Making up Numbers
A History of Invention in Mathematics

Ekkehard Kopp

Making up Numbers offers a detailed but accessible account of a wide range of mathematical ideas. Starting with elementary concepts, it leads the reader towards aspects of current mathematical research. Ekkehard Kopp adopts a chronological framework to demonstrate that changes in our understanding of numbers have often relied on the breaking of long-held conventions, making way for new inventions that provide greater clarity and widen mathematical horizons. Viewed from this historical perspective, mathematical abstraction emerges as neither mysterious nor immutable, but as a contingent, developing human activity.

Chapters are organised thematically to cover: writing and solving equations, geometric construction, coordinates and complex numbers, attitudes to the use of 'infinity' in mathematics, number systems, and evolving views of the role of axioms. The narrative moves from Pythagorean insistence on positive multiples to gradual acceptance of negative, irrational and complex numbers as essential tools in quantitative analysis.

Making up Numbers will be of great interest to undergraduate and A-level students of mathematics, as well as secondary school teachers of the subject. By virtue of its detailed treatment of mathematical ideas, it will be of value to anyone seeking to learn more about the development of the subject.

As with all Open Book publications, this entire book is available to read for free on the publisher's website. Printed and digital editions, together with supplementary digital material, can also be found at www.openbookpublishers.com

Cover images from Wikimedia Commons. For image details see captions in the book.

Cover Design by Anna Gatti.
CHAPTER 9

Counting beyond the finite

Let every student of nature take this as his rule, that whatever the mind seizes upon with particular satisfaction is to be held in suspicion.

Sir Francis Bacon, Novum Organum, 1620

Summary

In this chapter we begin with Georg Cantor’s work on the continuum, which reflects the abstract approach he and Richard Dedekind had shared in their models for the real number system. From the 1870s onward, their work was to have a profound influence on the development of mathematics. Within a decade, further investigation into the nature of the continuum would lead Cantor to focus on the nature of infinite sets, which sparked deep philosophical disagreements between leading groups of mathematicians about the nature of their subject. This would culminate in a profound conceptual revolution in prevailing views of the nature of mathematical truth.

Cantor’s perception of the continuum was to lead him to explore a general notion of ‘size’ for sets, prompted by ‘different kinds of infinity’ apparently represented by \( \mathbb{Q} \) and \( \mathbb{R} \). His investigation of trigonometric series, on the other hand, stimulated his development of the far-reaching concepts of transfinite ordinal and cardinal numbers as an abstract method of continuing the process of ‘counting’ beyond finite sets. His groundbreaking papers in both these areas laid the groundwork for an entirely new theory of sets, providing a basis for the whole of mathematics, while at the same time foreshadowing troubling paradoxes that would come to plague this new theory.

1. Cantor’s continuum

Georg Cantor’s initial motivation for investigating the continuum had nothing to do with concerns about the teaching of Calculus to students. Instead, it was his analysis of Fourier series that led him to his model for the real number system.
Given an infinite ‘point set’ \( P \) on the line (i.e. a subset of \( \mathbb{R} \)), he defined its derived set \( P' \) as the set of all its ‘limit points’.\(^1\) The point \( x \) belongs to \( P' \) precisely when infinitely many points of \( P \) ‘lie within any neighbourhood, however small’ of \( x \). This procedure can be iterated indefinitely, which led Cantor to his key result on Fourier series representations by defining two mutually exclusive ‘species’ of point sets. If, after \( n \) repetitions, the \( n^{th} \) derived set \( P^{(n)} \) is finite, then clearly its derived set \( P^{(n+1)} \) is empty – he would then call \( P \) a set of the \( n^{th} \) kind. A set \( P \) is of the ‘first species’ if it is a set of the \( n^{th} \) kind for some \( n \geq 1 \). Subsets of \( \mathbb{R} \) whose derived sets \( P^{(n)} \) were all infinite were placed in the second species.

This gave him the uniqueness criterion he was looking for: he showed that a real function \( f \) is represented uniquely by its Fourier series whenever, within any interval of length \( 2\pi \), the set of exceptional points (where the series fails to converge or the representation fails) is a set of the first species.

But his investigations had now led him into quite different territory: Although the set of exceptional points is typically infinite, one might expect it to be ‘small’ compared to the set of points within the interval where the Fourier representation of \( f \) is unique. Therefore, in what sense could one distinguish between different ‘sizes’ of infinite sets? How might one extend the notion of counting to such sets? These questions about the nature of the continuum lay at the root of Cantor’s extensive exploration of infinite sets.

### 1.1. Countably infinite subsets of \( \mathbb{R} \).

To pin down ‘how many’ elements a given set has, the natural first question concerns counting:

**How should we count the ‘number of elements’ of a set?**

Intuitively, a set \( A \) is finite precisely when we can ‘count off its members one by one’ in a list with a beginning and an end. In other words, there should be some natural number \( n \) such that we can ‘pair off’ all the elements of \( A \) one by one with the numbers \( 1, 2, 3, \ldots, n \). We would then say, quite naturally, that \( A \) has \( n \) elements, and write the list of its members as a finite sequence: \( a_1, a_2, a_3, \ldots, a_n \).

Such a pairing is a one-one correspondence between the sets \( \{1, 2, 3, \ldots, n\} \) and \( A \).

In this way, \( n \) serves to tell us ‘how many’ elements the set \( A \) has. Cantor expressed this as the power (‘Mächtigkeit’) of the set \( A \). Today we say that \( A \) has cardinality \( n \), and write this as \( |A| = n \). It is clear that no such finite pairing can exist for \( \mathbb{N} \), which is therefore not a finite set.

