Peter J. Cameron

Sets, Logic and Categories

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 $X_e \to Y_o$, $X_o \to Y_e$, 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 $X_e \to Y_o$ and $X_\infty \to Y_\infty$ bijectively. (This does not hold for $X_o \to Y_e$ since an element of Y_e may have no parent.)

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

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

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

2 Ordinal numbers

We have learned to pass with such facility from cardinal to ordinal number that the two aspects appear to us as one. To determine the plurality of a collection, that is, its cardinal number, we do not bother anymore to find a model collection with which we can match it – we *count* it ... The operations of arithmetic are based on the tacit assumption that *we can always pass from any number to its successor*, and this is the essence of the ordinal concept.

And so matching by itself is incapable of creating an art of reckoning. Without our ability to arrange things in ordered succession little progress could have been made. Correspondence and succession ... are woven into the very fabric of our number system.

Tobias Dantzig, *Number: The Language of Science* [12]

We learn numbers and counting as a process of succession. 'Eleven' has little real meaning to us except as 'the number after ten'. In this chapter, we use this process of succession to define the natural numbers – to do God's work, in Kronecker's phrase – starting from nothing (more precisely, the empty set) and progressing from one number to the next. As succession is the defining characteristic of natural numbers, so induction is the key proof technique. We can use it to define the arithmetic operations and to prove their basic properties.

But we do not have to stop there. By adding the principle of gathering up all the numbers so far constructed, we extend the ordinal number system into the infinite. We have transfinite induction to replace ordinary induction in our proofs. And, just as any finite set can be counted by a natural number, so any well-ordered set can be counted by a unique ordinal number. Moreover, when we come to define cardinal numbers in Chapter 6, we will see Dantzig's claim borne out: the only sets for which a satisfactory theory of cardinal number has been developed are those which can be well-ordered.

The ordinal numbers have another important role to play in the foundations. The strategy for developing a consistent set theory avoiding Russell's Paradox is to generate the sets out of nothing (that is, out of the empty set) in stages. In this way, the complete totality of sets is never formed, and Russell's 'set of all sets which are not members of themselves' is not defined. The stages in this process cannot just proceed through the natural numbers but must continue into the transfinite, at each stage admitting all subsets of the sets constructed at previous stages. The ordinals give a precise description of these stages. This is the most technical part of set theory which has to be developed before we begin the axiomatic approach. Having secured our theory from contradiction in this way, we follow standard mathematical practice by writing down axioms which capture the theory we have developed.

2.1 Well-order and induction

A *well-order* on a set X is a total order $<$ on X having the property that every non-empty subset of X has a least element (with respect to the restriction of $<$). This grammatically monstrous term is a back-formation from the term *well-ordered set*, which we apply to the ordered set $(X, <)$: strictly speaking we should talk about a 'good order', but the term 'well-order' has become firmly established and we continue to use it.

For example, the natural numbers (with the usual ordering) form a well-ordered set: every non-empty subset of natural numbers has a least element. Indeed, $(\mathbb{N}, <)$ is the simplest infinite well-ordered set. Any finite totally ordered set is well-ordered.

You should recognize that the well-ordering of the natural numbers is closely related to the idea of 'proof by induction'. One version of proof by induction which is commonly used is the 'minimal counterexample' technique, where we suppose that the proposition we are proving for all natural numbers is false, and argue on the smallest number which is a counterexample, using the fact that the proposition is true for all smaller numbers. Clearly this technique will work in any well-ordered set.

Theorem 2.1

Let $(X, <)$ be a well-ordered set. Suppose that Y is a subset of X with the property that, for all $x \in X$, if it holds that $y \in Y$ for all $y < x$, then it holds that $x \in Y$. Then $Y = X$.

Proof

Suppose that $Y \neq X$, so that $X \setminus Y \neq \varnothing$. Since X is well-ordered, $X \setminus Y$ has a least element, say x. By definition, any $y < x$ does not lie in $X \setminus Y$, and so lies in Y. The hypothesis of the theorem now shows that $x \in Y$, contrary to our choice of x. So it cannot be that $Y \neq X$. □

With a slight change, this becomes the *Principle of Induction*:

Theorem 2.2 (Principle of Induction)

Let $(X, <)$ be a well-ordered set. Let P be a property which may hold for elements of X. Suppose that, for all $x \in X$, if every element $y < x$ has property P, then x has property P. Then we conclude that every element of X has property P.

This follows on choosing Y to be the set of elements of X which have property P. Note that we don't have to do the base case of this induction: our hypotheses guarantee that the induction starts. For, if x is the smallest element of X, then there are no elements $y < x$, so vacuously all such elements have property P, whence x has property P by hypothesis.

2.2 The ordinals

We now develop the theory of *ordinals*, sometimes called *ordinal numbers*. These 'measure' well-ordered sets in the same way that natural numbers 'measure' finite sets. That is, given any finite set S, there is a unique natural number n such that a bijection exists between S and $\{1, 2, \ldots, n\}$: the number n is the *cardinality* of the set S. Inspired by this, we are going to prove the following theorem:

Theorem 2.3

Any well-ordered set is isomorphic to a unique ordinal.

The proof of this theorem is going to take the rest of this section. First, of course, we must define ordinals!

Given a totally ordered set $(X,<)$, and an element $a \in X$, we define the *section* X_a to consist of all elements of X which are less than a:

$$X_a = \{x \in X : x < a\}.$$

An *ordinal* is a well-ordered set $(X,<)$ with the property that $X_a = a$ for all $a \in X$. In other words, each element of X is the set of all its predecessors.

