Chapter 6

Sets and Infinity

6.1 Russell's Paradox

The first person to consider set theory in a rigorous way was the German mathematician Georg Cantor. He was led to the subject in a circuitous way, having begun his mathematical career by studying certain esoteric aspects of the theory of Fourier series. In the course of several papers he noticed that many of his proofs relied not on techniques from calculus, but rather on properties of certain sets of real numbers. This led him, in 1874, to publish the first paper in which a rigorous set theory was laid down.

Alas, he got off to a very bad start by defining a set as follows: "By a set we are to understand any collection into a whole M of definite and distinguishable elements of M." Before considering why this is such a poor definition, let us pause to understand what it means.

Intuitively, a set is just a collection of objects. There are no restrictions on the nature of the objects, which in principle could be anything. In mathematics we are most often interested in sets of numbers, but we could just as easily speak of sets of fruit, or people, or flowers, or anything else you care to think of. A set, therefore, is like a bag of groceries. The things in the bag are said to be the **elements** of the set. Generally speaking, when listing the elements of a set it is convenient to have every element listed exactly once. This is what Cantor had in mind by including the word "distinguishable" in his definition. By "definite" he meant that given any element we should be able to determine, for certain, whether that element is or is not in the set.

Let us now suppose that I have in mind a particular set and I desire to let you know precisely which set I am thinking of. How am I to do that? The simplest way is to list the elements of the set. That is fine, provided the set is not too big. But what if the set is very large, or infinite? In such a case, listing the elements is no longer possible. I would have to make do instead with a test that will allow a person to determine for himself whether a given element is actually in the set I am considering. If I say that I am thinking of the set of all positive even numbers, then I have not actually listed the elements of the set. But there is little doubt which set I have in mind.

So for any such description I provide, there should be a set consisting of all the elements satisfying that description. Perhaps I talk about the set of all left shoes or the set of all four-leaf clovers or the set of all abstract ideas. For each one of these descriptions there should be a set containing precisely those objects that satisfy the condition.

This all seems uncontroversial enough. So uncontroversial, in fact, that Cantor merrily developed a whole theory of sets based upon this definition. Much of this work consisted of detailed considerations on the nature of infinity, and we will consider much of that work in the remainder of this chapter. But first we pause to consider the luckless Gottlob Frege, whose two-volume work on set theory was based upon Cantor's definition.

Frege worked for years putting his treatise together. Just prior to the publication of the second volume, a young British mathematician named Bertrand Russell pointed out a serious flaw in Cantor's definition. The flaw he discovered is now known as Russell's paradox, in his honor.

We begin with the observation that every set either is, or is not, an element of itself. The set of all abstract ideas is itself an abstract idea. The set of all sets containing more than three elements is itself a set that contains more than three elements. These are example of sets that are elements of themselves.

Most sets are not elements of themselves. The set of all human beings on the Earth is not itself a human being, for example.

Therefore, we ought to be able to define a set S that contains precisely those sets that are not elements of themselves. According to Cantor's definition, this is perfectly acceptable. Also according to Cantor's definition, when confronted with any particular object, we should be able to determine whether that object is, or is not, an element of the set S.

The question is: Is S an element of itself?

The only sorts of things that S contains are sets that are not elements of themselves. It follows that if S is an element of the set S, then S must not be an element of itself. Since this is a contradiction, we are forced to consider the alternative.

Suppose that S is not an element of S. That means that S must be one of those sets that is not an element of itself. But sets that are not elements of themselves are precisely the sorts of things that S contains. So if we assume that S is not an element of itself, we are forced to the conclusion that S is indeed an element of itself. This is another contradiction.

No matter where we turn we arrive at a contradiction. This means there is a serious flaw in Cantor's definition. Russell pointed this out to Frege, who, to put it mildly, was rather nonplussed by the revelation.

So how do we resolve this difficulty? Well, there are professional mathematicians to this day who specialize in set theory, and they circumvent this problem by using an elaborate set of axioms to avoid just the sort of mathematical manure into which Cantor so innocently stepped. Alternatively, we can take the naive approach. This is where we simply take "set" and "element" to be undefined terms, confident that their meaning will be clear from the context in which they are used. That will be our resolution to the problem.