Bolzano had made similar observations some 40 years before Cantor, and gave examples of infinite sets which he could put into one-one correspondence with proper subsets of themselves. But his notes came to light

---

\(^1\)Now known as accumulation points. This notion was not new: it is implicit in Bolzano’s version of the Bolzano-Weierstrass theorem, discussed in Chapter 6.
only much later. As Dedekind, independently, had also realised, the concept of one-one correspondences between sets leads naturally to a definition of infinite sets:

An infinite set is one that can be put into a one-one correspondence with a proper subset of itself.

Obvious examples are the set $\mathbb{N}$ of all natural numbers and the set $\mathbb{E}$ of all even numbers (where $n \in \mathbb{N}$ corresponds uniquely to $2n \in \mathbb{E}$); or the set of all perfect squares (where $n$ corresponds to $n^2$); or the set of all prime numbers (although, in this example we can’t locate the $n^{th}$ prime number precisely when $n$ is large – as we saw in Chapter 7.)

Cantor observed that with one-one correspondences he could extend the notion of ‘counting’ to any set $B$ that can be written as an infinite sequence $b_1, b_2, b_3, ..., b_n, ...$. Listing its elements as a sequence establishes a one-one correspondence (or bijection) between the set $B$ and the set $\mathbb{N}$ of all natural numbers: for each $n = 1, 2, 3, ..., $ associating the element $b_n$ of $B$ with the natural number $n$. Thus: we will call a set $B$ countably infinite if it can be put in one-one correspondence with all of $\mathbb{N}$.

The term denumerable is often used instead, while a set is usually called countable if it is either finite or countably infinite. Cantor introduced the term ‘countable’ in 1883. Dedekind had called such sets simply infinite.

Cantor’s initial interest was to examine familiar subsets of the continuum to decide whether they are countably infinite. A striking example is the use of his first diagonal method to show that $\mathbb{Q}$ is countably infinite.

Restricting to positive rationals ($\mathbb{Q}^+$), we illustrate how $\mathbb{Q}^+$ can be written as a sequence in the following diagram, where we imagine an infinite square array containing all positive fractions:

```
1 1 → 2 1 → 3 1 → 4 1 → 5 1 → n 1 ...
↓ ↑ ↓ ↑ ↓ ↑ ...
1/2 2 2 → 3 2 → 4 2 → 5 2 → n 2 ...
↓ ↑ ↓ ↑ ↓ ↑ ...
1/3 2 3 → 3 3 → 4 3 → 5 3 → n 3 ...
↓ ↑ ↓ ↑ ↓ ↑ ...
1/4 2 4 → 3 4 → 4 4 → 5 4 → n 4 ...
↓ ↑ ↓ ↑ ↓ ↑ ...
1/5 2 5 → 3 5 → 4 5 → 5 5 → n 5 ...
↓ ↑ ↓ ↑ ↓ ↑ ...
... ... ... ... ... ...
1/n 2/n → 3/n → 4/n → 5/n → n/n ...
↓ ↑ ↓ ↑ ↓ ↑ ...
```

The arrows show how the distinct rational numbers in this array can be written as a sequence. From $1 = \frac{1}{1}$ move right to $2 = \frac{2}{1}$, then diagonally down left to $\frac{1}{2}$, down to $\frac{1}{3}$, then diagonally up right, (skipping $\frac{2}{2} = 1$)
to $3 = \frac{3}{1}$, then to $4 = \frac{4}{1}$, diagonally down left to $\frac{3}{2}, \frac{2}{1}, \frac{1}{4}$, down to $\frac{1}{5}$, then diagonally up right (skipping $\frac{2}{4}, \frac{3}{3}, \frac{4}{2}$) to $5 = \frac{5}{1}$, then right to $6 = \frac{6}{1}$, diagonally left again, and so on, zig-zagging through the whole array. Every ratio whose numerator and denominator have common factors is skipped (recall that a rational number is an equivalence class of fractions). For example, $\frac{1}{5}$ becomes the tenth term in our sequence (as we skipped $\frac{3}{2}$), while $\frac{5}{1}$ is the eleventh, since $\frac{2}{4}, \frac{3}{3}, \frac{4}{2}$ are skipped. This ensures that only distinct positive rational numbers are counted. We map this sequence of distinct members of $\mathbb{Q}^+$ one-to-one to the even numbers $2, 4, 6, ..., 2n, ...$. Similarly, the set $\mathbb{Q}^- = \{ -r : r \in \mathbb{Q}^+ \}$ of negative rationals is mapped one–to-one to the odd numbers $3, 5, ..., 2n + 1, ...$, and $0$ is mapped to $1$. Taken together, these provide a one-one correspondence between $\mathbb{Q}$ and $\mathbb{N}$.

A slight extension of the above argument shows that any countable union of countably infinite sets is countably infinite. Write this union as a sequence of sets, then write each set in the union as a sequence, list them below each other as in our array, and move through the array in the zig-zig manner indicated, skipping all elements encountered at an earlier stage. This process rearranges the elements and displays the countable union of sequences as a single sequence, in one-one correspondence with $\mathbb{N}$. Thus: an infinite sequence of infinite sequences can be re-arranged into a single infinite sequence.

Cantor continued his explorations of subsets of $\mathbb{R}$ by showing that the set $\mathbb{A}$ of all real algebraic numbers (see Chapter 8) also countably infinite.