It is not immediately clear that ordinals exist. But we can see how they start. Let X be an ordinal. It has a least element a. Since a is the least element, we have $X_a = \varnothing$. But X is an ordinal, so $X_a = a$. Thus $a = \varnothing$. So the smallest element of any non-empty ordinal is $\varnothing$. Moreover, $\varnothing$ is (vacuously) an ordinal. Continuing, if $X \neq \{a\}$, then the subset $X \setminus \{a\}$ has a least element b; and we have $b = X_b = \{a\} = \{\varnothing\}$, so $\{\varnothing\}$ is the second ordinal. Continuing, we find that the next ordinal is $\{\varnothing, \{\varnothing\}\}$, and so on.

We will identify these 'starting ordinals' with the natural numbers – indeed, this will be our *definition* of natural numbers. We take

$$\begin{aligned} 0 &= \varnothing \\ 1 &= \{\varnothing\} = \{0\} \\ 2 &= \{\varnothing, \{\varnothing\}\} = \{0,1\} \\ 3 &= \ldots = \{0,1,2\} \end{aligned}$$

and so on. In general, we have

$$n = \{0,1,2,\ldots,n-1\},$$

as we anticipated in the last chapter. So *every natural number is an ordinal.*

But the ordinals continue after the natural numbers leave off. If ω denotes the smallest ordinal which is not a natural number, then ω is the set of all natural numbers. Then the next ordinal after ω is $\omega \cup \{\omega\}$, and so on ...

Before we begin the proof of the theorem, it is convenient to have a result which shows that the above methods of constructing ordinals are typical.

Theorem 2.4

(a) If x is an ordinal, then so is $x \cup \{x\}$ (with $y < x$ for all $y \in x$).

(b) The union of a set of ordinals is an ordinal.

Proof

(a) The set $a = x \cup \{x\}$ (with order as specified in the statement of the theorem) has as sections all the sections of x and one additional one, namely a_x. But since all the elements of x are smaller than x, we have $a_x = x$. Moreover, for $y \in x$, we have $a_y = x_y = y$, since x is an ordinal.

(b) We defer the proof of this assertion until we know a little more about ordinals (after Lemma 2.7 below). □

The proof of the theorem requires a series of technical lemmas.

Lemma 2.1

If $(X,<)$ is well-ordered, $Y \subseteq X$, and $f : X \to Y$ is an isomorphism, then $f(x) \geq x$ for all $x \in X$.

Proof

Induction. Let $E = \{x \in X : f(x) < x\}$. If $E \neq \varnothing$, then E has a least element, x_0 say. Then $f(x_0) < x_0$. Since f is an isomorphism, $f(f(x_0)) < f(x_0)$. But this shows that $f(x_0) \in E$, whereas $f(x_0)$ is smaller than the smallest element $x_0 \in E$, a contradiction. So $E = \varnothing$. □

Lemma 2.2

There is at most one isomorphism between any two well-ordered sets.

Proof

Let $f, g : X \to Y$ be isomorphisms. Then $f \circ g^{-1}$ is an isomorphism from X to X, so $x \leq g^{-1}(f(x))$ for all $x \in X$ by Lemma 2.1, from which we see that $g(x) \leq f(x)$ since g is an isomorphism. But a similar argument shows that $f(x) \leq g(x)$ for all x, whence $f(x) = g(x)$ by antisymmetry. □

Lemma 2.3

There is no isomorphism from a well-ordered set to a section of itself.

Proof

If $f : X \to X_a$ is an isomorphism, then $f(a) \in X_a$, so $f(a) < a$, contradicting Lemma 2.1. □

Lemma 2.4

Let $(X, <)$ be a well-ordered set, and let $A = \{X_a : a \in X\}$ be the set of sections of X. Then $(A, \subset) \cong (X, <)$.

Proof

The isomorphism from X to A is given by $f(a) = X_a$. It is one-to-one since, if $a < b$, then $a \in X_b$ but $a \notin X_a$, so $X_a \neq X_b$; and clearly it is onto. Suppose that $a < b$. Then for any $x \in X_a$ we have $x < a$, so $x < b$, so $x \in X_b$; thus $X_a \subseteq X_b$. We already saw that these sets are not equal, so $X_a \subset X_b$. This shows that f is an isomorphism. □

Lemma 2.5

Every section of an ordinal is an ordinal.

Proof

Let X be an ordinal and X_a a section of X. What is $(X_a)_b$ for $b \in X_a$? It consists of all the elements $x \in X_a$ which are less than b. But any x which is less than b is automatically less than a. So $(X_a)_b = X_b = b$, the last equality holding since X is an ordinal. We conclude that X_a is an ordinal. □

Lemma 2.6

Y

If X and Y are ordinals and $Y \subset X$, then X is a section of X.

Proof

Take a to be the least element of $X \setminus Y$. Then $X_a \subseteq Y$. Choose any $y \in Y$. If $a < y$, then $X_y = y = Y_y$ contains a, so $a \in Y$, contrary to assumption. Also $y = a$ is impossible since $y \in Y$ and $a \notin Y$. So $y < a$, and $y \in X_a$. So we have $Y \subseteq X_a$. We conclude that $Y = X_a$. □

Lemma 2.7

Let X and Y be distinct ordinals. Then one is a section of the other.

Proof

First, $X \cap Y$ is an ordinal: for, if $a \in X \cap Y$, then $X_a = a = Y_a$, so all elements of a belong to both X and Y, and $a = (X \cap Y)_a$. Hence, by Lemma 2.6, $X \cap Y$ is a section of both X and Y.

Now suppose that $X \not\subset Y$ and $Y \not\subset X$. Then $X \cap Y = X_a$ for some $a \in X$, and $X \cap Y = Y_b$ for some $b \in Y$. But then

$$a = X_a = X \cap Y = Y_b = b \in X \cap Y,$$

a contradiction. □

Proof of Theorem 2.4

First, note that any member of an ordinal is an ordinal. For, if x is an ordinal, $y \in x$, and $z \in y$, then $y = x_y$, so $y_z = (x_y)_z = x_z = z$.

