SECTION 22

THE SCHRÖDER–BERNSTEIN THEOREM

The purpose of counting is to compare the size of one set with that of another; the most familiar method of counting the elements of a set is to arrange them in some appropriate order. The theory of ordinal numbers is an ingenious abstraction of the method, but it falls somewhat short of achieving the purpose. This is not to say that ordinal numbers are useless; it just turns out that their main use is elsewhere, in topology, for instance, as a source of illuminating examples and counterexamples. In what follows we shall continue to pay some attention to ordinal numbers, but they will cease to occupy the center of the stage. (It is of some importance to know that we could in fact dispense with them altogether. The theory of cardinal numbers can be constructed with the aid of ordinal numbers, or without it; both kinds of constructions have advantages.) With these prefatory remarks out of the way, we turn to the problem of comparing the sizes of sets.

The problem is to compare the sizes of sets when their elements do not appear to have anything to do with each other. It is easy enough to decide that there are more people in France than in Paris. It is not quite so easy, however, to compare the age of the universe in seconds with the population of Paris in electrons. For some mathematical examples, consider the following pairs of sets, defined in terms of an auxiliary set A: (i) $X = A$, $Y = A^+$; (ii) $X = \mathcal{P}(A)$, $Y = 2^A$; (iii) X is the set of all one-to-one mappings of A into itself, Y is the set of all finite subsets of A. In each case we may ask which of the two sets X and Y has more elements. The problem is first to find a rigorous interpretation of the question and then to answer it.

The well ordering theorem tells us that every set can be well ordered. For well ordered sets we have what seems to be a reasonable measure of

size, namely, their ordinal number. Do these two remarks solve the problem? To compare the sizes of X and Y, may we just well order each of them and then compare ord X and ord Y? The answer is most emphatically no. The trouble is that one and the same set can be well ordered in many ways. The ordinal number of a well ordered set measures the well ordering more than it measures the set. For a concrete example consider the set ω of all natural numbers. Introduce a new order by placing 0 after everything else. (In other words, if n and m are non-zero natural numbers, then arrange them in their usual order; if, however, $n = 0$ and $m \neq 0$, let m precede n.) The result is a well ordering of ω; the ordinal number of this well ordering is $\omega + 1$.

If X and Y are well ordered sets, then a necessary and sufficient condition that ord $X <$ ord Y is that X be similar to an initial segment of Y. It follows that we could compare the ordinal sizes of two well ordered sets even without knowing anything about ordinal numbers; all we would need to know is the concept of similarity. Similarity was defined for ordered sets; the central concept for arbitrary unordered sets is that of equivalence. (Recall that two sets X and Y are called equivalent, $X \sim Y$, in case there exists a one-to-one correspondence between them.) If we replace similarity by equivalence, then something like the suggestion of the preceding paragraph becomes usable. The point is that we do not have to know what size is if all we want is to compare sizes.

If X and Y are sets such that X is equivalent to a subset of Y, we shall write

$$X \precsim Y.$$

The notation is temporary and does not deserve a permanent name. As long as it lasts, however, it is convenient to have a way of referring to it; a reasonable possibility is to say that Y *dominates* X. The set of those ordered pairs (X, Y) of subsets of some set E for which $X \precsim Y$ constitutes a relation in the power set of E. The symbolism correctly suggests some of the properties of the concept that it denotes. Since the symbolism is reminiscent of partial orders, and since a partial order is reflexive, antisymmetric, and transitive, we may expect that domination has similar properties.

Reflexivity and transitivity cause no trouble. Since each set X is equivalent to a subset (namely, X) of itself, it follows that $X \precsim X$ for all X. If f is a one-to-one correspondence between X and a subset of Y, and if g is a one-to-one correspondence between Y and a subset of Z, then we may restrict g to the range of f and compound the result with f; the

conclusion is that X is equivalent to a subset of Z. In other words, if $X \precsim Y$ and $Y \precsim Z$, then $X \precsim Z$.

The interesting question is that of antisymmetry. If $X \precsim Y$ and $Y \precsim X$, can we conclude that $X = Y$? This is absurd; the assumptions are satisfied whenever X and Y are equivalent, and equivalent sets need not be identical. What then can we say about two sets if all we know is that each of them is equivalent to a subset of the other? The answer is contained in the following celebrated and important result.