Recall that we may take $\mathbb{A}$ as the set of all real roots of polynomials with integer coefficients, that is, solutions of equations of the form

$$c_m x^m + c_{m-1} x^{m-1} + ... + c_1 x + c_0 = 0,$$

where $c_i \in \mathbb{Z}$ ($i \leq m$) and $m \in \mathbb{N}$. Recalling that $|c_i|$ denotes the modulus of the integer $c_i$, define the height of $x$ as $h = m + |c_m| + |c_{m-1}| + ... + |c_0|$.

For any given $h$ there are only finitely many polynomials with height $h$, since clearly this would require $m$ and each $|c_k|$ to be at most $h$. Any polynomial of degree $m$ with integer coefficients has at most $m$ distinct real roots, so it contributes at most $m$ distinct algebraic numbers, since its factorisation includes the product of $k \leq m$ linear factors. The factors $(x - \alpha_1)(x - \alpha_2)...(x - \alpha_k)$ yield real roots $\alpha_1, ..., \alpha_k$. So any polynomial of height $h$ will contribute only finitely many algebraic numbers. Since all the real algebraic numbers are found as roots of such a polynomial for some height $h$, we can now write them down as a sequence, beginning with height $h$.

---

2 As noted in Chapter 4, in some cases we obtain quadratic factors that cannot be factorised further if we allow only real roots, such as $x^2 + 1 = 0$. These quadratic factors produce conjugate pairs of complex roots instead. In such cases $k = m - 2n$ for some $n \geq 1$. 

---
2 (there are none of height 1). Since each height contributes only finitely many new numbers, and all real algebraic numbers appear in the sequence, the set $A$ of all real algebraic numbers is countably infinite.

1.2. Uncountable subsets of $\mathbb{R}$. Despite his success in defining the real number system, a natural question worried Cantor: since the rationals are dense in the reals, so that between any two real numbers at least one rational number (indeed, infinitely many) can be found, how might one characterise the apparent difference in ‘size’ (or plurality) between these two sets? Since $\mathbb{Q}$ has gaps while $\mathbb{R}$ does not, intuitively there appear to be many more reals than rationals. On the other hand, the sets $\mathbb{Q}$ and $\mathbb{N}$ are in one-one correspondence, yet there are also many more rationals than natural numbers.

In November 1873 Cantor wrote to Dedekind, asking whether, in his view, one-one correspondences could be found between $\mathbb{R}$ and $\mathbb{N}$. Dedekind replied that he could offer no evidence that such a correspondence would be impossible. But Cantor soon solved the problem in dramatic fashion in a paper that appeared in Crelle’s Journal in 1874—this paper in effect launched set theory as a new subject. In this paper he proved that there can be no one-one correspondence between $\mathbb{N}$ and $\mathbb{R}$. (An infinite set would later be called uncountable if it could not be put into one-one correspondence with $\mathbb{N}$.)

Cantor’s proof (which we summarise below) was criticised as less than convincing by some of his peers. Undaunted, he boldly emphasised the significance of his result as follows: ‘Thus I have found the clear difference between a so-called continuum and a set of the nature of the entirety of the algebraic numbers.’

Rather than reproduce Cantor’s original proof, we consider a slight reformulation of his argument. Given an arbitrary sequence $(x_n)_{n \geq 1}$ of real numbers, construct nested closed intervals $([a_n, b_n]_{n \geq 1})$—that is, for each $i$, $[a_{i+1}, b_{i+1}]$ is a closed subinterval of $[a_i, b_i]$—each chosen to ensure that, for each $n \geq 1$, $x_n$ is not in $[a_n, b_n]$. Thus $x_1 \notin [a_1, b_1], x_2 \notin [a_2, b_2], ..., x_n \notin [a_n, b_n], ...$. For each $n$, the interval $[a_n, b_n]$ has been constructed to avoid the first $n$ points of our given sequence. By one version of the completeness property of $\mathbb{R}$ (see Footnote 9 in Chapter 8) the sequence of nested closed intervals has non-empty intersection $I = \cap_{n \geq 1} [a_n, b_n]$. No $x_n$ in our original sequence can belong to $I$, hence any member of $I$ is a real number not in our sequence, so that $\mathbb{R}$ cannot be countable.

A different proof, outlined by Cantor in 1891, is more common in textbooks today. It rests on his second diagonal argument: for simplicity, restrict attention to the infinite decimal expansions whose integral part is 0. The

---

[^3]: For $h = 2$ we have $x = 0$ or $2 = 0$; the latter equation is false, so 0 is the only algebraic number for height 2. You may check that height 3 yields just 1 and $-1$; height 4 yields $-2, -\frac{1}{2}, \frac{1}{2}, 2$; and that height 5 provides $\sqrt{2}$ as one of the first irrationals in this sequence.

proof is by contradiction: if these expansions could all be written down in a sequence, they would produce a doubly infinite array of the form

\[
\begin{align*}
\alpha_1 &= 0.a_{11}a_{12}a_{13}a_{14}a_{15}... \\
\alpha_2 &= 0.a_{21}a_{22}a_{23}a_{24}a_{25}... \\
\alpha_3 &= 0.a_{31}a_{32}a_{33}a_{34}a_{35}... \\
\alpha_4 &= 0.a_{41}a_{42}a_{43}a_{44}a_{45}... \\
\alpha_5 &= 0.a_{51}a_{52}a_{53}a_{54}a_{55}... \\
&\cdots\cdots\cdots
\end{align*}
\]