Now let X be a set of ordinals. By the above remark, $A = \bigcup X$ is also a set of ordinals, and so there is an irreflexive and antisymmetric relation $<$ defined on A by the rule that $x < y$ if x is a section of y. Now $<$ is an order: for, if $x < y < z$, then $y = z_y$ and so $x = y_x = (z_y)_x = z_x$, so $x < z$. Now Lemma 2.7 shows that $<$ is a total order. Moreover, it is a well-order: for given a non-empty subset B of A, choose any $b \in B$; if it is not least, then all smaller elements are sections of b, and there is a least element among them since b is well-ordered. (This argument shows that any set of ordinals is well-ordered.) Finally, choose any $a \in A$; suppose that $a \in x \in X$. Then $a = x_a$, so all elements of a are in x, and hence in A, and we have $a = A_a$ as required. So A is an ordinal. □

Lemma 2.8

If X and Y are isomorphic ordinals then $X = Y$.

Proof

Let $f : X \to Y$ be an isomorphism, and $E = \{x \in X : f(x) \neq x\}$. If $E = \varnothing$, then $X = Y$; so suppose not. If a is the least element of E, then $f(x) = x$ for all $x < a$, and so

$$a = X_a = Y_{f(a)} = f(a),$$

a contradiction. □

Proof of Theorem 2.3

We *claim* the following:

> If $(X, <)$ is a well-ordered set such that, for each $a \in X$, the section X_a is isomorphic to an ordinal, then X is isomorphic to an ordinal.

Let us first see that this claim suffices to prove the theorem. Let $P(a)$ be the property 'X_a is isomorphic to an ordinal'. Then, assuming the claim, $P(a)$ holds for all $a \in X$, by induction; appealing to the claim one last time, X itself is isomorphic to an ordinal. Finally, if X is isomorphic to two ordinals, they are isomorphic to one another, and hence are equal, by Lemma 2.8. So the theorem is proved from the claim, and we now only have to establish the claim. □

Proof of the Claim

Let $g_a : X_a \to Z(a)$ be an isomorphism for each $a \in X$, where $Z(a)$ is an ordinal. Note that $Z(a)$ and g_a are unique, by Lemmas 2.8 and 2.2. We can consider Z as a function on the set X. Let W be its range:

$$W = \{Z(a) : a \in X\}.$$

Now, if $x, y \in X$ and $x < y$, then $Z(x) \subset Z(y)$. For $Z(x)$ and $Z(y)$ are ordinals, and are not equal (since they are isomorphic to distinct sections of X); so one is a section of the other, by Lemma 2.7. It cannot be that $Z(y)$ is a section of $Z(x)$, else we could construct an isomorphism from X_y into its section X_x by composing g_y, the inclusion, and the inverse of g_x. So the function Z is a bijection, and indeed an isomorphism, from X to W (where W is ordered by inclusion). Thus W is well-ordered (being isomorphic to a well-ordered set). To finish the proof, we show that W is an ordinal. This holds because its members are ordinals, so any section W_a is equal to a. □

The ordinals thus form a sequence of well-ordered sets, each contained in the next, which go on for ever. One variant of Russell's Paradox, known as the *Burali-Forti paradox*, is the following assertion:

Theorem 2.5

The ordinal numbers do not form a set.

Proof

If there were a set O consisting of the ordinal numbers, then it would itself be an ordinal number, and so it would be a member of itself; but it is obviously greater than each of its members! □

Despite the fact that the ordinals do not form a set, it is still possible to think of them as if they were a set. In particular, the ordinals form an 'ordered class':

Theorem 2.6

For ordinals x and y, the following are equivalent:

(a) $x < y$;

(b) $x \in y$;

(c) $x \subset y$.

Moreover, exactly one of $x < y$, $x = y$, $y < x$ holds.

Proof

By Lemma 2.7, exactly one of $x \subset y$, $x = y$, $y \subset x$ holds. So the last assertion follows from the equivalence of (a) and (c). Now (a) requires a little interpretation: we write $x < y$ if x and y are members of some larger ordinal z for which this holds. Indeed, this doesn't depend on which ordinal z we take, since $x = z_x$ and $y = z_y$; and, if x is a section of y, then $z = y \cup \{y\}$ is an ordinal which contains both x and y. Now (a) and (b) are equivalent since $x < y$ if and only if $x \in z_y = y$. For (a) and (c), if $x < y$ then $x = z_x \subset z_y = y$; and if $x \not< y$ then $x = y$ or $y < x$, the latter implying $y \subset x$, so $x \not\subset y$. □

It is also possible (and important) to do *induction* over all the ordinals:

Theorem 2.7

Let P be a property of ordinals. Suppose that, whenever x is an ordinal for which $P(y)$ holds for all ordinals $y < x$, then $P(x)$ holds. Then $P(x)$ holds for all ordinals.

Proof

To prove $P(x)$, it suffices to do induction over an ordinal containing x, such as $x \cup \{x\}$. □

Although we have defined ordinals in a uniform way, it is convenient to subdivide them into three types; in many applications, the methods of handling these types are quite different.

The first type of ordinal consists only of zero, the smallest ordinal. (In inductive proofs, we are taught to make a separate argument for the base case where the value of the parameter is zero.)

The second type consists of *successor ordinals*. We observed above that, for any ordinal α, the set $\alpha \cup \{\alpha\}$ is also an ordinal, called the *successor* of α. The positive natural numbers are all successor ordinals: $n+1$ is the successor of n. In general, the successor of α is the smallest ordinal which is greater than α. We write the successor of α as $s(\alpha)$

A non-zero ordinal λ is called a *limit ordinal* if it is the union of all its predecessors:

$$\lambda = \bigcup_{\alpha<\lambda} \alpha.$$

A successor ordinal is not a limit ordinal: for if $\lambda = \alpha \cup \{\alpha\}$, then the ordinals smaller than λ are all contained in α, and so is their union.

Theorem 2.8

Any non-zero ordinal is either a successor ordinal or a limit ordinal.