6.2 Another Paradox

In the following section we will go on to consider some of Cantor's ruminations on the nature of infinity. First, we consider one further paradox arising from Cantor's definition.

Let us define the set S to be the set of all positive integers that can not be uniquely described by a sentence containing fewer than two hundred letters. Since there are infinitely many numbers but only finitely many sentences containing fewer than two hundred letters, we know that S is not empty. In other words, there really are positive integers that can not be described with any sentence containing fewer than two hundred letters. Since S is a set of positive integers, we also know that S must contain a unique, smallest element. Let us call that element x.

This element x is a positive integer. It is uniquely described by the sentence "x is the smallest positive integer that can not be described by a sentence containing fewer than two hundred letters." The trouble is, we have just described x with a sentence containing fewer than two hundred letters.

Weird, no?

6.3 Infinity

Most students encounter infinity for the first time when learning calculus. There they encounter phrases like "The limit as x approaches infinity," which creates the impression that infinity is a final destination for wayward variables. It also tends to give the erroneous impression that infinity is an actual number, which it surely is not.

The word "infinity" comes from Latin words meaning "having no boundary." In the context of calculus, stating that x approaches infinity is merely shorthand for saying that x takes on ever increasing values, with no upper bound to stunt its growth.

As it happens, however, the word "infinity" is not used all that much by mathematicians. Far more common is the term "infinite," and that is what we will consider in this section.

How many perfect squares are there? Infinitely many. How many natural numbers are there? Also infinitely many. Does that mean there are as many perfect squares as there are natural numbers? That doesn't seem quite right. After all, every perfect square is a natural number, but surely there are natural numbers that are not perfect squares.

Let us suppose that A and B are sets and that $A \subset B$. Further suppose that both A and B contain only finitely many elements. In this case it seems natural to conclude that A must have fewer elements than B. We could prove this fact, but that is likely to make a simple fact seem rather confusing. Think of it this way: if B is a finite set and we remove some of its elements and throw them in the garbage, the result is a strictly smaller set. Not much to prove when you put it that way.

But if our ruminations on the relative numbers of perfect squares and natural numbers are correct, then we are forced to conclude that what is self-evident for finite sets is false for infinite sets. The set of all perfect squares is a proper subset of the set of all natural numbers. Is it really possible that they have the same number of elements nonetheless?

To answer this question we must examine more closely what we mean when we say that two sets have the same number of elements. When dealing with finite sets there is no difficulty. We can simply count the number of elements in each set and see if we get the same number both times. Alas, this is not practical for infinite sets. Indeed, being infinite means precisely that no matter what positive integer you choose, the set contains more elements than that number.

There is another way of determining when two sets have the same number of elements. We could attempt to establish a one-to-one correspondence between them. For example, if we have a certain number of empty desks in a classroom, and some number of students who need places to sit, and we want to know if we have as many desks as students, we could simply seat as many people as possible, If all of the desks are taken but some students are still standing, then we know we have more students than desks. If everyone is seated but some desks remain empty, then we know we have more desks than students. But if we run out of desks and students at the same time, then we know that the numbers of desks and students are the same.

In general, by a one-to-one correspondence between two sets A and B we mean a way of pairing elements of A with elements of B in such a way that every element of A is paired with exactly one element of B, and every element of B is paired with exactly one element of A. If there is a one-to-one correspondence between A and B, then we conclude that they have the same number of elements. We also say that A and B have the same **cardinality**.

The nice thing about this definition is that it works equally well for infinite sets. Even if A and B are infinite sets we can still inquire as to whether there is a one-to-one correspondence between them. Seen in this light, our earlier reasoning about the relative numbers of perfect squares and positive integers begins to make sense. Let A be the set of positive integers. Let B be the set of perfect squares. We define the following correspondence: Given any positive integer x, pair it with the perfect square x^2. In other words, we have the pairing

$$(1,1) \quad (2,4) \quad (3,9) \quad (4,16) \quad (5,25) \quad (6,36) \quad \cdots$$

Every positive integer appears as the first element of some pair, while every perfect square appears as a second element. Since every positive integer has exactly one square, and every perfect square has exactly one positive square root, we have our one-to-one correspondence. In this sense it is perfectly reasonable to say there are the same number of perfect squares as there are natural numbers.