where all the \( a_{ij} \) are chosen from \( \{0, 1, 2, \ldots, 9\} \), and the sequence \( (\alpha_n)_{n \geq 1} \) would contain all infinite decimal expansions whose integral part is 0. To avoid duplication in the list we will write all terminating decimals in their ‘recurring nines’ form. Now construct another infinite decimal expansion \( \beta = 0.b_1b_2b_3... \), where, for each \( i \geq 1 \), the digit \( b_i \) is chosen to be different from \( a_{ii} \), and the digits 0 and 9 are not used. For each \( i \) this leaves seven alternative choices. This new expansion is different from each of the expansions in the sequence \( (\alpha_n)_{n \geq 1} \), since it differs from \( a_1 \) in the first digit, from \( a_2 \) in the second, and so on. None of the \( b_i \) are 0, so \( \beta \) cannot be a terminating expansion that coincides with a ‘recurring nines’ entry \( \alpha_n \) in the above list. But that means that the list does not contain all infinite decimal expansions, so our assumption that the set is countably infinite has led to a contradiction.

Hence the interval \([0, 1]\), and thus also the set \( \mathbb{R} \) of all (non-terminating) infinite decimal expansions, cannot be placed in one-one correspondence with \( \mathbb{N} \).

On the other hand, \( \mathbb{R} = \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q}) \) consists of all rationals together with all irrationals. The union of two countably infinite sets is countably infinite, and \( \mathbb{Q} \) is countably infinite. Therefore, the set of all irrationals is uncountable.

As described above, Cantor’s paper also verified that the set \( \mathbb{A} \) of all real algebraic numbers is countably infinite. Together with all transcendental numbers it again makes up the (uncountable) set \( \mathbb{R} \), so the set of transcendental numbers must be uncountable. In fact, as Cantor observes, his argument proves Liouville’s claim that every interval contains infinitely many transcendental numbers.

The fact that there are ‘many more’ real numbers that behave like \( \pi, e \) or Liouville’s constant \( L \), may be somewhat disconcerting. We have to accept that we have no way of ‘knowing’ (individually) most of the numbers that we represent by infinite decimal expansions. In fact, we cannot even define most of these decimal expansions in a meaningful way in finitely many words. Using a finite alphabet, there are only countably many possible sentences that we can form to articulate the definition of any particular number in words (or symbols).
These facts about the ‘familiar’ continuum help to explain why Cantor’s paper met with a hostile reception in some quarters, most notably from his former mentor Kronecker, who was to become a bitter enemy. Kronecker seems to have used his pre-eminent position in Berlin to block Cantor’s ambition for a post in Berlin or Göttingen. Cantor also believed that Kronecker successfully dissuaded journal editors from accepting his papers for publication – this does indeed appear to be have been the case, for example, with the prestigious Crelle’s Journal, where Kronecker was an editor.

2. Cantor’s transfinite numbers

By 1880 Cantor had become deeply involved in his formulation of the general theory of sets, and in particular in his extension of the number concept to include transfinite numbers, i.e. numbers ‘beyond the finite’. With this, Kronecker’s hostility intensified: his motivation was his adamant refusal to accept any notions of the actual infinite having a place in mathematics. In his ‘arithmetisation programme’ Kronecker had insisted that all mathematics should be capable of being based on a finite number of operations with integers. In his 1886 article ‘Über den Zahlbegriff’ (On the number concept) he objected strenuously to many widely accepted mathematical developments of his time, such as the Bolzano-Weierstrass theorem, claims for the existence of suprema and infima, and even the irrational numbers. For example, having read Lindemann’s proof that $\pi$ is transcendental, he commented:

Of what use is your beautiful investigation of $\pi$. Why study such problems when irrational numbers do not exist.

From this extreme perspective it is no surprise that he saw Cantor’s work as anathema. Sadly, the ensuing controversies and professional disappointments this created probably contributed to Cantor suffering a series of mental breakdowns that greatly hampered his mathematical research for significant periods. Ironically, the origin of the controversy is found in the notion of counting, with which I began this book, and which forms the basis of Pythagorean mathematics as well as of Kronecker’s own position.

Cantor first perceived a need to extend the process of counting ‘beyond the finite’ while dealing with the structural analysis of linear point sets through his concept of derived set. His conclusions were published as a series of six papers entitled “Über unendliche lineare Punktmannigfaltigkeiten (On infinite linear point sets) that appeared between 1879 and 1884.\(^5\) One starting point was his classification of point sets of the first and second species,

\(^5\)For Cantor’s original papers see his annotated collected works, edited in 1932 by Ernst Zermelo [46]. Zermelo comments that the ‘germ’ of Cantor’s theory of transfinite cardinals can be found in the transfinite sequence of derived sets.
where, as shown above, the first species could be divided into sets of the $n^{th}$ kind for $n \geq 1$.

As yet, he had defined no such subdivisions for the second species. In the second paper of his series he observed that, in the sequence of derived sets $P', P'', P''', \ldots$ of a set $P$, each member is a subset of the previous one. When $P$ is a set of the second species, its derived set, again denoted by $P'$, can therefore be written as a disjoint union $P' = Q \cup R$, where $Q$ consists of all points ‘lost’ when we construct a sequence of successive derived sets of $P$ (i.e. given $x \in Q$, there is a smallest $n \geq 1$ such that $x$ does not belong to $P^{(n)}$), while $R = \bigcap_{n \geq 1} P^{(n)}$ consists of all points that belong to every derived set $P^{(n)}$ for $n \in \mathbb{N}$. Since $P$ is not a set of the first species, he knew that $R \neq \emptyset$ and he denoted $R$ by $P^{(\infty)}$, as the ‘derived set of order $\infty$’.