Proof

Suppose that λ is non-zero and is not a successor ordinal. Let

$$\mu = \bigcup_{\alpha<\lambda} \alpha.$$

Since a union of ordinals is an ordinal, we see that μ is an ordinal; and clearly $\mu \subseteq \lambda$, that is, $\mu \leq \lambda$. Suppose that $\mu < \lambda$. Then μ is one of the sets in the union defining itself; so it is the greatest ordinal less than λ. Since $\lambda > \mu$, we have $\lambda \geq s(\mu)$; and we cannot have strict inequality here, or else $s(\mu)$ would be in the family whose union is μ, that is, $s(\mu) \leq \mu$, which is clearly false. So $\lambda = s(\mu)$ is a successor ordinal, a contradiction. We conclude that $\lambda = \mu$ and so λ is a limit ordinal. □

As we remarked earlier, many arguments about ordinals (especially induction arguments) require different methods for the cases of zero, successors, and limit ordinals. For example, we can re-formulate induction so that it looks more like ordinary induction (at least in the case of successor ordinals):

Theorem 2.9

Let P be a property of ordinals. Assume that

- $P(0)$ is true;
- $P(\alpha)$ implies $P(s(\alpha))$ for any ordinal α;
- if λ is a limit ordinal and $P(\beta)$ holds for all $\beta < \lambda$, then $P(\lambda)$ holds.

Then $P(\alpha)$ is true for all ordinals α.

Proof

We let Q be the property which holds at α if and only if $P(\beta)$ holds for all $\beta \leq \alpha$. Now we verify the hypotheses of Theorem 2.7 for Q. Suppose that $Q(\beta)$ is true for all $\beta < \alpha$. Then *a fortiori*, $P(\beta)$ holds for all $\beta < \alpha$. If $\alpha = 0$ or α is a limit ordinal, it follows from the hypotheses of the theorem that $P(\alpha)$ holds. Suppose that $\alpha = s(\beta)$. Then $P(\beta)$ holds so, by hypothesis, $P(\alpha)$ holds.

Thus $P(\alpha)$ is true in all cases. Since we know that $P(\beta)$ is true for all $\beta < \alpha$, we now deduce that $Q(\alpha)$ holds.

By Theorem 2.7, $Q(\alpha)$ (and hence $P(\alpha)$) is true for all ordinals α. □

2.3 The hierarchy of sets

Zermelo's hierarchy was an approach to set theory aimed at avoiding the paradoxes. With modifications suggested by Fraenkel, it is the approach most commonly used today. Zermelo's idea was to build the sets in well-ordered stages. If sets which are not members of themselves continue to appear at every stage, then there will be no stage at which they all exist and can be gathered into a set, so Russell's Paradox will not arise. The stages of the construction will be indexed by the ordinals (since any well-ordered set is order-isomorphic to a unique ordinal). Let V_α be the set of all sets constructed at stage α. Then the inductive definition is as follows:

$$\begin{aligned} V_0 &= \varnothing \\ V_{s(\alpha)} &= \mathcal{P}\, V_\alpha \\ V_\lambda &= \bigcup_{\alpha<\lambda} V_\alpha \text{ for limit ordinals } \lambda. \end{aligned}$$

Thus,

$$\begin{aligned} V_1 &= \{\varnothing\}; \\ V_2 &= \{\varnothing, \{\varnothing\}\}; \\ V_3 &= \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\varnothing, \{\varnothing\}\}\}; \end{aligned}$$

V_4 is a set with sixteen elements; and so on.

The main content of Zermelo's approach is that this procedure gives us all sets. That is, every set is contained in V_α for some ordinal α. Symbolically, we can write

$$V = \bigcup_{\alpha \in \mathrm{On}} V_\alpha,$$

where V is the 'class' of all sets and On the 'class' of all ordinal numbers. (This is only an *aide-mémoire*, not a mathematical expression!)

We derive now a couple of facts about the Zermelo hierarchy, using transfinite induction. First, it really is a hierarchy: the sets get larger as we progress. (This is not obvious; each set is the power set of its predecessor, and most sets X don't satisfy $X \subseteq \mathcal{P}\,X$.)

Theorem 2.10

$V_\alpha \subseteq V_\beta$ for $\alpha < \beta$.

Proof

The proof actually requires a double induction, which we separate into two steps.

Step 1

We *claim* that it suffices to prove that $V_\alpha \subseteq V_{s(\alpha)}$ for all ordinals α. For suppose that this holds, and suppose that $\alpha < \beta$ and $V_\alpha \not\subseteq V_\beta$. Let β be the smallest such ordinal (for given α). Clearly $\beta \neq 0$, so there are two cases:

Case 1: β is a successor ordinal, say $\beta = s(\gamma)$. Then $\alpha \leq \gamma$, so

$$V_\alpha \subseteq V_\gamma \subseteq V_{s(\gamma)} = V_\beta,$$

the second inclusion following from the claim.

Case 2: β is a limit ordinal. Then $V_\alpha \subseteq V_\lambda$ for all $\alpha \leq \lambda < \beta$, so

$$V_\alpha \subseteq \bigcup_{\lambda < \beta} V_\lambda = V_\beta.$$

Step 2

We prove that $V_\alpha \subseteq V_{s(\alpha)}$ for all ordinals α by induction.

Case 1: $\alpha = 0$. Then V_α is the empty set, which is a subset of any set.

Case 2: $\alpha = s(\gamma)$ for some γ. Take $x \in V_\alpha$. Then

$$x \subseteq V_\gamma \subseteq V_{s(\gamma)} = V_\alpha,$$

so $x \in \mathcal{P}\,V_\alpha = V_{s(\alpha)}$.

