $|X| \leq |Y|$ or

$|Y| \leq |X|$); and

- if $X \neq \varnothing$, there is an injective function from X to Y if and only if there is a surjective function from Y to X.

This being the case, it is important to see just what we can prove. At least the following two statements are true.

Theorem 1.8

If there is an injective function from X to Y, and $X \neq \varnothing$, then there is a surjective function from Y to X.

Proof

Let $f : X \to Y$ be injective. Let a be an arbitrary element of X. Now define a function $g : Y \to X$ by the rule

$$g(y) = \begin{cases} x & \text{if } f(x) = y; \\ a & \text{if no such } x \text{ exists.} \end{cases}$$

Since f is injective, if x exists, then it is unique; so the function is well-defined. Now for any $x \in X$, we have $x = g(f(x))$; so g is surjective. □

Theorem 1.9 (Schröder–Bernstein Theorem)

If there is an injective function from X to Y and an injective function from Y to X, then there is a bijective function from X to Y.

In other words, if $|X| \leq |Y|$ and $|Y| \leq |X|$, then $|X| = |Y|$.

Proof

We are given injective functions $f : X \to Y$ and $g : Y \to X$, and have to construct from them a bijection between the two sets. We give two similar proofs of this important result. The first is more intuitive, but uses some elementary properties of natural numbers, whereas the second uses nothing but set theory. Without loss of generality, X and Y are disjoint. (Given any sets X and Y, we can find disjoint sets X' and Y' bijective with X and Y: for example, take $X' = X \times \{1\}$ and $Y' = Y \times \{2\}$.) This dodge is not needed for the second proof.

First Proof

We say that $y \in Y$ is the *parent* of $x \in X$ if $g(y) = x$; dually, $x \in X$ is the *parent* of $y \in Y$ if $f(x) = y$. Each element of X or Y has at most one parent. An *ancestral chain* for $z \in X \cup Y$ is a tuple $(z_0, z_1, \ldots, z_n)$ such that $z_0 = z$ and z_{i+1} is the parent of z_i for $i = 0, \ldots, n-1$. (Its elements belong alternately to X and Y.) The *length* of such a chain is n (the number of *steps*, not the number of *elements*).

Now there are two possibilities for any element z. Either there are arbitrarily long ancestral chains for z, in which case we shall say that z has *infinite depth*; or there is a unique longest ancestral chain for z, ending with an element with no parent, in which case we say that the length of this chain is the *depth* of z. (The second possibility includes the case when z itself has no parent, in which case its depth is 0.) We let X_{e} denote the set of elements of X whose depth is even; X_{o} the set of elements of X with odd depth; and X_∞ the set of elements with infinite depth. We define Y_{e}, Y_{o}, and Y_∞ similarly.

If $x \in X$ has finite depth, then $f(x)$ has depth one greater than the depth of X; and if $x \in X$ has infinite depth, then so does $f(x)$. So f maps $X_{\mathrm{e}} \to Y_{\mathrm{o}}$, $X_{\mathrm{o}} \to Y_{\mathrm{e}}$, and $X_\infty \to Y_\infty$. A similar assertion holds for the action of g on elements in Y. Furthermore, elements of Y_{o} or Y_∞ have parents; so f maps $X_{\mathrm{e}} \to Y_{\mathrm{o}}$ and $X_\infty \to Y_\infty$ bijectively. (This does not hold for $X_{\mathrm{o}} \to Y_{\mathrm{e}}$ since an element of Y_{e} may have no parent.)

Define a map $h : X \to Y$ by

$$h(x) = \begin{cases} f(x) & \text{if } x \in X_{\mathrm{e}} \cup X_\infty, \\ g^{-1}(x) & \text{if } x \in X_{\mathrm{o}}. \end{cases}$$

Then it is easily seen that h is a bijection. □

Second Proof

We first prove a couple of lemmas.

Lemma 1.1

Let X be a set, and $p : \mathcal{P}X \to \mathcal{P}X$ a function which is monotonic, in the sense that if $A \subseteq B \subseteq X$, then $p(A) \subseteq p(B)$. Then there is a set $Z \subseteq X$ such that $p(Z) = Z$.

Proof

We set $Z = \bigcup\{A \subseteq X : A \subseteq p(A)\}$. Take $z \in Z$. Then there is a set $A \subseteq X$ such that $z \in A$ and $A \subseteq p(A)$. So $z \in p(A)$. Moreover, $A \subseteq Z$, so $p(A) \subseteq p(Z)$ by hypothesis. Thus $z \in p(Z)$. We have shown that $Z \subseteq p(Z)$. Again by hypothesis, $p(Z) \subseteq p(p(Z))$. So $p(Z)$ is one of the sets in the family whose union is Z. But this means that $p(Z) \subseteq Z$. So we have $Z = p(Z)$, as claimed. □

For the next lemma, we require some notation. If f is a function on a set X, and $A \subseteq X$, we define $f[A] = \{f(a) : a \in A\}$.

Lemma 1.2

Let X and Y be sets, and $f : X \to Y$ and $g : Y \to X$ be injective functions. For $A \subseteq X$, set $p(A) = X \setminus g[Y \setminus f[A]]$. Then, if $A \subseteq B \subseteq X$, we have $p(A) \subseteq p(B)$.

