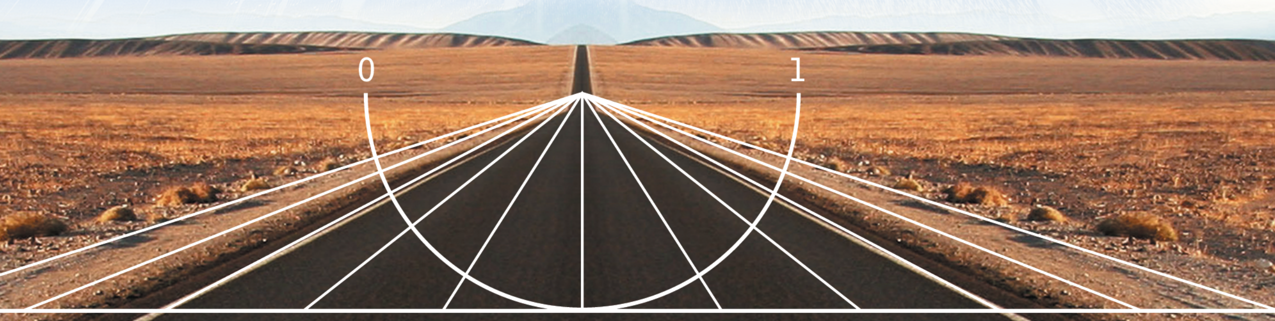
Roads to Infinity

The Mathematics of Truth and Proof



John Stillwell

Roads to Infinity

Roads to Infinity

The Mathematics of Truth and Proof

John Stillwell

A K Peters, Ltd.
Natick, Massachusetts

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20140703

International Standard Book Number-13: 978-1-4398-6550-7 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To Elaine

Contents

Preface

> ...it is hard to be finite upon an infinite subject, and all subjects are infinite.
>
> —Herman Melville (1850), p. 1170

Mathematics and science as we know them are very much the result of trying to grasp infinity with our finite minds. The German mathematician Richard Dedekind said as much in 1854 (see Ewald (1996), pp. 755–756):

> ...as the work of man, science is subject to his arbitrariness and to all the imperfections of his mental powers. There would essentially be no more science for a man gifted with an unbounded understanding—a man for whom the final conclusions, which we attain through a long chain of inferences, would be immediately evident truths.

Dedekind's words reflect the growing awareness of infinity in 19th-century mathematics, as reasoning about infinite processes (calculus) became an indispensable tool of science and engineering. He wrote at the dawn of an era in which infinity and logic were viewed as mathematical concepts for the first time. This led to advances (some of them due to Dedekind himself) that made all previous knowledge of these topics seem almost childishly simple.

Many popular books have been written on the advances in our understanding of infinity sparked by the set theory of Georg Cantor in the 1870s, and incompleteness theorems of Kurt Gödel in the 1930s. However, such books generally dwell on a single aspect of either set theory or logic. I believe it has not been made clear that the results of Cantor and Gödel form a seamless whole. The aim of this book is to explain the whole, in which set theory *interacts* with logic, and both begin to affect mainstream mathematics (the latter quite a recent development, not yet given much space in popular accounts).

In particular, I have taken some pains to tell the story of two neglected figures in the history of logic, Emil Post and Gerhard Gentzen. Post discovered incompleteness before Gödel, though he did not publish his proof until later. However, his proof makes clear (more so than Gödel's did) the origin of incompleteness in Cantor's set theory and its connections with the theory of computation. Gentzen, in the light of Gödel's theorem that the consistency of number theory depends on an assumption from outside number theory, found the *minimum* such assumption—one that grows out of Cantor's theory of ordinal numbers—paving the way for new insights into unprovability in number theory and combinatorics.

This book can be viewed as a sequel to my *Yearning for the Impossible*, the main message of which was that many parts of mathematics demand that we accept some form of infinity. The present book explores the consequences of accepting infinity, which are rich and surprising. There are many levels of infinity, ascending to heights that almost defy belief, yet even the highest levels have "observable" effects at the level of finite objects, such as the natural numbers 1, 2, 3, ... Thus, infinity may be more down-to-earth than certain parts of theoretical physics! In keeping with this claim, I have assumed very little beyond high school mathematics, except a willingness to grapple with alien ideas. If the notation of symbolic logic proves *too* alien, it is possible to skip the notation-heavy parts and still get the gist of the story.