Case 3: α is a limit ordinal. Take $x \in V_\alpha$. Then $x \in V_\delta$ for some $\delta < \alpha$; so

$$x \in V_{s(\delta)} \subseteq V_{s(\alpha)},$$

since $V_{s(\delta)} = \mathcal{P}\,V_\delta \subseteq \mathcal{P}\,V_\alpha = V_{s(\alpha)}$. □

Theorem 2.11

For any ordinal α, we have $\alpha \subseteq V_\alpha$, and hence $\alpha \in V_{s(\alpha)}$.

Proof

Again the proof is by induction.

Case 1: $\alpha = 0 = \varnothing$: then α is a subset of any set!

Case 2: $\alpha = s(\gamma) = \gamma \cup \{\gamma\}$. Now $\gamma \subseteq V_\gamma \subseteq V_\alpha$, and $\gamma \in V_{s(\gamma)} = V_\alpha$, both by the induction hypothesis; so $\alpha \subseteq V_\alpha$.

Case 3: α is a limit ordinal. Then

$$\alpha = \bigcup_{\delta < \alpha} \delta \subseteq \bigcup_{\delta < \alpha} V_\delta = V_\alpha.$$

□

There are two drawbacks with this simple approach to rigorous set theory. First, we appear to be defining sets in terms of ordinals, which are themselves sets: is our procedure not circular? Second, it is not easy, from this approach, to prove things about sets.

In Chapter 6, we will deduce from Zermelo's hierarchy a number of assertions about sets. We will then take these assertions as axioms for a formal theory of sets. As always in mathematics, axiomatization represents a development in a field which has already achieved some mathematical maturity. Zermelo's insight gives us confidence in our formal manipulations with the axioms.

2.4 Ordinal arithmetic

'Can you do Addition?' the White Queen asked. 'What's one and one and one and one and one and one and one and one and one and one?'

> ‘I don’t know,’ said Alice. ‘I lost count.’
> ‘She can’t do Addition,’ the Red Queen interrupted.
>
> Lewis Carroll, *Through the Looking-Glass, and what Alice found there*

The ordinals we have defined are a kind of numbers – indeed, they include the natural numbers – although they are designed for ‘counting’ well-ordered sets, not arbitrary sets. Accordingly, we would like to do arithmetic with them; in particular, to add and multiply them. There are two approaches to this.

First is the structural approach: we figure out how to ‘add’ and ‘multiply’ well-ordered sets, and use these to define the operations on ordinals. Essentially, we take the ‘ordered sum’ of two ordered sets to be their disjoint union, with each element of the first set preceding each element of the second set. Since the sets may not be disjoint (as will indeed happen if they are non-zero ordinals), we ‘tag’ them to make disjoint copies, as in the first proof of the Schröder–Bernstein theorem. (We take the tags to be the first two ordinals, $0 = \varnothing$ and $1 = \{\varnothing\}$, though in fact any two distinct tags would do.) For multiplication, we take the ‘lexicographic product’ (the cartesian product with the lexicographic order).

Definition 2.1

Let $(X, <_X)$ and $(Y, <_Y)$ be ordered sets. We define the *ordered sum* of these sets to be $(Z, <_Z)$, where

- $Z = (X \times \{0\}) \cup (Y \times \{1\})$;
- $(x_1, 0) <_Z (x_2, 0)$ if and only if $x_1 <_X x_2$;
- $(y_1, 1) <_Z (y_2, 1)$ if and only if $y_1 <_Y y_2$;
- $(x, 0) <_Z (y, 1)$ for all $x \in X$, $y \in Y$.

We define the *lexicographic product* of the sets to be $(W, <_W)$, where

- $W = X \times Y$;
- $(x_1, y_1) <_W (x_2, y_2)$ if $y_1 <_Y y_2$;
- $(x_1, y) <_W (x_2, y)$ if $x_1 <_X x_2$.

It can be shown that the ordered sum and lexicographic product of totally ordered sets are totally ordered sets; and ordered sum and lexicographic product of well-ordered sets are well-ordered sets.

Definition 2.2

Let α and β be ordinals. We define $\alpha + \beta$ to be the unique ordinal isomorphic to the ordered sum of α and β, and $\alpha \cdot \beta$ to be the unique ordinal isomorphic to their lexicographic product.

The second definition is more formal, using transfinite induction.

Definition 2.3

- $\alpha + 0 = \alpha$.
- $\alpha + s(\beta) = s(\alpha + \beta)$.
- If λ is a limit ordinal then $\alpha + \lambda = \bigcup_{\beta<\lambda} \alpha + \beta$.

- $\alpha \cdot 0 = 0$.
- $\alpha \cdot s(\beta) = \alpha \cdot \beta + \alpha$.
- If λ is a limit ordinal then $\alpha \cdot \lambda = \bigcup_{\beta<\lambda} \alpha \cdot \beta$.

It can be shown that this definition agrees with the previous one. (The proof, of course, is by induction.) The advantage of this approach is its flexibility. The last clause in the definition is essentially the same in both cases, and is easily modified for other situations. So if we want to define, for example, exponentiation of ordinals, we can proceed as follows:

Definition 2.4

- $\alpha^0 = 1$.
- $\alpha^{s(\beta)} = \alpha^\beta \cdot \alpha$.
- If λ is a limit ordinal then $\alpha^\lambda = \bigcup_{\beta<\lambda} \alpha^\beta$.

Now the *natural numbers* are just the ordinals less than ω, so we have incidentally defined them and shown how to add and multiply them. (The definitions are most easily obtained from Definition 2.3 by dropping the clauses about limit ordinals.) Now the properties of natural numbers that we used in Section 1.8 can be proved by induction: see Exercise 2.4. We use the usual notation for the natural numbers, so that, for example, $2 = \{\varnothing, \{\varnothing\}\}$. The first infinite ordinal (the set of all natural numbers) is usually denoted by ω.

Various properties of ordinal arithmetic can be proved (see Exercise 2.6 for some of these). Perhaps more interesting are the properties which are not true. For example,

$$1 + \omega = \omega \neq \omega + 1.$$

So the commutative law for addition and the cancellation law both fail. Indeed, if we take an infinite sequence and place a new element to the left, we still have an infinite sequence; but if we place a new element to the right, we get a different ordered set (one with a greatest element).