The derived set of $P^{(\infty)}$ would then be denoted by $P^{(\infty+1)}$. He defined the next derived set as $P^{(\infty+2)}$, and continued in this fashion, denoting the $n^{th}$-order derived set of $P^{(\infty)}$ by $P^{(\infty+n)}$. In this way, $P^{(\infty)}$ will also have a ‘derived set of order $\infty$’ consisting of points belonging to every $P^{(\infty+n)}$. This set would be denoted by $P^{(2\infty)}$. Continuing in this fashion, Cantor constructed the derived set of order ‘$n_0\infty + n_1$’ for any natural numbers $n_0, n_1$. This led to the next ‘limit set’ as the set $P^{(\infty^2)} = \bigcap_{n \geq 1} P^{(n\infty)}$, and then to ‘polynomial combinations’ of the symbol $\infty$ in the form $n_0\infty^n + n_1\infty^{n-1} + \ldots + n_v$. (Note that this was simply his notation to identify the ‘positions’ of sets in the list, without implying ‘arithmetical’ operations!) Treating the power $\nu$ as a variable, this leads, as before, to $P^{(\infty\infty)} = \bigcap_{\nu \geq 1} P^{(\infty^\nu)}$. This, in turn, generated new derived sets and the process could be continued indefinitely.

Although at this stage the infinite symbols served primarily as labels by which he could distinguish between various levels of derived sets, Cantor stated boldly that his successive definitions amounted to a ‘dialectical generation of concepts, which continues ever further and, free of any arbitrariness [Willkür], remains consistent and necessary in itself’.

2.1. Cardinal numbers. Thus the roots of Cantor’s investigations of different types of infinite sets are illustrated by two distinct aspects of his early work described above:

(i) his discovery that some familiar infinite sets are of different ‘sizes’;

(ii) his classification of sets of the first and second species (via their sequences of derived sets).

We now consider where these led him in turn.

Cantor’s definition of the cardinal number (or power) of a set was to be of crucial importance, extending the idea of one-one correspondences from $\mathbb{N}$ to sets in general. Readers familiar with modern set theory will be aware
that the definition of this concept has become considerably more sophisti-
cicated since Cantor’s time, depending on the particular axiom system em-
ployed to define the notion of set. We will avoid such issues here and restrict 
our attention essentially to Cantor’s perceptions. What remains basic to our 
(naive) setting is the following:

**Definition**

Two sets $M, N$ are *equipotent* (also called *equipollent*) if there is a one-one 
correspondence (also called a bijection) between them. In that case $M$ and 
$N$ are said to have the same *cardinal number* (or *power*) and we denote this 
by $M \sim N$.

Cantor argued that this concept is not restricted to sets of whole num-
bbers, but ‘ought to be considered as the most general genuine foundation [he used 
the German word ‘Moment’] of sets’. Here we see him already claiming set 
theory as the basis of all ‘pure’ (or, as he would have it, ‘free’) mathematics.

In his later papers (1895/1897) he used the notation $\overline{M}$ to denote the power 
of the set $M$, arguing that it was ‘that general concept which, with the help of 
our active thought-processes, arises from the set $M$, abstracting from the character 
of its various elements $m$ and from the order in which they occur’. Statements 
such as these reveal a philosophical stance that asserts the reality of mental 
constructs and the primacy of consciousness.

I will not use Cantor’s notation, but simply write $|M|$ to denote the car-
dinal number of $M$. For the purposes of discussing its cardinality $M$ can be 
replaced by any set equipotent to $M$. The cardinality of a set has nothing to 
do with any ‘ordering’ of its elements.

At this stage Cantor only had two distinct infinite examples (sets equipo-
tent with either $\mathbb{N}$ or $\mathbb{R}$), but he asserted confidently that the concept of 
power ‘is by no means restricted to linear point sets, but can be regarded as an at-
tribute of any well-defined collection, whatever may be the character of its elements’. 
He was soon to provide a more detailed justification of this claim.

In 1883 he published the fifth article in his series on point sets in the 
journal *Mathematische Annalen*. This was also published separately and is 
now known as his *Grundlagen* [Basics] paper. Here he developed his ideas 
about infinite sets in an abstract setting, using his previous results about de-
rived sets as a guiding example. He prefaced the presentation of his theory 
with an extensive discussion of philosophical and historical objections to 
the concept of *actual infinity*, a concept which he saw as central to his whole 
project.

His research in set theory had reached a point where further progress 
would depend upon a systematic ‘extension of the concept of real whole number 
beyond its previous boundaries’, and this extension had taken a direction that, 
to his knowledge, no-one had taken before him. Although he expressed the 
hope, indeed firm conviction, that the idea of the actual infinite would ‘in
time, have to be regarded as a thoroughly simple, appropriate and natural one’, he
was well aware that he was placing himself ‘in a certain opposition to wide-
spread views about the mathematical infinite and to frequently advanced opinions
on the nature of number’.

He pointed out that Aristotle’s notion of the potential infinite had habit-
ually been used to justify the Calculus. This relied crucially on the idea of
variable finite quantities that could grow or shrink beyond any assignable
bounds, while remaining finite at any particular stage. Cantor would now
refer to such notions as describing improper infinities. By way of contrast, the
widely accepted practice of postulating the existence of a ‘point at infinity’,
which was prominent in both projective and hyperbolic geometry and via
the Riemann sphere in complex function theory, had a quite different char-
acter. These were fixed ideal points, justifying their description as actual, or
as he wished to call them, proper infinities.