Schröder-Bernstein theorem. *If $X \precsim Y$ and $Y \precsim X$, then $X \sim Y$.*

REMARK. Observe that the converse, which is incidentally a considerable strengthening of the assertion of reflexivity, follows trivially from the definition of domination.

PROOF. Let f be a one-to-one mapping from X into Y and let g be a one-to-one mapping from Y into X; the problem is to construct a one-to-one correspondence between X and Y. It is convenient to assume that the sets X and Y have no elements in common; if that is not true, we can so easily make it true that the added assumption involves no loss of generality.

We shall say that an element x in X is the *parent* of the element $f(x)$ in Y, and, similarly, that an element y in Y is the parent of $g(y)$ in X. Each element x of X has an infinite sequence of *descendants*, namely, $f(x)$, $g(f(x))$, $f(g(f(x)))$, etc., and similarly, the descendants of an element y of Y are $g(y)$, $f(g(y))$, $g(f(g(y)))$, etc. This definition implies that each term in the sequence is a descendant of all preceding terms; we shall also say that each term in the sequence is an *ancestor* of all following terms.

For each element (in either X or Y) one of three things must happen. If we keep tracing the ancestry of the element back as far as possible, then either we ultimately come to an element of X that has no parent (these orphans are exactly the elements of $X - g(Y)$), or we ultimately come to an element of Y that has no parent ($Y - f(X)$), or the lineage regresses ad infinitum. Let X_X be the set of those elements of X that originate in X (i.e., X_X consists of the elements of $X - g(Y)$ together with all their descendants in X), let X_Y be the set of those elements of X that originate in Y (i.e., X_Y consists of all the descendants in X of the elements of $Y - f(X)$), and let X_∞ be the set of those elements of X that have no parentless ancestor. Partition Y similarly into the three sets Y_X, Y_Y, and Y_∞.

If $x \in X_X$, then $f(x) \in Y_X$, and, in fact, the restriction of f to X_X is a one-to-one correspondence between X_X and Y_X. If $x \in X_Y$, then x belongs to the domain of the inverse function g^{-1} and $g^{-1}(x) \in Y_Y$; in fact the restriction of g^{-1} to X_Y is a one-to-one correspondence between X_Y and Y_Y. If, finally, $x \in X_\infty$, then $f(x) \in Y_\infty$, and the restriction of f to X_∞ is a one-to-one correspondence between X_∞ and Y_∞; alternatively, if $x \in X_\infty$, then $g^{-1}(x) \in Y_\infty$, and the restriction of g^{-1} to X_∞ is a one-to-one correspondence between X_∞ and Y_∞. By combining these three one-to-one correspondences, we obtain a one-to-one correspondence between X and Y.

EXERCISE. Suppose that f is a mapping from X into Y and g is a mapping from Y into X. Prove that there exist subsets A and B of X and Y respectively, such that $f(A) = B$ and $g(Y - B) = X - A$. This result can be used to give a proof of the Schröder-Bernstein theorem that looks quite different from the one above.

By now we know that domination has the essential properties of a partial order; we conclude this introductory discussion by observing that the order is in fact total. The assertion is known as the comparability theorem for sets: it says that if X and Y are sets, then either $X \precsim Y$ or $Y \precsim X$. The proof is an immediate consequence of the well ordering theorem and of the comparability theorem for well ordered sets. Well order both X and Y and use the fact that either the well ordered sets so obtained are similar or one of them is similar to an initial segment of the other; in the former case X and Y are equivalent, and in the latter one of them is equivalent to a subset of the other.

SECTION 23

COUNTABLE SETS

If X and Y are sets such that Y dominates X and X dominates Y, then the Schröder-Bernstein theorem applies and says that X is equivalent to Y. If Y dominates X but X does not dominate Y, so that X is not equivalent to Y, we shall write

$$X \prec Y,$$

and we shall say that Y *strictly dominates* X.