Any set that can be put into one-to-one correspondence with the positive integers is said to be **countably infinite**. A set is said to be **countable** if it is either finite or countably infinite. Infinite sets that are not countable are said to be **uncountable**. Are there any uncountably infinite sets? For all we know right now, it might be the case that any two infinite sets can be placed into one-to-one correspondence.

As a candidate for a set that is not countable, let us consider the positive rational numbers. We notice that the positive integers are a subset of the set of rationals, but this tells us nothing about how large the positive rationals actually are. On the other hand, we know that the rational numbers are dense on the number line. By this I mean that in any interval on the line, no matter how small, you can always find infinitely many rational numbers residing inside the interval. This is in stark contrast to the natural numbers, which come at regularly spaced intervals. Does this imply that it is impossible to place the natural numbers into one-to-one correspondence with the positive rationals?

No, it does not. We will now show that the rational numbers are countably infinite. To do that, we must exhibit a one-to-one correspondence between the natural numbers and the rationals. This is accomplished by breaking up the set of all positive rational numbers into separate sets, depending on the sum of their numerator and denominator. Thus, first we list all the fractions whose top and bottom add up to two. Next to that we list all the fractions whose parts add up to three. For each possible sum, the individual fractions are listed from smallest numerator to largest numerator. Thus, we obtain the following list of positive rationals:

$$\frac{1}{1} \quad \| \quad \frac{1}{2}, \frac{2}{1} \quad \| \quad \frac{1}{3}, \frac{2}{2}, \frac{3}{1} \quad \| \quad \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1} \quad \| \quad \frac{1}{5}, \frac{2}{4}, \dots$$

It is clear that every rational number will appear somewhere in this list. For example, the fraction $\frac{17}{21}$ will be the seventeenth fraction listed among those whose top and bottom add up to 38. It is true that some fractions will be listed more than once, but this detail is not important.

Given our list of positive rationals, it is now easy to describe a pairing between them and the positive integers. We obtain the following correspondence:

$$\left(1, \frac{1}{1}\right), \quad \left(2, \frac{1}{2}\right), \quad \left(3, \frac{2}{1}\right), \quad \left(4, \frac{1}{3}\right), \dots$$

It follows that the rational numbers are countably infinite.

Incidentally, if you prefer a more explicit one-to-one correspondence, chew on this: Consider the fraction $\frac{i}{j}$. If $i + j$ is even pair this fraction with the integer $\dfrac{(i+j-1)(1+j-2)+2j}{2}$. If $i+j$ is odd then pair this fraction with the integer $\dfrac{(i+j-1)(i+j-2)+2i}{2}$. Once again we see that the rational numbers are countable.

We return to the original question: Do uncountable sets exist?

6.4 The Real Numbers are Uncountable

It turns out that it is flatly impossible to put the set of real numbers into one-to-one correspondence with the set of natural numbers. This fact was first proved by Cantor. His argument was so ingenious that it is now referred to simply as "Cantor's Diagonalization Argument."

Actually we will prove something even better. Instead of considering the set of all real numbers, let us focus our attention solely on those numbers that lie between zero and one. All such numbers can be represented by an infinite decimal expansion. The rational numbers lying between zero and one will have decimal expansions that either terminate (in which case they will have infinitely many zeros at the end of their expansion) or their expansions will repeat forever (for example, $\frac{1}{3} = .3333\cdots$). Irrational numbers have decimal expansions that never terminate and never repeat.

We will now proceed by contradiction. Suppose the real numbers can be placed in one-to-one correspondence with the natural numbers. In this case we would imagine that every real number is paired up with a particular natural number. Thus, we could talk about the first real number on the list, the second real number, the third, and so on. For this to be a one-to-one correspondence it must be the case that every real number gets paired up with some natural number. If we can show that every such list must, as a matter of logic, omit some real number between zero and one, then we will have shown that the real numbers can not be countable.

How are we to do that? The key realization is that two infinite decimal expansions represent different numbers precisely when they differ in at least one place. If x and y are the two decimal expansions, and the n-th digit of x is different from the n-th digit of y for even one value of n, then we know that $x \neq y$.