Proof

Suppose that $A \subseteq B$ but $p(A) \not\subseteq p(B)$. Choose an element $x \in p(A) \setminus p(B)$. Then, by definition of p, we have $x \in g[Y \setminus f[B]]$; say $x = g(y)$, where $y \in Y \setminus f[B]$. But $A \subseteq B$, so $f[A] \subseteq f[B]$, so $Y \setminus f[B] \subseteq Y \setminus f[A]$. We conclude that $y \in Y \setminus f[A]$, so that $x = g(y) \in g[Y \setminus f[A]]$; that is, $x \notin p(A)$, contrary to assumption. □

Now we prove the Schröder–Bernstein Theorem. Let X, Y, f, g be as in the theorem. Define the function $p : \mathcal{P}X \to \mathcal{P}X$ as in Lemma 1.2. Then the two lemmas imply the existence of a subset Z of X with $p(Z) = Z$. Define a function $h : X \to Y$ as in the first proof:

$$h(x) = \begin{cases} f(x) & \text{if } x \in Z, \\ g^{-1}(x) & \text{if } x \notin Z. \end{cases}$$

Note that, if $x \notin Z$, then $x \in g[Y \setminus f[Z]]$, so that $f = g(y)$ for some unique $y \in Y \setminus f[Z]$; this element y is what we have called $g^{-1}(x)$. Now it is readily checked that $h : X \to Y$ is a bijection. □

There is one particular case, noted by Cantor, where we can show that no bijection exists between two sets.

Theorem 1.10 (Cantor's Theorem)

For any set X, there is an injection from X to $\mathcal{P}X$ but no bijection between these sets; that is,

$$|X| < |\mathcal{P}X|.$$

Proof

We can define an injection $f : X \to \mathcal{P}X$ very simply: set $f(x) = \{x\}$. (By the Principle of Extension, if $\{x\} = \{y\}$ then $x = y$.)

For the second statement, we suppose for a contradiction that there is a bijection $h : X \to \mathcal{P}X$. Let

$$Y = \{x \in X : x \notin h(x)\}.$$

Since h is a bijection, there is a (unique) element $y \in X$ such that $Y = h(y)$. Now we ask: *Is* $y \in h(y)$*?* (Note the similarity with Russell's Paradox.) If $y \in h(y)$, then by definition $y \notin Y$; and similarly if $y \notin h(y)$ then $y \in Y$. But these contradict the fact that $Y = h(y)$. □

1.6 Finite sets

A set has n elements if it has a bijection with the set $\{1, \ldots, n\}$. (This is the basic principle of counting, that we learn at an early age and use every day.) It will be convenient to regard the number n as a 'standard' n-element set to use for counting. However, if we take n to be the set $\{1, \ldots, n\}$, then n is a member of itself; a feeling of nervousness about Russell's Paradox leads us to proceed slightly differently. Anticipating the next chapter, we take the number n to be the set $\{0, \ldots, n-1\}$. Also, we take the set $\mathbb{N}$ of natural numbers to include 0: that is, $\mathbb{N} = \{0, 1, 2, \ldots\}$.

So set theorists count

zero, one, two, ...

instead of

one, two, three, ...

like the rest of us.

We begin with a simple but important property: no two natural numbers have the same cardinality!

Theorem 1.11

If there is a bijection between n and m, then $n = m$.

Proof

We will prove this theorem by induction. (When we construct the natural numbers in the next chapter, induction will be available right from the start.)

Specifically, we use induction on n. If $n = 0$, then the set $n = \{0, \ldots, n-1\}$ is the empty set, and the only set bijective with it is the empty set. So the induction starts.

Suppose that $k > 0$, and that the result is true with $n = k-1$. Now suppose that we are given a bijection f from $\{0, \ldots, k-1\}$ to $\{0, \ldots, m-1\}$. Suppose first that $f(k-1) = m-1$. Then the restriction of f to $\{0, \ldots, k-2\}$ is a bijection to $\{0, \ldots, m-2\}$. By the inductive hypothesis, $k-1 = m-1$, so $k = m$.

If $f(k-1) \neq m-1$, then we adjust f to produce a bijection which does satisfy this. Let $f(k-1) = a$ and $f(b) = m-1$, and define a function f' by the rule

$$f'(x) = \begin{cases} f(x) & \text{if } x \neq b, k-1; \\ a & \text{if } x = b; \\ m-1 & \text{if } x = k-1. \end{cases}$$

(See Figure 1.2.) Then f' is a bijection and $f'(k-1) = m-1$, and we conclude as before that $k = m$. Thus the result is proved for $n = k$, and the induction goes through. □

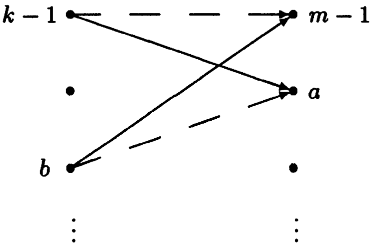


Fig. 1.2. Bijections

It follows that, for any set X, there is at most one natural number n such that X is bijective with n. We say that X is *finite* if such an n exists, and is *infinite* otherwise.

Sometimes the terms 'Peano finite' and 'Peano infinite' are used here. This is because an alternative definition was proposed by Dedekind: we say that a set X is *Dedekind infinite* if there is a bijection from X to a proper subset of itself, and is *Dedekind finite* if no such bijection exists.