Domination and strict domination can be used to express some of the facts about finite and infinite sets in a neat form. Recall that a set X is called finite in case it is equivalent to some natural number; otherwise it is infinite. We know that if $X \precsim Y$ and Y is finite, then X is finite, and we know that ω is infinite (§ 13); we know also that if X is infinite, then $\omega \precsim X$ (§ 15). The converse of the last assertion is true and can be proved either directly (using the fact that a finite set cannot be equivalent to a proper subset of itself) or as an application of the Schröder-Bernstein theorem. (If $\omega \precsim X$, then it is impossible that there exist a natural number n such that $X \sim n$, for then we should have $\omega \precsim n$, and that contradicts the fact that ω is infinite.)

We have just seen that a set X is infinite if and only if $\omega \precsim X$; next we shall prove that X is finite if and only if $X \prec \omega$. The proof depends on the transitivity of strict domination: if $X \precsim Y$ and $Y \precsim Z$, and if at least one of these dominations is strict, then $X \prec Z$. Indeed, clearly, $X \precsim Z$. If we had $Z \precsim X$, then we should have $Y \precsim X$ and $Z \precsim Y$ and hence (by the Schröder-Bernstein theorem) $X \sim Y$ and $Y \sim Z$, in contradiction to the assumption of strict domination. If now X is finite, then $X \sim n$ for some natural number n, and, since ω is infinite, $n \prec \omega$, so that $X \prec \omega$.

If, conversely, $X \prec \omega$, then X must be finite, for otherwise we should have $\omega \precsim X$, and hence $\omega \prec \omega$, which is absurd.

A set X is called *countable* (or *denumerable*) in case $X \precsim \omega$ and *countably infinite* in case $X \sim \omega$. Clearly a countable set is either finite or countably infinite. Our main purpose in the immediate sequel is to show that many set-theoretic constructions when performed on countable sets lead again to countable sets.

We begin with the observation that every subset of ω is countable, and we go on to deduce that every subset of each countable set is countable. These facts are trivial but useful.

If f is a function from ω *onto* a set X, then X is countable. For the proof, observe that for each x in X the set $f^{-1}(\{x\})$ is not empty (this is where the *onto* character of f is important), and consequently, for each x in X, we may find a natural number $g(x)$ such that $f(g(x)) = x$. Since the function g is a one-to-one mapping from X into ω, this proves that $X \precsim \omega$. The reader who worries about such things might have noticed that this proof made use of the axiom of choice, and he may want to know that there is an alternative proof that does not depend on that axiom. (There is.) The same comment applies on a few other occasions in this section and its successors but we shall refrain from making it.

It follows from the preceding paragraph that a set X is countable if and only if there exists a function from some countable set onto X. A closely related result is this: if Y is any particular countably infinite set, then a necessary and sufficient condition that a set X be countable is that there exist a function from Y onto X.

The mapping $n \to 2n$ is a one-to-one correspondence between ω and the set A of all even numbers, so that A is countably infinite. This implies that if X is a countable set, then there exists a function f that maps A onto X. Since, similarly, the mapping $n \to 2n + 1$ is a one-to-one correspondence between ω and the set B of all odd numbers, it follows that if Y is a countable set, then there exists a function g that maps B onto Y. The function h that agrees with f on A and with g on B (i.e., $h(x) = f(x)$ when $x \in A$ and $h(x) = g(x)$ when $x \in B$) maps ω onto $X \cup Y$. Conclusion: the union of two countable sets is countable. From here on an easy argument by mathematical induction proves that the union of a finite set of countable sets is countable. The same result can be obtained by imitating the trick that worked for two sets; the basis of the method is the fact that for each non-zero natural number n there exists a pairwise disjoint family $\{A_i\}$ $(i < n)$ of infinite subsets of ω whose union is equal to ω.

The same method can be used to prove still more. Assertion: there

exists a pairwise disjoint family $\{A_n\}$ $(n \in \omega)$ of infinite subsets of ω whose union is equal to ω. One way to prove this is to write down the elements of ω in an infinite array by counting down the diagonals, thus:

0	1	3	6	10	15	⋯
2	4	7	11	16	⋯	
5	8	12	17	⋯		
9	13	18	⋯			
14	19	⋯				
20	⋯					
⋯						

and then to consider the sequence of the rows of this array. Another way is to let A_0 consist of 0 and the odd numbers, let A_1 be the set obtained by doubling each non-zero element of A_0, and, inductively, let A_{n+1} be the set obtained by doubling each element of A_n, $n \geqq 1$. Either way (and there are many others still) the details are easy to fill in. Conclusion: the union of a countably infinite family of countable sets is countable. Proof: given the family $\{X_n\}$ $(n \in \omega)$ of countable sets, find a family $\{f_n\}$ of functions such that, for each n, the function f_n maps A_n onto X_n, and define a function f from ω onto $\bigcup_n X_n$ by writing $f(k) = f_n(k)$ whenever $k \in A_n$. This result combined with the result of the preceding paragraph implies that the union of a countable set of countable sets is always countable.