Let us return now to the list. It's first few entries will look something like this:

1	$.4756639067\cdots$
2	$.6836548963\cdots$
3	$.7957362204\cdots$
4	$.1029132537\cdots$
5	$.4568741001\cdots$
6	$.1112234323\cdots$

The particular decimals that I have listed here were chosen at random, simply to help me illustrate the argument. Any other decimals would work just as well. The point is that in our proposed one-to-one correspondence some real number must appear first, some other appears second, and so forth.

We will now show that there must be some real number between zero and one that is not on this list. To do that, I will show how to construct a specific example of such a number, which I will denote by x.

The first digit in x's expansion will be 5. Since the real number currently paired with one begins with 4, we can be sure that x is not the first number on the list.

The second digit in the expansion of x will be 9. This will guarantee that x is not the second number in the list, since the second digit in the expansion of *that* number is 8.

The third digit in x's expansion will be 4.

The fourth digit will be 0.

The fifth digit will be 8.

The sixth digit will be 4.

So x will begin $.594084\cdots$. Already we see that x can not possibly be equal to any of the first six numbers on the list. By continuing this construction, we can be certain that x will differ in at least one place from every decimal currently on our list. For example, perhaps you worry that in writing down x we have carelessly duplicated the 500th number on the list, which we will call y. We know that can not happen, however, because the 500th digit in the expansion of x will be different from the 500th digit of y.

The only possible conclusion is that x did not appear on the list. This shows that, in fact, our proposed listing was not a one-to-one correspondence after all. Therefore, the real numbers are not countable.

Even though the real numbers and the natural numbers are both infinite sets, we can now plausibly say that the set of real numbers is larger nonetheless.

6.5 The Power Set

We now have two sorts of infinities. There are countable infinities like the natural numbers and uncountable infinities like the real numbers. Is that all there is? In other words, is it necessarily the case that any infinite set can be put into one-to-one correspondence either with the natural numbers or with the real numbers? Or are there other infinities that can not be put into one-to-one correspondence with either of those sets?

It turns out that there are infinitely many kinds of infinite set. Just as we showed that the set of real numbers can plausibly be described as being larger than the set of natural numbers, we will now set about proving that there are infinite sets larger than the reals.

To do that, we need the concept of the power set. Given any set S (finite or infinite), we define its **power set** to be the set whose elements are all of the subsets of S. The power set of S is denoted by $\mathcal{P}(S)$.

For example, if $S = \{2\}$, then $\mathcal{P}(S) = \{\emptyset, \{2\}\}$. If $S = \{a, b\}$, then $\mathcal{P}(S) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$. If

$$S = \{\alpha, \beta, \gamma\}$$

then

$$\mathcal{P}(S) = \{\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\alpha, \beta\}, \{\alpha, \gamma\}, \{\beta, \gamma\}, \{\alpha, \beta, \gamma\}\}.$$

You might recall from chapter four that the number of subsets of a set with n elements is 2^n. It is easily shown that for $n > 1$, we have that $2^n > n$. This shows that for finite sets S, the power set $\mathcal{P}(S)$ is always larger.

If this holds true for infinite sets as well then we have a method for generating ever larger sorts of infinity. For example, suppose it turns out that the power set of an infinite set contains more elements than the set itself. Then it would have to be the case that the power set of the real numbers is a larger sort of infinity than the reals themselves. And the power set of the power set would be larger still. We could continue producing infinities forever.

Of course, that is assuming that what holds for finite sets also holds for infinite sets. As we have seen, that is not always a safe assumption. Happily, in this case it turns out to be a correct assumption, and that is what we shall prove in this section.

To do that we will mimic the proof we used in the last section. As before, we will proceed by contradiction. Let S be an infinite set. Let us suppose that it is possible to put the elements of S into one-to-one correspondence with the elements of $\mathcal{P}(S)$. Then we have paired up each element of S with one of the subsets of S in such a way that every subset appears somewhere on the list.

Now we must show that this is impossible. In other words, we must show that no matter how clever we have been in setting up our list, there must nonetheless be subsets that are not there. We will make use of an argument similar to Russell's paradox.

Let $x \in S$. Then in our listing x must be paired up with some element of $\mathcal{P}(S)$. Let us call that subset $f(x)$.