Ordinals soon grow to a point where it is not easy to imagine the resulting sets. For example, ω^2 is an infinite sequence of infinite sequences. Further along the sequence, we come to ω^3, ω^4, ..., and then

$$\omega^\omega = \bigcup_{n \in \mathbb{N}} \omega^n.$$

But we can continue, to reach ordinals like $\omega^{\omega^{\cdot^{\cdot^{\cdot^{\omega}}}}}$, (a 'tower' of $n+1$ omegas), which we denote by ω_n. More generally, we can define ω_α by

- $\omega_0 = \omega$,
- $\omega_{s(\alpha)} = \omega^{\omega_\alpha}$,
- $\omega_\lambda = \bigcup_{\alpha<\lambda} \omega_\alpha$ for limit ordinals λ.

Then ω_ω is an infinite 'tower' $\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}$. Eventually, we reach the unimaginably large $\epsilon = \omega_{\omega_{\cdot_{\cdot_{\cdot}}}}$.

However, vast as these ordinals are, *they are all countable*, since each is a countable union of countable ordinals. Somewhere, even further down the line, lies the first uncountable ordinal ...

EXERCISES

2.1 Prove that the ordered sum and lexicographic product of totally ordered (resp., well-ordered) sets is totally ordered (resp., well-ordered).

2.2 Let X be any set, and define X^* to be the set of all finite sequences of elements of X. Prove that, if X can be well-ordered, then so can X^*. [*Hint*: $X^* = \bigcup_{n \in \mathbb{N}} X^n$; arrange the n-tuples in dictionary order.] Show that dictionary order on the set X^* is never a well-ordering if $|X| > 1$.

2.3 According to our definition, any natural number can be described in symbols as a sequence whose terms are the empty set $\varnothing$, opening and closing curly brackets { and }, and commas ,. For example, the number 4 is

$$\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}\}$$

with eight occurrences of $\varnothing$, eight of each sort of bracket, and seven commas. How many occurrences of each symbol are there in the expression for the number n?

2.4 Prove the properties of addition and multiplication of natural numbers used in Section 1.8.

2.5 Prove that the two definitions of ordinal addition and multiplication agree.

2.6 Prove the following properties of ordinal arithmetic:

(a) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

(b) $(\alpha + \beta) \cdot \gamma = \alpha \cdot \gamma + \beta \cdot \gamma$.

(c) $\alpha^{\beta+\gamma} = \alpha^\beta \cdot \alpha^\gamma$.

2.7 (a) Show that, if $\gamma + \alpha = \gamma + \beta$, then $\alpha = \beta$.

[*Hint*: The identity map from $\gamma + \alpha$ to $\gamma + \beta$ maps γ to γ and induces an isomorphism from α to β.]

(b) Show that, if $\gamma \cdot \alpha = \gamma \cdot \beta$ and $\gamma \neq 0$, then $\alpha = \beta$.

2.8 Let $(X_i)_{i \in I}$ be a family of non-empty sets. Prove that, under either of the following conditions, the cartesian product $\prod_{i \in I} X_i$ is non-empty:

(a) $X_i = X$ for all $i \in I$;

(b) X_i is well-ordered for all $i \in I$.

2.9 Let X be a subset of the set of real numbers, which is well-ordered by the natural order on $\mathbb{R}$. Prove that X is finite or countable.

[*Hint*: Let $X = \{x_\beta : \beta < \alpha\}$ for some ordinal α, and assume that $\beta < \gamma$ implies $x_\beta < x_\gamma$. Choose a real number q_β in the interval $(x_\beta, x_{s(\beta)})$ for all $\beta < \alpha$. Prove that these rational numbers are all distinct.]

Case 2: $I \subset R(P)$. By Theorem 6.2 (which uses the Axiom of Choice), there is a maximal ideal J of $B(P)$ containing I. Since $B(P)$ is a commutative ring with identity, $B(P)/J$ is a field, all of whose elements satisfy the polynomial $x^2 - x$ of degree 2. So $|B(P)/J| = 2$, and $B(P)/J \cong R_2$. So there is a homomorphism $v : B(P) \to R_2$ whose kernel is J. Thus, there is a valuation v such that $v(1+s) = 0$ (in other words, $v(s) = 1$) for all $s \in \Sigma$; that is, Σ is satisfiable, as required. □

Remark We have seen that various mathematical facts (such as the infinite Four-Colour Theorem, and the fact that every set can be totally ordered) can be proved using the Propositional Compactness Theorem. So these are also consequences of the Axiom of Choice. It is known, however, that Propositional Compactness is a 'weaker' principle than AC: there is no proof of AC using Propositional Compactness, and indeed models of set theory have been constructed in which Propositional Compactness is true but AC fails. See the chapter by John Truss in Kaye and Macpherson [29] for a survey of this.

6.3 Cardinals

We now develop a theory of cardinal numbers. As in the case of the ordinals, we can state at the start the theorem that we want to prove from our definition. Since cardinal numbers should measure the size of arbitrary sets, we require a theorem which says:

Theorem 6.6

Every set has a bijection to a unique cardinal number.

It has to be said that no really adequate theory of cardinal numbers exists in ZF. Bertrand Russell attempted a definition in which, for example, the number 2 is the class of all 2-element sets. With this definition, however, 2 is not even a set, and certainly not a 2-element set! However, with the Axiom of Choice, things are much simpler. We work in ZFC for the rest of this section. (In ZF, this theory applies to those sets which can be well-ordered.)

Definition 6.1

A *cardinal* is an ordinal α with the property that there is no bijection between α and any section of α.

Note that, according to this definition, all finite ordinals (that is, all natural numbers) are cardinals; and ω is a cardinal, since it is infinite but all its sections are finite. However, $\omega + 1$ is not a cardinal, since it is countable (that is, has a bijection to its section ω).