I have tried to ease the reader into the technicalities of set theory and logic by tracing a single thread in each chapter, beginning with a natural mathematical question and following a sequence of historic responses to it. Each response typically leads to new questions, and from them new concepts and theorems emerge. At the end of the chapter there is a longish section called "Historical Background," which attempts to situate the thread in a bigger picture of mathematics and its history. My intention is to present key ideas in closeup first, then to revisit and reinforce them by showing a wider view. But this is not the only way to read the book. Some readers may be impatient to get to the core theorems, and will skip the historical background sections, at least at first reading. Others, in search of a big picture from the beginning, may begin by reading the historical background and then come back to fill in details.

The book has been developing in my unconscious since the 1960s, when I was an undergraduate at the University of Melbourne and a graduate student at MIT. Set theory and logic were then my passions in mathematics, but it took a long time for me to see them in perspective—I apologize to my teachers for the late return on their investment! I particularly want to thank Bruce Craven of Melbourne for indulging my interest in a field outside his specialty, and my MIT teachers Hartley Rogers Jr. and Gerald Sacks for widening my horizons in logic and set theory.

In recent times I have to thank Jeremy Avigad for bringing me up to date, and Cameron Freer for a very detailed criticism of an earlier draft of this book. John Dawson also made very helpful comments. If errors remain, they are entirely my fault. The University of San Francisco and

Monash University provided valuable support and facilities while I was writing and researching.

I also want to thank my friend Abe Shenitzer for his perennial help with proofreading, and my sons Michael and Robert for sharing this onerous task. Finally, I thank my wife Elaine, as always.

John Stillwell
University of San Francisco and Monash University
March 2010

- Chapter 1 -

The Diagonal Argument

Preview

Infinity is the lifeblood of mathematics, because there is no end to even the simplest mathematical objects—the positive integers 1, 2, 3, 4, 5, 6, 7, … . One of the oldest and best arguments about infinity is Euclid's proof that the prime numbers 2, 3, 5, 7, 11, 13, … form an infinite sequence. Euclid succeeds despite knowing virtually nothing about the sequence, by showing instead that any *finite* sequence of primes is incomplete. That is, he shows how to find a prime p different from any given primes $p_1, p_2, \ldots, p_n$.

A set like the prime numbers is called *countably infinite* because we can order its members in a list with a first member, second member, third member, and so on. As Euclid showed, the list is infinite, but each member appears at some finite position, and hence gets "counted."

Countably infinite sets have always been with us, and indeed it is hard to grasp infinity in any way other than by counting. But in 1874 the German mathematician Georg Cantor showed that infinity is more complicated than previously thought, by showing that the set of real numbers is *uncountable*. He did this in a way reminiscent of Euclid's proof, but one level higher, by showing that any countably infinite list of real numbers is incomplete.

Cantor's method finds a real number x different from any on a given countable list $x_1, x_2, x_3, \ldots$ by what is now called the *diagonal argument*, for reasons that will become clear below. The diagonal argument (which comes in several variations) is logically the simplest way to prove the existence of uncountable sets. It is the first "road to infinity" of our title, so we devote this chapter to it. A second road—via the *ordinals*—was also discovered by Cantor, and it will be discussed in Chapter 2.

1.1 Counting and Countability

> If I should ask further how many squares there are, one might reply truly that there are as many as the corresponding number of roots, since every square has its own root and every root has its own square, while no square has more than one root and no root more than one square.
>
> —Galileo Galilei,
> *Dialogues Concerning the Two New Sciences*, First day.

The process of counting 1, 2, 3, 4, ... is the simplest and clearest example of an *infinite process*. We know that counting never ends, because there is no last number, and indeed one's first thought is that "infinite" and "neverending" mean the same thing. Yet, in a sense, the endless counting process *exhausts* the set $\{1, 2, 3, 4, \ldots\}$ of positive integers, because each positive integer is eventually reached. This distinguishes the set of positive integers from other sets—such as the set of points on a line—which seemingly cannot be "exhausted by counting." Thus it may be enlightening to dwell a little longer on the process of counting, and to survey some of the infinite sets that can be exhausted by counting their members.

First, what do we mean by "counting" a set of objects? "Counting" objects is the same as arranging them in a (possibly infinite) *list*—first object, second object, third object, and so on—so that each object in the given set appears on the list, necessarily at some positive integer position. For example, if we "count" the squares by listing them in increasing order,

$$1,\ 4,\ 9,\ 16,\ 25,\ 36,\ 49,\ 64,\ 81,\ 100,\ 121,\ 144,\ 169,\ 196,\ 225,\ \ldots,$$

then the square 900 appears at position 30 on the list. Listing a set is mathematically the same as assigning the positive integers in some way to its members, but it is often easier to visualize the list than to work out the exact integer assigned to each member.