Now a set which is Peano finite is Dedekind finite. For suppose that X is bijective with $\{0, \ldots, n-1\}$, and that X is bijective with a proper subset of itself. We may assume that $X = \{0, \dots, n-1\}$. We show by induction that this is impossible: the induction clearly starts when $n = 0$. So suppose that $n > 0$, and that no bijection from $\{0, \dots, n-1\}$ to a proper subset can exist, but there is a bijection f from $\{0, \dots, n\}$ to a proper subset. Just as in the proof of the above theorem, we can modify f to ensure that $f(n-1) = n-1$. Then the restriction of f to $\{0, \dots, n-2\}$ is a bijection from this set to a proper subset, contradicting the inductive hypothesis.

What about the converse? Suppose that X is Peano infinite. Then choose distinct elements $x_0, x_1, \dots$ of X: we never get stuck since if $X \setminus \{x_0, \dots, x_{n-1}\}$ were empty, then X would be bijective with $n = \{0, \dots, n-1\}$, and so would be Peano finite. Then define a map $f : X \to X$ by the rule

$$f(x) = \begin{cases} x_{n+1} & \text{if } x = x_n; \\ x & \text{if } x \neq x_i \text{ for all } i. \end{cases}$$

Then f is a bijection from X to $X \setminus \{x_0\}$, and so X is Dedekind infinite.

This proof involves making infinitely many choices, and so it requires the Axiom of Choice. It is not true that Peano's and Dedekind's definitions of finiteness are equivalent without the assumption that this principle is true.

We will briefly touch on some counting results which hold in finite sets. For the remainder of this section, 'finite' means 'Peano finite'. We use the notation $|X| = n$ to mean that X is bijective with the set $n = \{0, \dots, n-1\}$.

If n and m are natural numbers, we let $\binom{n}{m}$ denote the number of subsets of X of cardinality m, where X is a set of cardinality n. (Of course, it does not matter which set of cardinality n we choose.) As is standard,

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

where $n! = 1 \cdot 2 \cdots n$.

Theorem 1.12 (Principle of Inclusion and Exclusion)

Let $(A_i : i \in I)$ be a family of subsets of the finite set U, with $|I| = n$. For any non-empty $J \subseteq I$, let $A_J = \bigcap_{i \in J} A_i$, and let $A_\emptyset = U$. Then the number of elements of U which lie in none of the sets A_i is equal to

$$\sum_{J \subseteq I} (-1)^{|J|} |A_J| .$$

Proof

The summation in the theorem can be regarded as being made up of a contribution, for each element x of U, of the sum (over all J for which x lies in A_J) of $(-1)^{|J|}$. Now there are two cases:

- If x lies in none of the sets A_i, then the only set J such that $x \in A_J$ is $J = \varnothing$, and so the contribution of x to the sum is 1.
- Suppose that x does lie in some set A_i, and let $K = \{i \in I : x \in A_i\}$, with $k = |K|$. Then the sets J for which $x \in A_J$ are just the subsets of K; and the contribution of x to the sum is

  $$\sum_{J \subseteq K} (-1)^{|J|} = \sum_{j=0}^{k} \binom{k}{j} (-1)^j = (1-1)^k = 0,$$

  the second equality being an instance of the Binomial Theorem.

So the sum does indeed count the points lying in no set A_i. □

Theorem 1.13

Let X and Y be finite sets with $|X| = m$ and $|Y| = n$. Then

(a) the number of functions from X to Y is n^m;

(b) the number of one-to-one functions from X to Y is $n(n-1)\cdots(n-m+1)$;

(c) the number of surjective functions from X to Y is

$$\sum_{j=0}^{n} (-1)^j \binom{n}{j} (n-j)^m.$$

Proof

We can assume that $X = \{0, \ldots, m-1\}$. A function f from X to Y is specified completely by the m-tuple

$$(f(0), \ldots, f(m-1))$$

of values. Now each value can be chosen to be any of the n elements of Y, and these choices are independent; so the number of functions is n^m, proving (a). If we require f to be one-to-one, then $f(0)$ may still be any of the n elements of X; $f(1)$ may be any element of X except $f(0)$ (giving $n-1$ choices); and so on, so that there are $n-m+1$ choices for $f(m)$; multiplying these numbers gives (b). (This argument is not valid if $m > n$, since we cannot have negative numbers of choices; but, if $m > n$, then there are no possible choices for $f(m-1)$, and so no one-to-one functions exist, and the formula in (b) correctly gives zero.)

Part (c) is proved using the Principle of Inclusion and Exclusion. We let U be the set of all functions from X to Y. For each element $y \in Y$, we let A_y be the set of all functions which do not take the value y. Now a function is onto if and only if it lies in none of the sets A_y, for $y \in Y$.

If $J \subseteq Y$, then A_J consists of all functions which do not take any value in the subset J. These are precisely the functions from X to $Y \setminus J$, and there are $(n-j)^m$ of them, if $j = |J|$. Now there are $\binom{n}{j}$ choices for the subset J of cardinality J, and each of them contributes a term $(-1)^j(n-j)^m$ to the sum in the Inclusion–Exclusion Principle. Applying this principle proves the result. □

These results quantify, for finite sets, the facts that there is a one-to-one function from X to Y if and only if $|X| \leq |Y|$, and an onto function if and only if $|X| \geq |Y|$. However, it is not at all obvious that the formula in (c) yields zero when $m < n$. Try a few small values to check that it does.