An interesting and useful corollary is that the Cartesian product of two countable sets is also countable. Since

$$X \times Y = \bigcup_{y \in Y} (X \times \{y\}),$$

and since, if X is countable, then, for each fixed y in Y, the set $X \times \{y\}$ is obviously countable (use the one-to-one correspondence $x \to (x, y)$), the result follows from the preceding paragraph.

EXERCISE. Prove that the set of all finite subsets of a countable set is countable. Prove that if every countable subset of a totally ordered set X is well ordered, then X itself is well ordered.

On the basis of the preceding discussion it would not be unreasonable to guess that every set is countable. We proceed to show that that is not so; this negative result is what makes the theory of cardinal numbers interesting.

Cantor's theorem. *Every set is strictly dominated by its power set, or, in other words,*

$$X \prec \mathcal{P}(X)$$

for all X.

PROOF. There is a natural one-to-one mapping from X into $\mathcal{P}(X)$, namely, the mapping that associates with each element x of X the singleton $\{x\}$. The existence of this mapping proves that $X \precsim \mathcal{P}(X)$; it remains to prove that X is not equivalent to $\mathcal{P}(X)$.

Assume that f is a one-to-one mapping from X onto $\mathcal{P}(X)$; our purpose is to show that this assumption leads to a contradiction. Write $A = \{x \in X : x \notin f(x)\}$; in words, A consists of those elements of X that are not contained in the corresponding set. Since $A \in \mathcal{P}(X)$ and since f maps X onto $\mathcal{P}(X)$, there exists an element a in X such that $f(a) = A$. The element a either belongs to the set A or it does not. If $a \in A$, then, by the definition of A, we must have $a \notin f(a)$, and since $f(a) = A$ this is impossible. If $a \notin A$, then, again by the definition of A, we must have $a \in f(a)$, and this too is impossible. The contradiction has arrived and the proof of Cantor's theorem is complete.

Since $\mathcal{P}(X)$ is always equivalent to 2^X (where 2^X is the set of all functions from X into 2), Cantor's theorem implies that $X \prec 2^X$ for all X. If in particular we take ω in the role of X, then we may conclude that the set of all sets of natural numbers is *uncountable* (i.e., not countable, non-denumerable), or, equivalently, that 2^ω is uncountable. Here 2^ω is the set of all infinite sequences of 0's and 1's (i.e., functions from ω into 2). Note that if we interpret 2^ω in the sense of ordinal exponentiation, then 2^ω is countable (in fact $2^\omega = \omega$).

SECTION 24

CARDINAL ARITHMETIC

One result of our study of the comparative sizes of sets will be to define a new concept, called *cardinal number*, and to associate with each set X a cardinal number, denoted by card X. The definitions are such that for each cardinal number a there exist sets A with card $A = a$. We shall also define an ordering for cardinal numbers, denoted as usual by $\leqq$. The connection between these new concepts and the ones already at our disposal is easy to describe: it will turn out that card $X =$ card Y if and only if $X \sim Y$, and card $X <$ card Y if and only if $X \prec Y$. (If a and b are cardinal numbers, $a < b$ means, of course, that $a \leqq b$ but $a \neq b$.)

The definition of cardinal numbers can be approached in several different ways, each of which has its strong advocates. To keep the peace as long as possible, and to demonstrate that the essential properties of the concept are independent of the approach, we shall postpone the basic construction. We proceed, instead, to study the arithmetic of cardinal numbers. In the course of that study we shall make use of the connection, described above, between cardinal inequality and set domination; that much of a loan from the future will be enough for the purpose.