One of the first interesting things to be noticed about infinite sets is that *counting a part may be "just as infinite" as counting the whole*. For example, the set of *positive even numbers* 2, 4, 6, 8, ... is just a part of the set of positive integers. But the positive even numbers (in increasing order) form a list that matches the list of positive integers completely, item by item. Here they are:

$$\begin{array}{cccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & \ldots \\ 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20 & 22 & 24 & 26 & \ldots \end{array}$$

Thus listing the positive even numbers is a process completely parallel to the process of listing the positive integers. The reason lies in the item-by-item matching of the two lists, which we call a *one-to-one correspondence*. The function $f(n) = 2n$ encapsulates this correspondence, because it matches each positive integer n with exactly one positive even number $2n$, and each positive even number $2n$ is matched with exactly one positive integer n.

So, to echo the example of Galileo quoted at the beginning of this section: if I should ask how many even numbers there are, one might reply truly that there are as many as the corresponding positive integers. In both Galileo's example, and my more simple-minded one, one sees a one-to-one correspondence between the set of positive integers and a part of itself. This unsettling property is the first characteristic of the world of infinite sets.

COUNTABLY INFINITE SETS

A set whose members can be put in an infinite list—that is, in one-to-one correspondence with the positive integers—is called *countably infinite*. This common property of countably infinite sets was called their *cardinality* by Georg Cantor, who initiated the general study of sets in the 1870s. In the case of finite sets, two sets have the same cardinality if and only if they have the same number of elements. So the concept of cardinality is essentially the *same* as the concept of number for finite sets.

For countably infinite sets, the common cardinality can also be regarded as the "number" of elements. This "number" was called a *transfinite number* and denoted $\aleph_0$ ("aleph zero" or "aleph nought") by Cantor. One can say, for instance, that there are $\aleph_0$ positive integers. However, one has to bear in mind that $\aleph_0$ is more elastic than an ordinary number. The sets $\{1, 2, 3, 4, \ldots\}$ and $\{2, 4, 6, 8, \ldots\}$ both have cardinality $\aleph_0$, even though the second set is a strict subset of the first. So one can also say that there are $\aleph_0$ even numbers.

Moreover, the cardinality $\aleph_0$ stretches to cover sets that at first glance seem much larger than the set $\{1, 2, 3, 4, \ldots\}$. Consider the set of dots shown in Figure 1.1. The grid has infinitely many infinite rows of dots, but nevertheless we can pair each dot with a different positive integer as shown in the figure. Simply view the dots along a series of finite diagonal lines, and "count" along the successive diagonals, starting in the bottom left corner.

There is a very similar proof that the set of (positive) fractions is countable, since each fraction m/n corresponds to the pair (m, n) of positive integers. It follows that the set of positive rational numbers is countable, since each positive rational number is given by a fraction. Admittedly, there are many fractions for the same number—for example the number $1/2$ is also given by the fractions $2/4$, $3/6$, $4/8$, and so on—-but we can

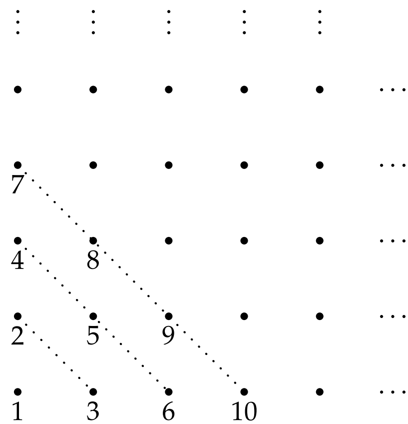


Figure 1.1. Counting the dots in an infinite grid.

list the positive rational numbers by going through the list of fractions and omitting all fractions that represent previous numbers on the list.

1.2 Does One Infinite Size Fit All?

A nice way to illustrate the elasticity of the cardinality $\aleph_0$ was introduced by the physicist George Gamow (1947) in his book *One, Two, Three, …, Infinity*. Gamow imagines a hotel, called *Hilbert's hotel*, in which there are infinitely many rooms, numbered 1, 2, 3, 4, … . Listing the members of an infinite set is the same as accommodating the members as "guests" in Hilbert's hotel, one to each room.

The positive integers can naturally be accommodated by putting each number n in room n (Figure 1.2):

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

Figure 1.2. Standard occupancy of Hilbert's hotel.