2.2. Ordinal numbers. The transfinite numbers he would now define
had the latter character. He had worked with them for some years without
fully realising that they constituted ‘concrete numbers of real meaning’. In con-
trast to a single ideal ‘point at infinity’, he would introduce, successively, an
infinite collection of infinite numbers, all differing from one another. Their
construction would be based on two distinct principles of generation together
with a limitation principle that would serve to distinguish between different
classes of numbers within this collection.

Thus Cantor was ready for the second fundamental innovation sug-
gested by his earlier work on sets of limit points: he would extend the fi-
nite ordinals (or ordinal numbers), indicated by their position in the sequence
1, 2, 3, ..., n, ..., indefinitely beyond the finite. These ideas constitute his sec-
ond major breakthrough, initiating an entirely new subject of study.

In this area also, mathematics has evolved substantially since Cantor’s
time, but his basic ideas largely remain intact. We will see below how—as
Cantor himself described in his later papers—the concept of well-ordering
(which, for \( \mathbb{N} \), follows from the induction principle—see Chapter 7) was to
become fundamental to his theory.

Today, one common procedure is to identify an ordinal as the set of all
ordinals that precede it in the given ordering – as is done for finite ordinals
in von Neumann’s model for \( \mathbb{N}_0 \) in Chapter 7. Thus ordinals ‘label’ the posi-
tions of elements of a set, whose order type is then given by the least ordinal
that is not a member of that set. By way of contrast, cardinals only express

---

6For the last of these, imagine ‘bending’ the Gaussian plane into a sphere, so that the
origin 0 forms the South Pole, while the boundaries of the four quadrants meet at the North
Pole, which we then treat as the ‘point at infinity’, denoted by \( \infty \). Points in the plane with very
small coordinates map to points near 0 while points with very large coordinates map to points
‘near’ \( \infty \).
the ‘size’ of the set. As Cantor pointed out in the construction we describe below, many different infinite ordinals will have the same cardinal number, since the latter takes no account of the ordering of the elements of two sets being compared, only the existence of a one-one correspondence between them.

Cantor’s starting point was to consider how counting, i.e. starting at the unit 1 and successively adding a unit each time, enables one to create the set of all natural numbers, which he denoted by (I) (the ‘first number class’) for this purpose. Its elements were the finite ordinal numbers. Although it would be contradictory to speak of a ‘largest number’ in this set, nothing prevented him from defining a new number \( \omega \) that expressed the natural, regular order of the set (I) as a whole.\(^7\) The symbol \( \omega \) would represent the first transfinite ordinal number, the first number to follow the entire sequence of natural numbers \( \nu \). He argued that it was legitimate to think of \( \omega \) as a ‘limit’ to which the natural numbers \( \nu \) ‘tend’, provided that we meant by this that \( \omega \) should be the first number that follows all the natural numbers.

But, having defined \( \omega \), he could now continue adding units successively, creating new transfinite ordinal numbers

\[
\omega + 1, \omega + 2, \ldots, \omega + \nu, \ldots
\]

again producing a new sequence without a largest element. Nonetheless, the ordinals \( \omega + \nu (\nu \in \mathbb{N}) \) are all equipotent to \( \omega \) (as sets!) so they all have the same cardinal number. Applying the same logic as when discussing derived sets, he defined a new transfinite ordinal number \( 2\omega \) as denoting the set consisting of all numbers of the form \( \nu \) or \( \omega + \nu \), with \( \nu \) taken from (I).\(^8\) As he had done for his derived sets, Cantor now repeated the use of his two generating principles—in describing this we use his original notation. The first principle is the successive addition of units, while the second only comes into play for a ‘definite succession of defined whole numbers...for which there is no largest’. In such situations the new number created is the ‘next number larger than all of them’, the first two examples being \( \omega \) and \( 2\omega \). Using these two principles repeatedly, he could first reach transfinite numbers of the form \( 2\omega + \nu \) for all \( \nu \) in (I), to be followed immediately by \( 3\omega \), then all \( 3\omega + \nu, \ldots, \mu\omega + \nu, \ldots \), etc., so that the ordinal immediately following all these is

\(^7\)He commented in a footnote that he would now use \( \omega \) rather than \( \infty \), precisely because the latter symbol was frequently used to signify the potential infinite (as in \( x = \lim_{n \to \infty} x_n \)) rather than denoting an actual infinite number, as was required here.

\(^8\)In the 1890s Cantor use the notation \( \omega^2 \) instead of \( 2\omega \), and this is the notation used today. One way of envisaging \( \omega^2 \) (or \( \omega + \omega \)) is as two copies of \( \omega \), (e.g.) representing the infinite sequence

\[
1, 3, 5, 7, 9, \ldots; 2, 4, 6, 8, 10, \ldots,
\]

since, just like the set used to define \( \omega^2 \), it has two numbers (1 and 2) which are not immediate successors of any number, and each of the sequences \( 1, 3, 5, \ldots \) and \( 2, 4, 6, \ldots \) can be put in one-one correspondence with \( \mathbb{N} \). What matters here is the order structure rather than any specific ‘labels’ (including any ‘arithmetical’ notation) used to identify individual elements.
denoted by \( \omega^2 \). Using the symbol + to indicate how the successions proceed in each case, he could describe all transfinite numbers of ‘polynomial’ form as \( \nu_0 \omega^\mu + \nu_1 \omega^{\mu-1} + \ldots + \nu_\mu \), where \( \mu, \nu_k \) are natural numbers, while the collection of all of these would be followed by \( \omega^\omega \), and so on!