If a and b are cardinal numbers, and if A and B are disjoint sets with card $A = a$ and card $B = b$, we write, by definition, $a + b =$ card $(A \cup B)$. If C and D are disjoint sets with card $C = a$ and card $D = b$, then $A \sim C$ and $B \sim D$; it follows that $A \cup B \sim C \cup D$, and hence that $a + b$ is unambiguously defined, independently of the arbitrary choice of A and B. Cardinal addition, thus defined, is commutative $(a + b = b + a)$, and associative $(a + (b + c) = (a + b) + c)$; these identities are immediate consequences of the corresponding facts about the formation of unions.

EXERCISE. Prove that if a, b, c, and d are cardinal numbers such that $a \leqq b$ and $c \leqq d$, then $a + c \leqq b + d$.

There is no difficulty about defining addition for infinitely many summands. If $\{a_i\}$ is a family of cardinal numbers, and if $\{A_i\}$ is a correspondingly indexed family of pairwise disjoint sets such that card $A_i = a_i$ for each i, then we write, by definition,

$$\sum_i a_i = \text{card } (\bigcup_i A_i).$$

As before, the definition is unambiguous.

To define the product ab of two cardinal numbers a and b, we find sets A and B with card $A = a$ and card $B = b$, and we write $ab =$ card $(A \times B)$. The replacement of A and B by equivalent sets yields the same value of the product. Alternatively, we could have defined ab by "adding a to itself b times"; this refers to the formation of the infinite sum $\sum_{i \in I} a_i$, where the index set I has cardinal number b, and where $a_i = a$ for each i in I. The reader should have no difficulty in verifying that this proposed alternative definition is indeed equivalent to the one that uses Cartesian products. Cardinal multiplication is commutative $(ab = ba)$ and associative $(a(bc) = (ab)c)$, and multiplication distributes over addition $(a(b + c) = ab + ac)$; the proofs are elementary.

EXERCISE. Prove that if a, b, c, and d are cardinal numbers such that $a \leqq b$ and $c \leqq d$, then $ac \leqq bd$.

There is no difficulty about defining multiplication for infinitely many factors. If $\{a_i\}$ is a family of cardinal numbers, and if $\{A_i\}$ is a correspondingly indexed family of sets such that card $A_i = a_i$ for each i, then we write, by definition,

$$\prod_i a_i = \text{card } (\bigtimes_i A_i).$$

The definition is unambiguous.

EXERCISE. If $\{a_i\}$ $(i \in I)$ and $\{b_i\}$ $(i \in I)$ are families of cardinal numbers such that $a_i < b_i$ for each i in I, then $\sum_i a_i < \prod_i b_i$.

We can go from products to exponents the same way as we went from sums to products. The definition of a^b, for cardinal numbers a and b, is most profitably given directly, but an alternative approach goes via repeated multiplication. For the direct definition, find sets A and B with card $A = a$ and card $B = b$, and write $a^b =$ card A^B. Alternatively, to define a^b "multiply a by itself b times." More precisely: form $\prod_{i \in I} a_i$, where the index set I has cardinal number b, and where $a_i = a$ for each i

in I. The familiar laws of exponents hold. That is, if a, b, and c are cardinal numbers, then

$$a^{b+c} = a^b a^c,$$

$$(ab)^c = a^c b^c,$$

$$a^{bc} = (a^b)^c.$$

EXERCISE. Prove that if a, b, and c are cardinal numbers such that $a \leqq b$, then $a^c \leqq b^c$. Prove that if a and b are finite and c is infinite, then $a^c = b^c$.

The preceding definitions and their consequences are reasonably straightforward and not at all surprising. If they are restricted to finite sets only, the result is the familiar finite arithmetic. The novelty of the subject arises in the formation of sums, products, and powers in which at least one term is infinite. The words "finite" and "infinite" are used here in a very natural sense: a cardinal number is *finite* if it is the cardinal number of a finite set, and *infinite* otherwise.

If a and b are cardinal numbers such that a is finite and b is infinite, then

$$a + b = b.$$

For the proof, suppose that A and B are disjoint sets such that A is equivalent to some natural number k and B is infinite; we are to prove that $A \cup B \sim B$. Since $\omega \lesssim B$, we may and do assume that $\omega \subset B$. We define a mapping f from $A \cup B$ to B as follows: the restriction of f to A is a one-to-one correspondence between A and k, the restriction of f to ω is given by $f(n) = n + k$ for all n, and the restriction of f to $B - \omega$ is the identity mapping on $B - \omega$. Since the result is a one-to-one correspondence between $A \cup B$ and B, the proof is complete.