The $\aleph_0$ positive integers fill every room in Hilbert's hotel, so we might say that $\aleph_0$ is the "size" of Hilbert's hotel, and that occupancy by more than $\aleph_0$ persons is unlawful. Nevertheless there is room for one more (say, the number 0). Each guest simply needs to move up one room, leaving the first room free (Figure 1.3):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	$\cdots$

Figure 1.3. Making room for one more.

Thus $\aleph_0$ can always stretch to include one more: in symbols, $\aleph_0 + 1 = \aleph_0$. In fact, there is room for another countable infinity of "guests" (say, the negative integers $-1, -2, -3, \ldots$). The guest in room n can move to room $2n$, leaving all the odd numbered rooms free (Figure 1.4):

	1		2		3		4		5		6		7		8		9		10	$\cdots$

Figure 1.4. Making room for a countable infinity more.

In symbols: $\aleph_0 + \aleph_0 = \aleph_0$.

There is even room for a countable infinity of countable infinities of guests. Suppose, say, that the guests arrive on infinite buses numbered 1, 2, 3, 4, ..., and that each bus has guests numbered 1, 2, 3, 4 , ... The guests in bus 1 can be accommodated as follows:

put guest 1 in room 1; then skip 1 room; that is,
put guest 2 in room 3; then skip 2 rooms; that is,
put guest 3 in room 6; then skip 3 rooms; that is,
put guest 4 in room 10; then skip 4 rooms; ... .

Thus the first bus fills the rooms shown in Figure 1.5:

1		2			3				4					5						$\cdots$

Figure 1.5. Making room for a countable infinity of countable infinities.

After the first bus has been unloaded, the unoccupied rooms are in blocks of 1, 2, 3, 4, ... rooms, so we can unload the second bus by putting its guests in the leftmost room of each block. After that, the unoccupied rooms are *again* in blocks of 1, 2, 3, 4, ... rooms, so we can repeat the process with the third bus, and so on. (You may notice that each busload occupies a sequence of rooms numbered the same as a row in Figure 1.1.)

The result is that the whole series of $\aleph_0$ busloads, each with $\aleph_0$ guests, can be packed into Hilbert's hotel—with exactly one guest per room. In symbols: $\aleph_0 \times \aleph_0 = \aleph_0$.

The equations of "cardinal arithmetic" just obtained,

$$\aleph_0 + 1 = \aleph_0,$$
$$\aleph_0 + \aleph_0 = \aleph_0,$$
$$\aleph_0 \times \aleph_0 = \aleph_0,$$

show just how elastic the transfinite number $\aleph_0$ is. So much so, one begins to suspect that cardinal arithmetic has nothing to say except that any infinite set has cardinality $\aleph_0$. And if all transfinite numbers are the same it is surely a waste of time to talk about them. But fortunately they are *not* all the same. In particular, the set of points on the line has cardinality strictly *larger* than $\aleph_0$. Cantor discovered this difference in 1874, opening a crack in the world of the infinite from which unexpected consequences have spilled ever since. There is, after all, a lot to say about infinity, and the purpose of this book is to explain why.

1.3 Cantor's Diagonal Argument

Before studying the set of points on the line, we look at a related set that is slightly easier to handle: the set of all sets of positive integers. A set S of positive integers can be described by an infinite sequence of 0s and 1s, with 1 in the nth place just in case n is a member of S. Table 1.1 shows a few examples:

subset	1	2	3	4	5	6	7	8	9	10	11	...
even numbers	0	1	0	1	0	1	0	1	0	1	0	...
squares	1	0	0	1	0	0	0	0	1	0	0	...
primes	0	1	1	0	1	0	1	0	0	0	1	...

Table 1.1. Descriptions of positive integer sets.

Now suppose that we have $\aleph_0$ sets of positive integers. That means we can form a list of the sets, $S_1, S_2, S_3, \ldots$, whose nth member S_n is the set paired with integer n. We show that such a list can never include *all* sets of positive integers by describing a set S different from each of $S_1, S_2, S_3, \ldots$.

This is easy: for each number n, put n in S just in case n is *not* in S_n. It follows that *S differs from each S_n with respect to the number n: if n is S_n, then n is not in S; if n is not S_n, then n is in S.* Thus S is not on the list $S_1, S_2, S_3, \ldots$, and hence no such list can include all sets of positive integers.

subset	1	2	3	4	5	6	7	8	9	10	11	…
S_1	**0**	1	0	1	0	1	0	1	0	1	0	…
S_2	1	**0**	0	1	0	0	0	0	1	0	0	…
S_3	0	1	**1**	0	1	0	1	0	0	0	1	…
S_4	1	0	1	**0**	1	0	1	0	1	0	1	…
S_5	0	0	1	0	**0**	1	0	0	1	0	0	…
S_6	1	1	0	1	1	**0**	1	1	0	1	1	…
S_7	1	1	1	1	1	1	**1**	1	1	1	1	…
S_8	0	0	0	0	0	0	0	**0**	0	0	0	…
S_9	0	0	0	0	0	0	0	0	**1**	0	0	…
S_{10}	1	0	0	1	0	0	1	0	0	**1**	0	…
S_{11}	0	1	0	0	1	0	0	1	0	0	**0**	…
$\vdots$												
S	**1**	**1**	**0**	**1**	**1**	**1**	**0**	**1**	**0**	**0**	**1**	…

Table 1.2. The diagonal argument.