1.7 Countable sets

We will briefly discuss countable sets in this section. The ideas here will be generalized later to arbitrary infinite sets.

Recall that a set is finite if it has the same cardinality as n, for some natural number n. We say that a set is *countable* if it has the same cardinality as the set $\mathbb{N}$.

Theorem 1.14

A set is at most countable if and only if it is finite or countable.

Proof

The statement involves a small adaptation of our earlier terminology: a set X is *at most countable* if there is an injective function from X to $\mathbb{N}$. Now if f is such a function, we modify it to form a function $g : X \to \mathbb{N}$ as follows: $g(x) = n$ if $f(x)$ is the nth element in the image of f. More formally, we define (by induction) $g(x) = 0$ if $f(x)$ is the least element of $f[X]$, and $g(x) = n$ if $f(x)$ is the smallest element of $f[X \setminus g^{-1}\{0, \ldots, n-1\}]$. Now either this inductive procedure terminates because there are no elements left (in which case $f[X] = \{0, \ldots, n-1\}$, and X is finite), or it continues for ever and defines a bijection between X and $\mathbb{N}$. □

We see that, if a set X is at most countable, then we can write it as $\{x_0, x_1, \ldots\}$, where either the list of elements stops at x_{n-1} for some n (if X has n elements), or it continues for ever (if X is countable).

The difficulties described at the end of Section 1.5 do not arise for sets which are at most countable. Any two such sets are comparable. (Two countable sets have the same cardinality; if X is finite and Y is countable then $|X| < |Y|$; and, if X has n elements and Y has m elements, then $|X| \leq |Y|$ if and only if $m \leq n$.) Moreover:

Theorem 1.15

There is a surjection from $\mathbb{N}$ to X if and only if X is at most countable.

Proof

We already know that the 'if' statement holds. So suppose that $g : \mathbb{N} \to X$ is a surjective function. For each $x \in X$, the set $g^{-1}(x) = \{n \in \mathbb{N} : g(n) = x\}$ is non-empty, so has a smallest element m_x. Now the function $f : X \to \mathbb{N}$ defined by $f(x) = m_x$ is injective since, if $x \neq y$, then $g(m_x) = x \neq y = g(m_y)$, so $m_x \neq m_y$. □

Various constructions applied to sets which are at most countable yield sets with the same properties.

Theorem 1.16

(a) The union of at most countably many at most countable sets is at most countable.

(b) The cartesian product of two at most countable sets is at most countable.

Proof

We observe first that a set is countable if and only if we can arrange its elements in an infinite sequence $(x_0, x_1, x_2, \ldots)$ so that each element occurs exactly once in the sequence. (This says no more or less than that the function $f : \mathbb{N} \to X$ defined by $f(n) = x_n$ is a bijection.)

We prove first that the cartesian product of two countable sets is countable. For this, it is enough to prove that $\mathbb{N} \times \mathbb{N}$ is countable, so we have to arrange the set of ordered pairs of natural numbers in a sequence. This we do as follows. First we break $\mathbb{N} \times \mathbb{N}$ into finite sets $S_0, S_1, S_2, \ldots$, where

$$S_k = \{(i, j) \in \mathbb{N} \times \mathbb{N} : i + j = k\}.$$

The set S_k contains just $k+1$ pairs, namely

$$S_k = \{(0,k),(1,k-1),\ldots,(k,0)\}.$$

Now we arrange the list by writing first the one element of S_0, then the two elements of S_1, then the three elements of S_2, and so on. (See Figure 1.3.)

$$\begin{tikzcd}
(0,0) & (0,1) \arrow[dl] & (0,2) \arrow[dl] & (0,3) \arrow[dl] & (0,4) \arrow[dl] & \cdots \\
(1,0) & (1,1) \arrow[dl] & (1,2) \arrow[dl] & (1,3) \arrow[dl] & (1,4) \arrow[dl] & \cdots \\
(2,0) & (2,1) \arrow[dl] & (2,2) \arrow[dl] & (2,3) \arrow[dl] & (2,4) \arrow[dl] & \cdots \\
(3,0) & (3,1) \arrow[dl] & (3,2) \arrow[dl] & (3,3) \arrow[dl] & (3,4) \arrow[dl] & \cdots \\
(4,0) & (4,1) & (4,2) & (4,3) & (4,4) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{tikzcd}$$

Fig. 1.3. Arranging $\mathbb{N} \times \mathbb{N}$ in a sequence

It is possible to work out an explicit formula for the bijection f from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$, though (as we have stressed) this is not necessary. The element (i,j) is the $(i+1)$st in S_{i+j}, and so is preceded by i elements of S_{i+j} together with all the elements of $S_0 \cup \ldots \cup S_{i+j-1}$ (in number $1+2+\ldots+(i+j) = (i+j)(i+j+1)/2$). So

$$f((i,j)) = (i+j)(i+j+1)/2 + i.$$

Now assertion (b) of the theorem follows: for, if X and Y are at most countable, then there are countable sets X' and Y' containing X and Y; and $|X \times Y| \leq |X' \times Y'|$.