Next: if a is an infinite cardinal number, then

$$a + a = a.$$

For the proof, let A be a set with card $A = a$. Since the set $A \times 2$ is the union of two disjoint sets equivalent to A (namely, $A \times \{0\}$ and $A \times \{1\}$), it would be sufficient to prove that $A \times 2$ is equivalent to A. The approach we shall use will not quite prove that much, but it will come close enough. The idea is to approximate the construction of the desired one-to-one correspondence by using larger and larger subsets of A.

Precisely speaking, let $\mathfrak{F}$ be the collection of all functions f such that the domain of f is of the form $X \times 2$, for some subset X of A, and such that f is a one-to-one correspondence between $X \times 2$ and X. If X is a countably infinite subset of A, then $X \times 2 \sim X$. This implies that the collection $\mathfrak{F}$ is not empty; at the very least it contains the one-to-one correspondences between $X \times 2$ and X for the countably infinite subsets X of A. The collection $\mathfrak{F}$ is partially ordered by extension. Since a straightforward verification shows that the hypotheses of Zorn's lemma are satisfied, it follows that $\mathfrak{F}$ contains a maximal element f with ran $f = X$, say.

Assertion: $A - X$ is finite. If $A - X$ were infinite, then it would include a countably infinite set, say Y. By combining f with a one-to-one correspondence between $Y \times 2$ and Y we could obtain a proper extension of f, in contradiction to the assumed maximality.

Since card X + card X = card X, and since card A = card X + card $(A - X)$, the fact that $A - X$ is finite completes the proof that card A + card A = card A.

Here is one more result in additive cardinal arithmetic: if a and b are cardinal numbers at least one of which is infinite, and if c is equal to the larger one of a and b, then

$$a + b = c.$$

Suppose that b is infinite, and let A and B be disjoint sets with card $A = a$ and card $B = b$. Since $a \leqq c$ and $b \leqq c$, it follows that $a + b \leqq c + c$, and since $c \leqq$ card $(A \cup B)$, it follows that $c \leqq a + b$. The result follows from the antisymmetry of the ordering of cardinal numbers.

The principal result in multiplicative cardinal arithmetic is that if a is an infinite cardinal number, then

$$a \cdot a = a.$$

The proof resembles the proof of the corresponding additive fact. Let $\mathfrak{F}$ be the collection of all functions f such that the domain of f is of the form $X \times X$ for some subset X of A, and such that f is a one-to-one correspondence between $X \times X$ and X. If X is a countably infinite subset of A, then $X \times X \sim X$. This implies that the collection $\mathfrak{F}$ is not empty; at the very least it contains the one-to-one correspondences between $X \times X$ and X for the countably infinite subsets X of A. The collection $\mathfrak{F}$ is partially ordered by extension. The hypotheses of Zorn's lemma are easily verified, and it follows that $\mathfrak{F}$ contains a maximal element f with ran $f = X$, say. Since (card X)(card X) = card X, the proof may be completed by showing that card X = card A.

Assume that card $X <$ card A. Since card A is equal to the larger one of card X and card $(A - X)$, this implies that card A = card $(A - X)$, and hence that card $X <$ card $(A - X)$. From this it follows that $A - X$

has a subset Y equivalent to X. Since each of the disjoint sets $X \times Y$, $Y \times X$, and $Y \times Y$ is infinite and equivalent to $X \times X$, hence to X, and hence to Y, it follows that their union is equivalent to Y. By combining f with a one-to-one correspondence between that union and Y, we obtain a proper extension of f, in contradiction to the assumed maximality. This implies that our present hypothesis (card $X <$ card A) is untenable and hence completes the proof.

EXERCISE. Prove that if a and b are cardinal numbers at least one of which is infinite, then $a + b = ab$. Prove that if a and b are cardinal numbers such that a is infinite and b is finite, then $a^b = a$.

SECTION 25

CARDINAL NUMBERS

We know quite a bit about cardinal numbers by now, but we still do not know what they are. Speaking vaguely, we may say that the cardinal number of a set is the property that the set has in common with all sets equivalent to it. We may try to make this precise by saying that the cardinal number of X is equal to the set of all sets equivalent to X, but the attempt will fail; there is no set as large as that. The next thing to try, suggested by analogy with our approach to the definition of natural numbers, is to define the cardinal number of a set X as some particular carefully selected set equivalent to X. This is what we proceed to do.