The argument we have just made is called a *diagonal* argument because it can be presented visually as follows. Imagine an infinite table whose rows encode the sets $S_1, S_2, S_3, \ldots$ as sequences of 0s and 1s, as in the examples above. We might have, say, the sets shown in Table 1.2.

The digit (1 or 0) that signals whether or not n belongs to S_n is set in bold type, giving a diagonal sequence of bold digits

$$\mathbf{00100010110}\ldots.$$

The sequence for S is obtained by switching each digit in the diagonal sequence. Hence the sequence for S is necessarily different from the sequences for all of $S_1, S_2, S_3, \ldots$.

The cardinality of the set of all sequences of 0s and 1s is called $2^{\aleph_0}$. We use this symbol because there are two possibilities for the first digit in the sequence, two possibilities for the second digit, two possibilities for the third, and so on, for all the $\aleph_0$ digits in the sequence. Thus it is reasonable to say that there are $2 \times 2 \times 2 \times \cdots$ ($\aleph_0$ factors) possible sequences of 0s and 1s, and hence there are $2^{\aleph_0}$ sets of positive natural numbers.

The diagonal argument shows that $2^{\aleph_0}$ is strictly greater than $\aleph_0$ because there is a one-to-one correspondence between the positive integers and certain sets of positive integers, but not with *all* such sets. As we have just seen, if the numbers $1, 2, 3, 4, \ldots$ are assigned to sets $S_1, S_2, S_3, S_4, \ldots$ there will always be a set (such as S) that fails to be assigned a number.

THE LOGIC OF THE DIAGONAL ARGUMENT

> Many mathematicians aggressively maintain that there can be no doubt of the validity of this proof, whereas others do not admit it. I personally cannot see an iota of appeal in this proof ... my mind will not do the things that it is obviously expected to do if this is indeed a proof.
>
> —P .W. Bridgman (1955), p. 101

P. W. Bridgman was an experimental physicist at Harvard, and winner of the Nobel prize for physics in 1946. He was also, in all probability, one of the smartest people *not* to understand the diagonal argument. If you had any trouble with the argument above, you can rest assured that a Nobel prize winner was equally troubled. On the other hand, I do not think that any mathematically experienced reader *should* have trouble with the diagonal argument. Here is why.

The logic of the diagonal argument is really very similar to that of Euclid's proof that there are infinitely many primes. Euclid faced the difficulty that the totality of primes is hard to comprehend, since they follow no apparent pattern. So, he avoided even considering the totality of primes by arguing instead that *any finite list of primes is incomplete*.

Given a finite list of primes $p_1, p_2, \dots, p_n$, one forms the number

$$N = p_1 p_2 \cdots p_n + 1,$$

which is obviously not divisible by any one of $p_1, p_2, \dots, p_n$ (they each leave remainder 1). But N is divisible by *some* prime number, so the list $p_1, p_2, \dots, p_n$ of primes is incomplete. Moreover, we can find a specific prime p not on the list by finding the smallest number ≥ 2 that divides N.

An uncountable set is likewise very hard to comprehend, so we avoid doing so and instead suppose that we are given a countable list $S_1, S_2, S_3, \dots$ of members of the set. The word "given" may be interpreted as strictly as you like. For example, if $S_1, S_2, S_3, \dots$ are sequences of 0s and 1s, you may demand a *rule* that gives the mth digit of S_n at stage $m + n$. The diagonal argument still works, and it gives a *completely specific* S not on the given list. (Indeed, it also leads to some interesting conclusions about rules for computing sequences, as we will see in Chapter 3.)

THE SET OF POINTS ON THE LINE

> The goal of set theory is to answer the question of highest importance: whether one can view the line in an atomistic manner, as a set of points.
>
> —Nikolai Luzin (1930), p. 2.

By the "line" we mean the number line, whose "points" are known as the *real numbers*. Each real number has a *decimal expansion* with an infinite