Assertion (a) also follows easily. By a similar argument, it is enough to prove that the union of countably many countable sets is countable. So let $X_0, X_1, \ldots$ be countable sets, and let $X_i = \{x_{i0}, x_{i1}, \ldots\}$ for each i. Then the map f defined by $f((i,j)) = x_{ij}$ is a surjection from $\mathbb{N} \times \mathbb{N}$ to $\bigcup_{i \in \mathbb{N}} X_i$; so the union is countable. (This function need not be a bijection, since the sets X_i may not be disjoint.) □

As a illustration of this result, we show that the set of finite subsets of $\mathbb{N}$ is countable. (By contrast, we know from Cantor's Theorem that the set of all subsets of $\mathbb{N}$ is not countable.) Let $\mathcal{P}_{\text{fin}}\,\mathbb{N}$ denote the set of finite subsets of $\mathbb{N}$. Then

$$\mathcal{P}_{\text{fin}}\,\mathbb{N} = \textstyle\bigcup_{n \in \mathbb{N}} \mathcal{P}_n\,\mathbb{N},$$

where $\mathcal{P}_n \mathbb{N}$ is the set of n-element subsets of $\mathbb{N}$. Now $\mathcal{P}_0 \mathbb{N} = \{\varnothing\}$ has just one element. For $n > 0$, there is an injective map from $\mathcal{P}_n \mathbb{N}$ to $\mathbb{N}^n$: take any n-element set, and map it to the n-tuple obtained by writing its elements in increasing order. So $\mathcal{P}_n \mathbb{N}$ is countable, and $\mathcal{P}_{\text{fin}} \mathbb{N}$ is a countable union of at most countable sets, whence countable.

Here is another illustration. In the next section, we will outline the formal construction of the integers, rational, real and complex numbers; but we know enough about them to decide whether there are countably or uncountably many of each sort of number.

Theorem 1.17

The sets $\mathbb{Z}$ of integers and $\mathbb{Q}$ of rational numbers are both countable.

Proof

We have

$$\mathbb{Z} = \mathbb{N} \cup \{-n : n \in \mathbb{N}\},$$

so $\mathbb{Z}$ is the union of two countable sets, hence countable.

Any rational number can be written in its lowest terms as a/b, where a and b are coprime integers and $b > 0$; and this representation is unique. So the map $g : \mathbb{Q} \to \mathbb{Z} \times \mathbb{Z}$ given by $g(a/b) = (a, b)$ where a/b are as above, is an injection; and so $\mathbb{Q}$ is countable. □

By contrast, we have:

Theorem 1.18

The set $\mathbb{R}$ of real numbers is not countable.

Proof

Suppose for a contradiction that $\mathbb{R}$ is countable. Then the subset $(0, 1)$ of $\mathbb{R}$ (the open unit interval) is also countable. So we can write its elements as a sequence:

$$(0, 1) = \{r_0, r_1, r_2, \ldots\}.$$

Each real number in the unit interval can be written as an infinite decimal (possibly terminating, in which case we let it continue forever with zeros). Let

$$r_i = 0.x_{i0}x_{i1}x_{i2}\ldots$$

be the decimal expansion of r_i, where $x_{ij} \in \{0,1,2,\ldots,9\}$. Define

$$y_i = \begin{cases} 7 & \text{if } x_{ii} \neq 7, \\ 3 & \text{if } x_{ii} = 7. \end{cases}$$

Then $y_i \neq x_{ii}$ for all i. Let r be the real number given by

$$r = 0.y_0y_1y_2\ldots$$

Then $r \neq r_i$, since their decimal expansions differ in the ith place. So r is a real number in the unit interval not included in the sequence, contrary to assumption. □

1.8 The number systems

Leopold Kronecker said, 'God made the natural numbers: the rest is the work of man.' In the next chapter, we will see how we can construct the natural numbers themselves out of nothing, in mathematics built on set theory. In this section, we consider briefly how we build the rest of mathematics from the natural numbers. Later, we will look at Peano's alternative approach to the natural numbers, as an axiomatic system.

We assume, then, that we have the set $\mathbb{N}$ of natural numbers, on which the operations of addition and multiplication and the usual order relation are defined. Among the properties which hold are the following:

- addition and multiplication are commutative and associative;
- zero is the identity for addition, and 1 is the identity for multiplication;
- the *cancellation laws* hold for addition and for multiplication (except by zero): that is,

$$\begin{aligned} a+c=b+c &\quad \text{implies} \quad a=b, \\ ac=bc,\ c\neq 0 &\quad \text{implies} \quad a=b; \end{aligned}$$

- if $a < b$, then $a+c < b+c$ and (if $c \neq 0$) $ac < bc$.

Subtraction is not everywhere defined; that is, the equation $a + x = b$ has a solution (for given a, b) only if $a \leq b$. Our first task is to extend the natural numbers to the *integers*, so that subtraction is everywhere defined. Accordingly, we have to add solutions to all equations of this form. Each ordered pair (a, b) should determine a unique integer x satisfying $a + x = b$. Accordingly, we will *represent* this x by the ordered pair. Different ordered pairs should determine the same x; so in fact x should be represented by an equivalence class of ordered pairs. What is the corresponding equivalence relation? In other words, when will $b - a = d - c$ hold? This equation can be written as $a + d = b + c$, which makes sense in $\mathbb{N}$. Accordingly we proceed as follows.