Proof of Theorem 6.6 (in ZFC)

Let X be a set. By WO, X can be well-ordered; that is, there is a bijection from X to some ordinal. Now there is a smallest ordinal α in the set of ordinals bijective with X. And α is a cardinal; for, if there was a bijection from α to a section β, then there would be a bijection from X to β, contrary to the choice of α.

Now, if X has a bijection to two cardinal numbers α and β, then there is a bijection between α and β, contradicting the fact that the smaller is a section of the larger. □

We denote the cardinal of the set X (the unique cardinal bijective with X) by $|X|$. Note that, if α is a cardinal, then $|\alpha| = \alpha$.

Cantor introduced the *aleph notation* for infinite cardinals. (The letter $\aleph$, 'aleph', is the first letter of the Hebrew alphabet.) This is a function from ordinals to cardinals, defined by transfinite recursion as follows:

- $\aleph_0 = \omega$;
- $\aleph_{s(\alpha)}$ is the smallest cardinal greater than $\aleph_\alpha$;
- if λ is a limit ordinal then

$$\aleph_\lambda = \bigcup_{\beta<\lambda} \aleph_\beta.$$

It is not obvious that $\aleph_\lambda$ is a cardinal. It is certainly an ordinal, since it is a union of ordinals. Suppose that it were bijective with a section of itself. This section could not contain all $\aleph_\beta$; but if some $\aleph_\beta$ does not lie in the section, then the restriction of the bijection takes $\aleph_\beta$ into a section of itself, a contradiction.

So we have two notations for the ordinal describing the infinite sequence of natural numbers, namely ω and $\aleph_0$. We use the first if we are thinking of it as an ordinal, and the second when we regard it as a cardinal. Note that $\aleph_1$ is the first uncountable ordinal.

There is an order relation defined on cardinals, since they are special kinds of ordinals. In fact, we have:

$$|X| \leq |Y| \text{ if and only if there is an injective function from } X \text{ to } Y.$$

For certainly $|X| \leq |Y|$ implies that $|X|$ (as an ordinal) is a subset of $|Y|$. Suppose that an injective function $f : \alpha \to \beta$ exists, where α and β are cardinals. Then there is a bijective function from α to an ordinal not exceeding β (since the image of f is well-ordered and contained in β); so the cardinal α is not greater than β.

The *Schröder–Bernstein Theorem* can be written in terms of cardinals as follows.

Theorem 6.7

For any two sets X and Y, if $|X| \leq |Y|$ and $|Y| \leq |X|$ then $|X| = |Y|$.

Now we will define arithmetic operations (addition, multiplication and exponentiation) of cardinals. The simplest way to do this is to mirror the operations of disjoint union, cartesian product, and set of functions: that is, for cardinals α and β, we define

- $\alpha + \beta = |(\alpha \times \{0\}) \cup (\beta \times \{1\})|$;
- $\alpha \cdot \beta = |\alpha \times \beta|$;
- $\alpha^\beta = |\alpha^\beta|$;

where in the third (confusing) equation, on the right-hand side A^B means the set of all functions from B to A, and not the ordinal exponentiation defined in Chapter 2. We can write these definitions as statements about the cardinalities of arbitrary sets, as follows:

- $|A \cup B| = |A| + |B|$ if A and B are disjoint;
- $|A \times B| = |A| \cdot |B|$;
- $|A^B| = |A|^{|B|}$.

It turns out that cardinal addition and multiplication tables are very easy to learn!

Theorem 6.8

Let α and β be non-zero cardinals, at least one of which is infinite. Then

$$\alpha + \beta = \alpha \cdot \beta = \max\{\alpha, \beta\}.$$

Proof

We claim that it is enough to prove that

$$\alpha \cdot \alpha = \alpha$$

for any infinite cardinal α. For, if $\beta \leq \alpha$, then there is an obvious injection from $\alpha \times \beta$ to $\alpha \times \alpha$; and there is an injection from $(\alpha \times \{0\}) \cup (\beta \times \{1\})$ to

$$(\alpha \times \{0\}) \cup (\alpha \times \{1\}) = \alpha \times 2,$$

where $2 = \{0, 1\}$. So $\alpha \cdot \beta \leq \alpha \cdot \alpha$ and $\alpha + \beta \leq \alpha \cdot \alpha$. On the other hand, clearly $\alpha \leq \alpha + \beta$ and $\alpha \leq \alpha \cdot \beta$ (the latter if $\beta \neq 0$).

We have already proved in Chapter 1 that $\aleph_0 \cdot \aleph_0 = \aleph_0$. The general proof follows similar lines but is rather more complicated. We suppose that the theorem is false. In that case, there will be a smallest cardinal α such that $\alpha \cdot \alpha > \alpha$. We let $P = \alpha \times \alpha$. Now recall that α is an ordinal, which means that it is a set of ordinals. In the following argument, we use *ordinal addition*! We define, for ordinals $\beta < \alpha$, the subset P_β of P by the rule

$$P_\beta = \{(x, y) \in P : x + y = \beta\}.$$

These sets correspond to the north-east to south-west arrows in Figure 1.3. We claim that the sets P_β for $\beta < \alpha$ form a partition of P. Clearly they are pairwise disjoint. To show that every point lies in one of them, we need to show that, if $x, y < \alpha$, then $x + y < \alpha$: but this follows from the fact that the theorem is true for cardinals less than α. (The ordinal $x + y$ has cardinality $|x| + |y| = \max\{|x|, |y|\}$.)

Now we well-order each 'diagonal strip' P_β by the 'lexicographic' rule

$$(x, y) < (x', y') \text{ if either } x < x', \text{ or } x = x', y < y'.$$

This is easily seen to be a well-ordering – it is the ordering induced on P_β as a subset of $\alpha \times \alpha$. So now we can well-order all of P by putting $(x, y) < (x', y')$ if either $(x, y) \in P_\beta$, $(x', y') \in P_\gamma$ with $\beta < \gamma$, or $(x, y) < (x', y')$ within P_β under the ordering already defined. This gives a well-ordering of P.