For each set X there are too many other sets equivalent to X; our first problem is to narrow the field. Since we know that every set is equivalent to some ordinal number, it is not unnatural to look for the typical sets, the representative sets, among ordinal numbers.

To be sure, a set can be equivalent to many ordinal numbers. A hopeful sign, however, is the fact that, for each set X, the ordinal numbers equivalent to X constitute a set. To prove this, observe first that it is easy to produce an ordinal number that is surely greater, strictly greater, than all the ordinal numbers equivalent to X. Suppose in fact that γ is an ordinal number equivalent to the power set $\mathcal{P}(X)$. If α is an ordinal number equivalent to X, then the set α is strictly dominated by the set γ (i.e., card $\alpha <$ card γ). It follows that we cannot have $\gamma \leqq \alpha$, and, consequently, we must have $\alpha < \gamma$. Since, for ordinal numbers, $\alpha < \gamma$ means the same thing as $\alpha \in \gamma$, we have found a set, namely γ, that contains every ordinal number equivalent to X, and this implies that the ordinal numbers equivalent to X do constitute a set.

Which one among the ordinal numbers equivalent to X deserves to be singled out and called the cardinal number of X? The question has only one natural answer. Every set of ordinal numbers is well ordered; the

least element of a well ordered set is the only one that seems to clamor for special attention.

We are now prepared for the definition: a *cardinal number* is an ordinal number α such that if β is an ordinal number equivalent to α (i.e., card α = card β), then $\alpha \leqq \beta$. The ordinal numbers with this property have also been called *initial numbers*. If X is a set, then card X, the cardinal number of X (also known as the *power* of X), is the least ordinal number equivalent to X.

EXERCISE. Prove that each infinite cardinal number is a limit number.

Since each set is equivalent to its cardinal number, it follows that if card X = card Y, then $X \sim Y$. If, conversely, $X \sim Y$, then card $X \sim$ card Y. Since card X is the least ordinal number equivalent to X, it follows that card $X \leqq$ card Y, and, since the situation is symmetric in X and Y, we also have card $Y \leqq$ card X. In other words card X = card Y if and only if $X \sim Y$; this was one of the conditions on cardinal numbers that we needed in the development of cardinal arithmetic.

A finite ordinal number (i.e., a natural number) is not equivalent to any finite ordinal number distinct from itself. It follows that if X is finite, then the set of ordinal numbers equivalent to X is a singleton, and, consequently, the cardinal number of X is the same as the ordinal number of X. Both cardinal numbers and ordinal numbers are generalizations of the natural numbers; in the familiar finite cases both the generalizations coincide with the special case that gave rise to them in the first place. As an almost trivial application of these remarks, we can now calculate the cardinal number of a power set $\mathcal{P}(A)$: if card $A = a$, then card $\mathcal{P}(A) = 2^a$. (Note that the result, though simple, could not have been stated before this; till now we did not know that 2 is a cardinal number.) The proof is immediate from the fact that $\mathcal{P}(A)$ is equivalent to 2^A.

If α and β are ordinal numbers, we know what it means to say that $\alpha < \beta$ or $\alpha \leqq \beta$. It follows that cardinal numbers come to us automatically equipped with an order. The order satisfies the conditions we borrowed for our discussion of cardinal arithmetic. Indeed: if card $X <$ card Y, then card X is a subset of card Y, and it follows that $X \lesssim Y$. If we had $X \sim Y$, then, as we have already seen, we should have card $X =$ card Y; it follows that we must have $X \prec Y$. If, finally, $X \prec Y$, then it is impossible that card $Y \leqq$ card X (for similarity implies equivalence), and hence card $X <$ card Y.

As an application of these considerations we mention the inequality

$$a < 2^a,$$

valid for all cardinal numbers a. Proof: if A is a set with card $A = a$, then $A \prec \mathcal{P}(A)$, hence card $A <$ card $\mathcal{P}(A)$, and therefore $a < 2^a$.

EXERCISE. If card $A = a$, what is the cardinal number of the set of all one-to-one mappings of A onto itself? What is the cardinal number of the set of all countably infinite subsets of A?