Today, ordinal numbers such as \( \omega, \omega^2 \) or \( \omega^\omega \), which are neither 0 nor successor ordinals (i.e. produced by adding 1 to an earlier ordinal) are called limit ordinals.

His formulation led Cantor to a limitation principle [‘Hemmungsprinzip’], whereby he would identify breaks in the seemingly endless process of number creation: repeating the essence of the proof that the set \( A \) of all real algebraic numbers is countable, the above collection of numbers of ‘polynomial’ form is shown to be countable, as \( \mu, \nu_k \) are natural numbers.

In his hierarchy of number classes, class (I) comprises the natural numbers. He now defined the second number class, denoted by (II), as:

‘the collection of all numbers, increasing in definite succession, which can be formed by means of the two principles of generation:

\[
\omega, \omega + 1, \ldots, \nu_0 \omega^\mu + \nu_1 \omega^{\mu-1} + \ldots + \nu_\mu, \ldots, \omega^\omega, \ldots, \alpha, \ldots
\]

subject to the condition that all numbers preceding \( \alpha \) (from 1 on) constitute a set of the power of the first number-class (I).’

In other words, ‘initial segments’ of the second number class (II) were to remain countable sets, just as initial segments of class (I) were finite. Cantor went on to prove that number class (II) has a higher power than number class (I), and also that this power immediately follows that of the first.\(^9\) For the last of these claims he needs to make use of the smallest ordinal number in class (III). This is the only occasion (in the *Grundlagen*) where he mentions the third number class (III): it consists of all numbers ‘generated’ by repeating the above process, starting with number class (II) instead of with class (I). Nonetheless, he asserts without further ado that the process of generating new transfinite numbers can be continued indefinitely, creating an unlimited collection of number classes, each subject to his limitation principle, which in its general form states that new transfinite numbers can be created by use of the two generating principles ‘only if the totality of all preceding numbers has the power, in its whole extent, of an already defined number class’.

---

\(^9\)Cantor’s proofs of these claims are quite unwieldy, but already contain the seeds of the notion of well-ordering, which became the centrepiece of his later reformulation of the theory of transfinite numbers. He published his new approach only in 1897. In the *Grundlagen* Cantor simply states without proof that any non-empty sub-collection of the collection of all ordinals has a first element. In [46] his editor Zermelo provides a straightforward proof of this fact.
Basing his treatment on the order concepts he had introduced, he ended the paper by developing the arithmetic of transfinite ordinals in considerable detail, showing that their algebraic properties are quite different from those elaborated earlier for \( \mathbb{N} \), \( \mathbb{Z} \) and \( \mathbb{Q} \). In particular, the commutative laws for addition and multiplication break down even in simple cases. For example, \( 2 + \omega \neq \omega + 2 \). To see why this is so, note that we can write these two sets as follows:

\[
2 + \omega = \{ 1, 2, a_1, a_2, ..., a_\nu, ... \}, \\
\omega + 2 = \{ a_1, a_2, ..., a_\nu, ... ; 1, 2 \}.
\]

These two sets are equipotent, but are not equal as ordinal numbers, as the orderings do not correspond: in the first, only one element, \( 1 \), has no immediate predecessor; in the second there are two such elements, \( a_1 \) and \( 1 \). Similar arguments show that multiplication is not commutative in general: for transfinite ordinals \( \alpha, \beta \) we find that \( \beta \alpha \neq \alpha \beta \), since \( \alpha \) copies of \( \beta \) need not have the same order structure as \( \beta \) copies of \( \alpha \). But we will not delve further into the arithmetic of transfinite ordinals here.

### 3. Comparison of cardinals

A widely used notation for the cardinality of the various number classes, introduced by Cantor in the 1890s, uses the Hebrew letter \( \aleph \) (aleph). It lists the cardinality of the first number class as \( \aleph_0 \) and that of the second number class as \( \aleph_1 \). The precise relationship between \( \aleph_1 \) and the cardinal number of the continuum was to occupy much of his subsequent work. Recall that in the Grundlagen he had shown that the cardinalities of his number classes (I) and (II) were distinct. He had also claimed correctly that the cardinality of number class (II) was the next greatest after that of class (I) in this sequence. Here his arguments, based on what Ernst Zermelo (1871-1953) describes as a ‘purely constructive’ definition of the two ‘generating principles’, lacked much of the clarity of the treatment provided when Cantor revisited the matter in his Beiträge papers in 1895 and 1897. The claim that the alephs constitute an infinite number of distinct cardinal numbers would need further clarification.

A fundamental question concerned the comparability of transfinite cardinal numbers. For any two distinct finite numbers \( m, n \) we know that either \( m < n \) or \( n < m \) will hold; this is what we called the trichotomy for the (total) ordering of \( \mathbb{N} \). To extend this to transfinite cardinals (and thus to justify their designation as ‘numbers’), Cantor suggested in 1887 that, given two sets \( M, N \), the inequality \( |M| < |N| \) should mean that there is a proper subset \( N' \) of \( N \) that is equipotent to \( M \), while no subset of \( M \) is equipotent to \( N \). This ordering of cardinal numbers is easily shown to be transitive (see MM).

Cantor proved that at most one of \( |M| < |N| \), \( |M| = |N| \), \( |N| < |M| \) can hold. This is straightforward: by definition equality cannot hold at the same
time as either of the other relations. But if $|M| < |N|$ then some $N_1 \subset N$ is equipotent to $M$, which means that we cannot have $|N| < |M|$.