Define a relation $\sim$ on $\mathbb{N} \times \mathbb{N}$ by the rule that $(a, b) \sim (c, d)$ if and only if $a + d = b + c$. It is a simple exercise to check (using the rules given above) that $\sim$ is an equivalence relation. Now we define the *integers* to be the set

$$\mathbb{Z} = (\mathbb{N} \times \mathbb{N})/\sim$$

of equivalence classes of this relation. We let $[a, b]$ denote the equivalence class containing (a, b). Now we define addition, multiplication and order on $\mathbb{Z}$ as follows:

- $[a, b] + [c, d] = [a + c, b + d]$;
- $[a, b] \cdot [c, d] = [ac + bd, ad + bc]$;
- $[a, b] \leq [c, d]$ if and only if $a + d \leq b + c$.

Where do these definitions come from? They are obtained from our intended interpretation, that $[a, b]$ will be the integer $a - b$. For example, $(a - b)(c - d) = (ac + bd) - (ad + bc)$.

We have first to show that these definitions are good ones; that is, that a different choice of representatives of the equivalence classes would not change the object defined. For example, suppose that $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$. Then a short calculation shows that $(a + c, b + d) \sim (a' + c', b' + d')$ and $(ac + bd, ad + bc) \sim (a'c' + b'd', a'd' + b'c')$.

Then it is not difficult to show that the usual arithmetic properties hold in $\mathbb{Z}$: for example, it is a commutative ring with identity and has no divisors of zero. (See Wallace [46], Chapters 3 and 5, for the concepts of ring theory.)

Moreover, the map that takes a to $[a, 0]$ is an injective function from $\mathbb{N}$ to $\mathbb{Z}$, and preserves addition, multiplication, and order. So, if we are working with $\mathbb{Z}$, we can now throw away our usual definition of $\mathbb{N}$ and regard a natural number as a special kind of integer (one which happens to satisfy $a \geq 0$).

The following constructions will be treated a bit more briefly; they are mostly similar to this one.

To obtain the rational numbers from the integers, we must add solutions to equations $ax = b$ with $a \neq 0$; and the equations $ax = b$ and $cx = d$ should have the same solution if $ad = bc$. So we define a relation $\sim$ on the set

$$\{(a, b) \in \mathbb{Z}^2 : a \neq 0\}$$

by the rule that $(a, b) \sim (c, d)$ if and only if $ad = bc$, and define the rational numbers to be the equivalence classes of this relation. Then we add, multiply and compare equivalence classes by the rules

- $[a,b]+[c,d]=[ad+bc,bd]$;
- $[a,b]\cdot[c,d]=[ac,bd]$;
- $[a,b]\leq[c,d]$ if and only if $abd^2\leq b^2cd$.

(These are the rules for fractions!) Prove that the operations are well-defined and that the set $\mathbb{Q}$ is an ordered field. Also, the map from $\mathbb{Z}$ to $\mathbb{Q}$ taking a to $[a,1]$ is an injective function which preserves addition, multiplication and order, so we can regard $\mathbb{Z}$ as a subset of $\mathbb{Q}$.

This procedure is known to algebraists as constructing the *field of fractions* of an integral domain.

The ordered set of rational numbers has many 'gaps'; useful numbers like $\sqrt{2}$, π and e are missing. (We can approximate them as closely as we choose by rational numbers, but cannot express them exactly.) The construction of the real numbers is designed to 'fill these gaps'.

The construction is a little more complicated than earlier ones. As we have seen, it increases the cardinality, so it cannot be done just by taking equivalence classes of pairs. Two procedures are commonly used, involving *Cauchy sequences* and *Dedekind cuts* respectively. We outline the second method. The idea is that any real number r cuts the rationals into two sets, namely $\{x\in\mathbb{Q}:x\leq r\}$ and $\{x\in\mathbb{Q}:x>r\}$. Different real numbers give different cuts of $\mathbb{Q}$, since between any two real numbers there is a rational. So we can identify the real number with the corresponding cut.

A *Dedekind cut* is a partition of $\mathbb{Q}$ into two subsets L and R with the properties

- every element of L is smaller than every element of R;
- R has no least element.

Then $\mathbb{R}$ is the set of all Dedekind cuts. We write a cut as an ordered pair (L,R).

Defining the arithmetic is not quite straightforward. As a first attempt, we would put $(L,R)+(L',R')=(L+L',R+R')$, where

$$L+L'=\{x+y:x\in L,y\in L'\}.$$

However, this may not be a Dedekind cut, since we may be missing one rational. (If the corresponding reals are $\sqrt{2}$ and $-\sqrt{2}$, for example, then 0 would lie in neither set.) In this case, we add the missing rational to $L+L'$. Multiplication is even more complicated, since positive and negative numbers have to be handled separately. Order, fortunately, is easy: $(L,R)\leq(L',R')$ if and only if $L\subseteq L'$.

Now with some work it can be shown that $\mathbb{R}$ is an ordered field and satisfies the *Principle of the Supremum*:

> Every non-empty set of real numbers which has an upper bound has a least upper bound (or supremum).

From these facts, the standard theorems of real analysis can be derived.

The construction of $\mathbb{C}$ from $\mathbb{R}$ can be done most easily by taking a complex number to be a pair of real numbers (namely, its real and imaginary parts). That is, $\mathbb{C} = \mathbb{R} \times \mathbb{R}$. We define addition and multiplication by

- $(a,b) + (c,d) = (a+c, b+d)$;
- $(a,b) \cdot (c,d) = (ac - bd, ad + bc)$.