Now define the following subset of S:

$$A = \{x \in S \mid x \notin f(x)\}.$$

In other words, every element x gets paired with some subset $f(x)$. Thus, $f(x)$ is a subset of S. As such, it either contains x or does not contain x. In A we gather all those elements of S that get paired with subsets that do not contain themselves.

The set A is a perfectly legitimate subset of S. Therefore, it must appear somewhere on our list. Let us suppose that α is the lucky element of S that gets paired with A. Then we must have either $\alpha \in A$ or $\alpha \notin A$. We will show that both of these possibilities lead to contradictions.

Suppose that $\alpha \in A$. Then α is one of those elements of S that gets paired with a subset containing itself. But A contains only those sets that do not get paired with themselves. Since this is a contradiction, we must consider the alternative.

Perhaps you see where this is leading. Now suppose that $\alpha \notin A$. Then α is not one of those elements that gets paired with a subset containing itself. It therefore satisfies the requirement for being in A. This implies that $\alpha \in A$, which is another contradiction.

The only way out of this morass is to conclude that our listing is not a one-to-one correspondence after all.

Of course, since every element of S can be viewed as a one-element subset of S, there is no danger that $\mathcal{P}(S)$ is actually smaller than S itself. It follows that the power set of S is always larger than S itself. So the power set of the real numbers is strictly larger than the real numbers. The power set of that power set would be larger still. It follows that there is no largest infinity.

6.6 Yet Another Paradox

Before moving on we really must mention yet another problem with Cantor's definition of a set. As he defined things, it should be legitimate to discuss the set of all sets. Let us denote the set of all sets by S. Of course, S is infinite. We know from the previous section that $\mathcal{P}(S)$ must be strictly larger than S itself. But $\mathcal{P}(S)$ is itself a set. Therefore, it must be contained in S. It follows that $\mathcal{P}(S) \subseteq S$. But this would imply that $\mathcal{P}(S)$ is not larger than S, which is a contradiction. Ugh!

6.7 The Continuum Hypothesis

Our ruminations about the power set have shown that there is no largest infinity. Given any infinite set S, we know that $\mathcal{P}(S)$ is strictly larger than S. So far, however, we have only actually constructed two specific sorts of infinity. There are the natural numbers and the real numbers. As already mentioned, sets that can be put into one-to-one correspondence with the natural numbers are said to be countably infinite, while everything else is said to be uncountable.

You might be wondering whether there are infinite sets that are strictly larger than the natural numbers but strictly smaller than the real numbers. The assertion that there is no such infinity is known as **the continuum hypothesis**. Do not waste your time trying to prove this, however.

Recall that in the wake of Russell's paradox, it became necessary to develop a rigorous set of axioms for set theory. It was essential that crazy, self-referential gutter sets like the one in Russell's paradox be excluded from consideration. Sets that were too big, like the set of all sets, had to be excluded as well. Such a set of axioms was developed by two mathematicians named Zermelo and Fraenkel in the early part of the twentieth century. Thus, when we talk about proving the continuum hypothesis, our intention is to prove it with respect to these axioms. (There is one additional axiom, called the axiom of choice, that is sometimes added to the list compiled by Zermelo and Fraenkel, but that is a detail with which we will not concern ourselves).

It was subsequently shown by Kurt Gödel that no contradiction is obtained by assuming that the Zermelo-Fraenkel axioms and the continuum hypothesis are simultaneously true. However, this does not show that the continuum hypothesis is, in fact, true. Then, in 1963, Paul Cohen proved the reverse. In other words, he showed that no contradiction is obtained by assuming the Zermelo-Fraenkel axioms are true, while the continuum hypothesis is false. As a result, we conclude that the hypothesis is independent of the other axioms of set theory.

Since the proof techniques used by Gödel and Cohen are far more complex than anything we are considering in this book, we will instead consider a simpler question. We will prove that the power set of the natural numbers has the same size as the set of real numbers.

Doing that requires that we establish a one-to-one correspondence between the set of subsets of the natural numbers and the set of real numbers. Actually, we will only use the real numbers between zero and one for that purpose.

Begin by recalling that every real number can be expressed as an infinite decimal. Then recall that we could just as easily use a base-two system instead of a decimal system. In other words, in a base ten system every digit gets multiplied by some power of ten. Thus, the decimal .438 is the same as $(4 \times 10^{-1}) + (3 \times 10^{-2}) + (8 \times 10^{-3})$. But there is no reason why we could not use powers of two in place of powers of ten. In such a system, the number .1011 would be interpreted as $(1 \times 2^{-1}) + (0 \times 2^{-2}) + (1 \times 2^{-3}) + (1 \times 2^{-4})$.