The facts about the ordering of ordinal numbers are at the same time facts about the ordering of cardinal numbers. Thus, for instance, we know that any two cardinal numbers are comparable (always either $a < b$, or $a = b$, or $b < a$), and that, in fact, every set of cardinal numbers is well ordered. We know also that every set of cardinal numbers has an upper bound (in fact, a supremum), and that, moreover, for every set of cardinal numbers, there is a cardinal number strictly greater than any of them. This implies of course that there is no largest cardinal number, or, equivalently, that there is no set that consists exactly of all the cardinal numbers. The contradiction, based on the assumption that there is such a set, is known as *Cantor's paradox*.

The fact that cardinal numbers are special ordinal numbers simplifies some aspects of the theory, but, at the same time, it introduces the possibility of some confusion that it is essential to avoid. One major source of difficulty is the notation for the arithmetic operations. If a and b are cardinal numbers, then they are also ordinal numbers, and, consequently, the sum $a + b$ has two possible meanings. The cardinal sum of two cardinal numbers is in general not the same as their ordinal sum. All this sounds worse than it is; in practice it is easy to avoid confusion. The context, the use of special symbols for cardinal numbers, and an occasional explicit warning can make the discussion flow quite smoothly.

EXERCISE. Prove that if α and β are ordinal numbers, then card $(\alpha + \beta)$ = card α + card β and card $(\alpha\beta)$ = (card α)(card β). Use the ordinal interpretation of the operations on the left side and the cardinal interpretation on the right.

One of the special symbols for cardinal numbers that is used very frequently is the first letter ($\aleph$, aleph) of the Hebrew alphabet. Thus in particular the smallest transfinite ordinal number, i.e., ω, is a cardinal number, and, as such, it is always denoted by $\aleph_0$.

Every one of the ordinal numbers that we have explicitly named so far is countable. In many of the applications of set theory an important role is played by the smallest uncountable ordinal number, frequently denoted by Ω. The most important property of ω is that it is an infinite well or-

dered set each of whose initial segments is finite; correspondingly, the most important property of Ω is that it is an uncountably infinite well ordered set each of whose initial segments is countable.

The least uncountable ordinal number Ω clearly satisfies the defining condition of a cardinal number; in its cardinal role it is always denoted by $\aleph_1$. Equivalently, $\aleph_1$ may be characterized as the least cardinal number strictly greater than $\aleph_0$, or, in other words, the immediate successor of $\aleph_0$ in the ordering of cardinal numbers.

The arithmetic relation between $\aleph_0$ and $\aleph_1$ is the subject of a famous old problem about cardinal numbers. How do we get from $\aleph_0$ to $\aleph_1$ by arithmetic operations? We know by now that the most elementary steps, involving sums and products, just lead from $\aleph_0$ back to $\aleph_0$ again. The simplest thing we know to do that starts with $\aleph_0$ and ends up with something larger is to form $2^{\aleph_0}$. We know therefore that $\aleph_1 \leqq 2^{\aleph_0}$. Is the inequality strict? Is there an uncountable cardinal number strictly less than $2^{\aleph_0}$? The celebrated *continuum hypothesis* asserts, as a guess, that the answer is no, or, in other words, that $\aleph_1 = 2^{\aleph_0}$. All that is known for sure is that the continuum hypothesis is consistent with the axioms of set theory.

For each infinite cardinal number a, consider the set $c(a)$ of all infinite cardinal numbers that are strictly less than a. If $a = \aleph_0$, then $c(a) = \varnothing$; if $a = \aleph_1$, then $c(a) = \{\aleph_0\}$. Since $c(a)$ is a well ordered set, it has an ordinal number, say α. The connection between a and α is usually expressed by writing $a = \aleph_\alpha$. An equivalent definition of the cardinal numbers $\aleph_\alpha$ proceeds by transfinite induction; according to that approach $\aleph_\alpha$ (for $\alpha > 0$) is the smallest cardinal number that is strictly greater than all the $\aleph_\beta$'s with $\beta < \alpha$. The *generalized continuum hypothesis* is the conjecture that $\aleph_{\alpha+1} = 2^{\aleph_\alpha}$ for each ordinal number α.

INDEX