Now it can be shown that $\mathbb{C}$ is a field, and is *algebraically closed* (that is, any polynomial equation over $\mathbb{C}$ has a solution).

1.9 Shoes and socks

The people up in Eppalock
Perambulate in one odd sock.
The other socks they hide away
To use as gifts on Christmas Day.

Michael Dugan and Walter Stackpool, *Nonsense Places: An Absurd Australian Alphabet*

Bertrand Russell posed the following problem:

(a) Suppose that a drawer contains infinitely many pairs of shoes. Construct a set containing one shoe from each pair.

(b) Same question for pairs of socks.

Part (a) of Russell's problem has an easy solution: you could simply take the right shoe from each pair. Part (b), however, is more problematic, since there is no natural rule for choosing one sock from each pair as there is for shoes.

More formally, the question asks whether there is a set S containing one element from each of an infinite set of two-element sets. This is rather different from the problem about 'the set of all sets which are not members of themselves' that we encountered earlier. It is not that S is 'too large' to be a set, since it is a subset of something we know to be a set (the union of all the two-element sets). Rather, it is that there is no obvious rule to guide the choice. There is nothing self-contradictory in the assumption that the set exists, nor (somewhat surprisingly) in the assumption that it doesn't exist.

Note that a set containing one sock from each pair is the image of a choice function for the family of pairs (an element of the cartesian product of the family). So Russell's problem, generalized, asks whether the cartesian product of any family of non-empty sets is non-empty.

We have already met the Axiom of Choice, which will guarantee that such sets do exist. However, this principle is not an *axiom* as Euclid used the term, a self-evident truth accepted by everybody. We will see that it has some paradoxical consequences! Because of its connection with cartesian products, Russell refers to it as the *Multiplicative Axiom*. We will examine it further in Chapter 6.

We could try to solve (b) in special cases. If the set of socks is countable, then we have an enumeration of it, as $\{s_0, s_1, \ldots\}$, and we could choose the sock in each pair with the smaller number. If the set of *pairs of socks* is countable, however, then we are in no better shape than we were without any information. This should make you look again at the proof of Theorem 1.16(a), that the union of countably many countable sets is countable. The proof actually required us to have an enumeration of each of the countable sets! So we were using the Axiom of Choice in this proof without realizing it, since having an enumeration of each pair of socks is exactly the same information as having a way of choosing one sock from each pair.

EXERCISES

1.1 Show that the empty set is a subset of every set.

1.2 Which of the following equations are true? If the equation is not true, is one side a subset of the other?

(a) $\bigcup \mathcal{P} X = X$.

(b) $\mathcal{P} \bigcup X = X$.

(c) $\bigcup \mathcal{P} X = \mathcal{P} \bigcup X$.

(d) $\mathcal{P}(X \times Y) = \mathcal{P} X \times \mathcal{P} Y$.

(e) $\mathcal{P}(X \cup Y) = \mathcal{P} X \cup \mathcal{P} Y$.

1.3 Prove that each of the following is *not* a suitable definition of the ordered pair (x, y):

(a) $(x, y) = \{x, \{y\}\}$.

(b) $(x, y) = \{\{x\}, \{y\}\}$.

1.4 Which of the following would be a suitable definition of the ordered triple (x, y, z)?

(a) $(x, y, z) = \{(x, y), (y, z)\}$.

(b) $(x, y, z) = ((x, y), (y, z))$.

(c) $(x, y, z) = \{\{x\}, \{x, y\}, \{x, y, z\}\}$.

1.5 This exercise describes a very fruitful source of equivalence relations in mathematics. (See Wallace [46], Chapter 6, for more details.) Let G be a group, and X a set. An *action* of G on X is a function $\mu : X \times G \to X$ satisfying the rules

- $\mu(x, gh) = \mu(\mu(x, g), h)$ for all $g, h \in G$, $x \in X$;
- $\mu(x, 1) = x$ for all $x \in X$, where 1 is the identity element of G.

(a) Prove that

- $\mu(\mu(x, g), g^{-1}) = \mu(\mu(x, g^{-1}), x) = x$ for all $x \in X$, $g \in G$.

(b) Define a relation $\sim$ on X by the rule that $x \sim y$ if and only if $\mu(x, g) = y$ for some $g \in G$. Show that $\sim$ is an equivalence relation. [*Note*: The equivalence classes of this relation are called the *orbits* of G in X (for the given action).]

(c) Show that each of the following equivalence relations on the set of all $m \times n$ real matrices arises from a group action:

- row-equivalence (A and B are row-equivalent if some sequence of elementary row operations transforms A to B);
- equivalence (A and B are equivalent if some sequence of elementary row and column operations transforms A to B);
- conjugacy, for $m = n$ (A and B are conjugate if $B = P^{-1}AP$ for some invertible matrix P);
- congruence, for $m = n$ (A and B are congruent if $B = P^{\top}AP$ for some invertible matrix P, where $P^{\top}$ is the transpose of P).

(d) If H is a subgroup of G, and H acts on G by *right multiplication* (that is, $\mu(x, h) = xh$), then the orbits of H are its *left cosets* in G.

(e) If G acts on itself by *conjugation* (that is, $\mu(x, g) = g^{-1}xg$), then the orbits of G are the *conjugacy classes*.