If we use a binary, or base-two, system, then every real number between zero and one can be interpreted as an infinite decimal all of whose digits are zero and one. We will now show how to pair up every such decimal with a subset of the natural numbers.

The trick is easily done. Let us suppose that we have before us a particular decimal such as:

$$.0011101101000011\cdots$$

We will interpret this as a recipe for building a certain subset of the natural numbers. If there is a zero in the n-th decimal place of the number, then the natural number n will not be in our subset. If there is a one in the n-th place, then the natural number will be in our subset. Thus the numbers above would correspond to the subset

$$\{3, 4, 5, 7, 8, 10, 15, 16, \cdots\}.$$

Note that since zero is not a natural number, we begin our count with decimal place number one.

This correspondence is easily reversed. Given a subset of the natural numbers it is easy to generate the appropriate sequence of zeros and ones. It follows that there are as many subsets of the natural numbers as there are real numbers between zero and one.

6.8 Problems

1. Determine whether each of the following sets is finite, countably infinite or uncountable:

   (a) $\{x \in \mathbb{R} \mid 1 < x < 2\}$

   (b) $\{x \in \mathbb{Q} \mid 1 < x < 2\}$

   (c) $\{\frac{m}{n} \mid m, n \in \mathbb{N}, m < 100, 5 < n < 105\}$

   (d) $\{(a, b) \in \mathbb{Q} \times \mathbb{Q} \mid a + b = 1\}$.

2. Determine whether each of the following sets is countable or uncountable:

   (a) The integers greater than 10.

   (b) The odd, negative integers.

   (c) The real numbers between zero and two.

3. The next five problems involve Hilbert's Hotel. This is a hotel with a countably infinite number of rooms, numbered with the positive integers. Suppose that all of the rooms are filled. Then a guest walks into the lobby, asking for a room. Show how it is possible to accommodate him without forcing someone to leave the hotel.

4. Now suppose that x people show up at the hotel, where x is some positive integer. Is it still necessarily possible to accommodate them all?

5. What if a countably infinite number of people show up? Can all of these people be accommodated without forcing someone to leave the hotel?

6. Suppose that every night for a countably infinite number of nights, a countably infinite number of guests shows up looking for rooms. Does there ever come a point where we can no longer accommodate them?

7. Suppose that you have a countably infinite family of sets. Further suppose that each set contains a countably infinite number of elements. Use the result of the previous problem to show that the union of all the sets in the family has only a countably infinite number of elements. (This result is often expressed by saying, "A countable union of countable sets is countable.")

8. Prove that the set of all real numbers between zero and one whose decimal expansions use only 2's and 7's is uncountable.

9. Let S be the set of real numbers between zero and one whose decimal expansions consist only of 2's and 7's, with only finitely many 2's. What is the cardinality of S?

10. Prove that the set of perfect squares and the set of perfect cubes have the same cardinality.

11. On the real number line, prove that the interval $(0, 1)$ has the same cardinality as the interval $(0, 2)$.

12. Prove that if n and m are positive integers, then the set of multiples of n has the same cardinality as the multiples of m.

13. Let k be a positive integer. Prove that the set of positive integers has the same cardinality as the set $\{k, k+1, k+2, k+3, \ldots\}$.

14. Prove that if A and B are countable sets, then $A \times B$ is countable.

15. Prove that if A and B are countable sets then $A \cup B$ is countable as well.

16. We define the *difference* of two sets Y and Z to be $Y - Z = \{y \in Y | y \notin Z\}$. Use the previous exercise to show that if X is an uncountable set and A is a countable subset of X then $X - A$ is uncountable.

17. Let S_1 be a sphere of radius one and let S_2 be a sphere of radius 100. Prove that the set of points on the surface of S_1 has the same cardinality as the set of points on the surface of S_2. Hint: Imagine placing S_1 entirely inside S_2 so that the centers of the two spheres reside at the same point. Can you think of a method for pairing up the points on the surface of S_1 with the points on the surface of S_2?