Let θ be the unique ordinal isomorphic to P. We have $\theta > \alpha$, since $|P| > \alpha$ by assumption. So there is a point (u, v) in P such that the section (u, v) is isomorphic to α. Suppose that $(u, v) \in P_\beta$: that is, $u + v = \beta$. Then all points $(x, y) \in P$ with $(x, y) < (u, v)$ satisfy $x + y \leq \beta$, whence $x, y \leq \beta$; so this entire section of P is contained in $s(\beta) \times s(\beta)$. We conclude that

$$|s(\beta) \times s(\beta)| = \alpha > |s(\beta)|,$$

a contradiction since $s(\beta) < \alpha$. (An infinite successor ordinal cannot be a cardinal, and certainly $|\beta| < \alpha$.)

The theorem is proved. □

The theorem implies, in particular, that the union of at most α sets, each of cardinality at most α, has cardinality at most α, for any infinite cardinal α. So, using sums and products, we cannot build ever larger sets.

Exponentiation is much less trivial, however, and certainly has the ability to construct larger sets. We give a brief introduction. First, we have

Theorem 6.9

(a) For any set X, $|\mathcal{P}\,X| = 2^{|X|}$.

(b) $|\mathbb{R}| = 2^{\aleph_0}$.

Proof

(a) We have to produce a bijection between the set of subsets of X and the set of functions from X to $2 = \{0,1\}$. We do this by representing a subset Y of X by its *characteristic function* χ_Y, defined as follows:

$$\chi_Y(x) = \begin{cases} 1 & \text{if } x \in Y, \\ 0 & \text{if } x \notin Y. \end{cases}$$

Then distinct sets have distinct characteristic functions; and any function $F : X \to 2$ is the characteristic function of some set, namely the set $\{x \in X : f(x) = 1\}$. So we have a bijection as required.

(b) First, note that the cardinality of $\mathbb{R}$ is the same as that of the unit interval $(0,1)$: the map $f(x) = \tan \pi(x - \frac{1}{2})$ is a bijection from $(0,1)$ to $\mathbb{R}$. Moreover, we can regard an element of $2^{\mathbb{N}}$ as an infinite sequence of zeros and ones. Now we have an injection from $2^{\mathbb{N}}$ to $(0,1)$ by regarding the infinite sequence as an infinite decimal expansion (which happens to use only zeros and ones); and an injection from $(0,1)$ to $2^{\mathbb{N}}$ by taking the base 2 expansion of a number in the unit interval (resolving ambiguities by assuming that the base 2 expansion of a rational whose denominator is a power of 2 ends with infinitely many zeros rather than with infinitely many ones). By the Schröder–Bernstein Theorem, the cardinalities are equal. □

Thus, Cantor's Theorem (Theorem 1.10) can be translated into the form

Theorem 6.10

For any cardinal α, $2^\alpha > \alpha$. □

Cardinal arithmetic satisfies some (but of course not all) of the laws of the arithmetic of the natural numbers. In particular, it is true that $(\alpha^\beta)^\gamma = \alpha^{\beta\cdot\gamma}$, as is shown by producing a bijection between these sets (see Exercise 6.1). This simple observation has the following consequence:

Theorem 6.11

Let α and β be cardinals, with α infinite and $2 \le \beta \le 2^\alpha$. Then $\beta^\alpha = 2^\alpha$.

Proof

This follows from

$$2^\alpha \le \beta^\alpha \le (2^\alpha)^\alpha = 2^{\alpha\cdot\alpha} = 2^\alpha,$$

on using the Schröder–Bernstein Theorem. □

So much of the mystery of cardinal arithmetic lies in the function $\alpha \to 2^\alpha$. By Cantor's Theorem, we have $2^{\aleph_\alpha} \ge \aleph_{s(\alpha)}$ for any ordinal α. Do we have equality or not? The famous *Continuum Hypothesis* asserts that $2^{\aleph_0} = \aleph_1$. This was one of the problems posed in 1900 to the mathematical community by David Hilbert, to guide the development of mathematics in the twentieth century. In Hilbert's words (as translated by Dr Mary Winston Newson in the *Bulletin of the American Mathematical Society*), quoted in [8],

> Two assemblages, i.e. two assemblages of ordinary real numbers or points, are said to be (according to Cantor) equivalent or of *equal cardinal number*, if they can be brought into a relation to one another such that to every number of the one assemblage corresponds one and only one definite number of the other. The investigations of Cantor on such assemblages of points suggest a very plausible theorem, which nevertheless, in spite of the most strenuous efforts, no one has succeeded in proving. This is the theorem:
>
> Every system of infinitely many real numbers, i.e. every assemblage of numbers (or points), is either equivalent to the assemblage of natural integers, $1, 2, 3, \ldots$, or to the assemblage of all real numbers and therefore to the *continuum*, that is, to the points on a line: *as regards equivalence there are, therefore, only two assemblages of numbers, the countable assemblage and the continuum.*

Here Hilbert is using the term 'assemblage' for our 'infinite set', and considering subsets of $\mathbb{R}$. Since $|\mathbb{R}| = 2^{\aleph_0}$, he asks whether it is true that there is no infinite cardinal strictly between $\aleph_0$ and $2^{\aleph_0}$, that is, whether $2^{\aleph_0} = \aleph_1$. Its plausibility was reinforced when Gödel proved that it cannot be disproved in ZFC. Thirty years later, however, Hilbert was answered in a way he would not have expected by Cohen, who showed that it cannot be proved in ZFC either. By a new technique known as *forcing*, he constructed a model of ZFC in which $2^{\aleph_0} = \aleph_2$.