1.6 Suppose that μ is an action of the group G on the set X, as defined in the preceding exercise. Show that, for any $g \in G$, the function $x \mapsto \mu(x, g)$ is a bijection from X to X.

1.7 Show that the composition of injective functions is injective, and the composition of surjective functions is surjective.

1.8 Let $X \neq \varnothing$, let $f : X \to Y$ be a function, and let i_X and i_Y be the identity functions on X and Y respectively. Prove that

(a) f is injective if and only if there exists a function $g : Y \to X$ such that $f \circ g = i_X$;

(b) f is surjective if and only if there exists a function $h : Y \to X$ such that $h \circ f = i_Y$.

Where (if anywhere) have you used the Axiom of Choice in this proof?

1.9 Let $X \neq \varnothing$ and let $f : X \to Y$ be a function.

(a) Prove that f is injective if and only if $h_1 \circ f = h_2 \circ f$ implies $h_1 = h_2$, for any two functions $h_1, h_2 : Y \to X$.

(b) Prove that f is surjective if and only if $f \circ g_1 = f \circ g_2$ implies $g_1 = g_2$, for any two functions $h_1, h_2 : Y \to X$.

Where (if anywhere) have you used the Axiom of Choice in this proof?

1.10 Let R be a relation between X and Y. Define the *converse* of R to be the relation between Y and X defined by reversing all the pairs in R:

$$R^* = \{(y, x) : (y, x) \in R\}.$$

Show that the converse of a function f is a function if and only if f is bijective (in which case f^* is the inverse of f).

1.11 Let X and Y be finite sets, with m and n elements respectively. How many elements are there in each of the following sets?

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

(b) $X \times Y$.

(c) The set of relations from X to Y.

(d) The set $\prod_{y \in Y} X_y$, where $X_y = X$ for all $y \in Y$.

1.12 Show that

(a) any finite partially ordered set has a minimal element;

(b) any two (strict) total orders on a finite set are isomorphic;

(c) any (strict) partial order on a finite set X is contained in a (strict) total order on X.

1.13 Let R be a reflexive and transitive relation on a set X.

(a) Define a relation S on X by

$$S = \{(x, y) : (x, y), (y, x) \in R\}.$$

Show that S is an equivalence relation on X.

(b) Define a relation T on the set X/S of S-classes in X by

$$T = \{(S(x), S(y)) : (x, y) \in R\}.$$

Show that T is a non-strict order on X/S.

1.14 (a) Show that the cartesian product of finitely many copies of $\mathbb{N}$ is countable.

(b) Let X be a countable set. Show that the set X^* of all finite sequences of elements of X is countable.

(c) Prove that the set of *algebraic numbers* (those which satisfy some polynomial equation with integer coefficients) is countable. Prove that the set of *transcendental numbers* (those real numbers which are not algebraic) is uncountable.

1.15 This exercise completes our investigation of the cardinalities of the number systems.

(a) Show that there is a bijection between $\mathbb{R}$ and the open interval $(0, 1)$. [*Hint*: There is an analytic bijection.]

(b) Show that there is a bijection between the interval $(0, 1)$ and the interior of the unit square. [zemphHint: Interleave decimal expansions.]

(c) Deduce that $\mathbb{C}$ has the same cardinality as $\mathbb{R}$.

1.16 Let $(X, <)$ be a countable totally ordered set. Suppose that

(a) X is *dense*, that is, if $x < y$, then there exists z with $x < z < y$.

(b) X has no least or greatest element.

Prove that X is order-isomorphic to $\mathbb{Q}$.

[*Hint*: Enumerate $X = (x_0, x_1, \ldots)$ and $\mathbb{Q} = (q_0, q_1, \ldots)$. Now define, inductively, a map $f : X \to \mathbb{Q}$ as follows:

(a) $f(x_0) = q_0$.

(b) Suppose that $f(x_0), \ldots, f(x_{n-1})$ have been defined. Then the n points $x_0, \ldots, x_{n-1}$ divide X into $n+1$ intervals (including two semi-infinite intervals); x_n lies in one of these intervals, say (x_i, x_j). Now the corresponding interval $(f(x_i), f(x_j))$ in $\mathbb{Q}$ is non-empty. Choose the rational number q_h with smallest index in this interval, and define $f(x_n) = q_h$.

Prove that f is an order-preserving bijection.]

1.17 Use the same method to prove that any countable totally ordered set is isomorphic to a subset of $\mathbb{Q}$.

1.18 For $n > 0$, define a function $f : \mathcal{P}_n(\mathbb{N}) \to \mathbb{N}$ by the rule

$$f(\{x_0, x_1, \ldots, x_{n-1}\}) = \binom{x_0}{1} + \binom{x_1}{2} + \cdots + \binom{x_{n-1}}{n},$$

where $x_0 < x_1 < \cdots < x_{n-1}$. Prove that f is a bijection.

1.19 Prove that the following two statements are equivalent. (You may reason informally: when we discuss axioms for set theory in Chapter 6, you can check whether your solution is valid in axiomatic set theory. Note that statement (a) is what we have defined as the Axiom of Choice; because of the equivalence, we could take (b) instead if we preferred.)

(a) The cartesian product of any family of non-empty sets is non-empty.

(b) Let P be a partition of X. Then there is a subset Y of X which contains exactly one element from each member of P.

1.20 Use the Axiom of Choice to show that, if there is a surjection from Y to X, then there is an injection from X to Y.