However, the same could not be said for the claim that at least one of the above relations must hold for arbitrary sets $M, N$. For given $M, N$ there are two further possible outcomes in addition to the relations $|M| < |N|$ and $|N| < |M|$:

(i) $M$ is equipotent to a subset of $N$ and $N$ is equipotent to a subset of $M$,

(ii) $M$ is equipotent to no subset of $N$ and $N$ is equipotent to no subset of $M$.

Cantor claimed that case (i) would ensure that $M$ and $N$ are equivalent, but he never proved this. It was proved, independently in 1897, by Ernst Schröder (1841-1902) and by his student Felix Bernstein (1878-1956), who had corrected an error in Schröder’s original claim of this result, published in 1896.

The Schröder-Bernstein theorem:

If each of $M, N$ is equipotent to a subset of the other, then $M$ is equipotent to $N$.

Although the proof of this theorem does not require advanced tools, it is by no means obvious (see $MM$). It eluded Cantor until after his principal papers on set theory had been published.

Case (ii) above implies that $M$ and $N$ are not comparable by the relation $<$, which would mean that it is not a total order. Initially, Cantor was unable to exclude this possibility for the cardinals of infinite sets. However, in his Beitrag papers of 1895/97 he provided a complete reformulation of his number classes, based on the concept of a well-ordered set, instead of the more nebulous ‘generating principles’ presented in the Grundlagen.

The modern definition echoes and generalises the Well-Ordering property (WO) proved in Chapter 7 for $\mathbb{N}$:

A set $M$ is well-ordered in a given ordering if every non-empty subset of $M$ has a first element in that ordering.

Cantor’s definition was more elaborate. He first defined a set as simply ordered if for any two of its members one can always be shown to precede the other. Two simply ordered sets $M, N$ are similar if there is a one-one correspondence $\phi$ between them that respects order, i.e. (denoting their orderings by $<_M, <_N$) if $m_1 <_M m_2$ then $\phi(m_1) <_N \phi(m_2)$. The two sets are said to have the same order type – Cantor wrote this as $M = N$ – if and only if they are similar.\(^4\)

He distinguished between number [Zahl] and numbering [Anzahl]. The former relates only to the size (i.e. cardinality) of the set, the latter takes the ordering of the
elements into account, insisting that the one-one correspondence between the sets should preserve the ordering. For finite sets, of course, the two notions coincide, so this distinction would suffice to characterise actual infinite sets. He argued that the centuries-old confusion about potential and actual infinities might have had its origin in the fact that finite numbers function in this dual sense.

The upshot of his reasoning was that the first transfinite number $\aleph_0$ could be taken as that of the first number class (I), in other words, the ordinal $\omega$. The second number class (II) was defined as ‘the entirety of all order types $\alpha$ of well-ordered sets of cardinality $\aleph_0$’. By showing that this is a well-ordered set, he could define the second transfinite number $\aleph_1$ as its least element, and prove the inequality $\aleph_0 < \aleph_1$.

In much the same way as for cardinal numbers, we would today define an order type $\mu$ as any representative of a class of mutually similar sets. Clearly two sets with the same order type define the same cardinal number, but the converse is false in general.

It is clear that well-ordering is an intrinsic part of any counting procedure, as we saw when discussing $\mathbb{N}$. By 1883, Cantor had become aware of the centrality of well-ordered sets for his entire set theory, but he did not prove that his set of transfinite cardinals could be well-ordered. Instead, in the third section of the Grundlagen he made the claim that ‘any well-defined set can be brought into the form of a well-ordered set’. He regarded this as ‘a basic law of thought with far-reaching consequences especially remarkable for its general validity’ to which he promised to return in a later paper. However, by the 1890s he had realised that his bold claim was by no means self-evident. This led him to a thorough reformulation of his transfinite ordinals, published in Part II of the Beiträge (1897), and at last enabled him to resolve the awkward question of the comparability of his alephs.

3.1. Cantor’s second diagonal argument. The publication of Cantor’s Beiträge in 1895 and 1897 met a more receptive audience than had his earlier work in the Grundlagen. His primary critic, Kronecker, had died in 1891, and the younger generation of mathematicians throughout Europe showed greater willingness to grapple with the fundamental questions Cantor’s work had raised. The Beiträge were soon translated widely. They also proved to be more accessible, providing firmer foundations for some of Cantor’s claims that had been the case in earlier work.

Opposition to Cantor’s ideas had not gone away, however. For example, the great French mathematician, Henri Poincaré (1854-1912) remained a stern and influential critic of transfinite numbers, calling the theory a ‘disease’ from which mathematics would eventually recover!

In the late 1880s, disappointed at his failure to obtain the recognition and prestigious position he had hoped for, Cantor was active in campaigning for a new professional body for German mathematicians. The established professional organisation, representing mathematics and medicine,
seemed to him moribund, personifying the academic establishment that had blocked his publications and career aspirations. His advocacy of an alternative resulted in the formation of the Deutsche Mathematiker-Vereinigung (DMV) [German Mathematicians’ Union] which elected Cantor as its first President at its inaugural meeting, held in Halle in 1891.\textsuperscript{10}

Cantor used this occasion to present what has become one of his most distinctive and important contributions. We have already seen an application of his ‘second diagonal argument’ in the proof of the uncountability of the reals. His own description of his simple, yet groundbreaking technique, published under the unassuming title: ‘Über eine elementare Frage der Mannigfaltigkeitslehre (On an elementary question in set theory) and taking up just three pages of the first volume of the DMV’s annual reports, makes interesting reading. As he pointed out, this proof was the first to be entirely independent of the definition of irrational numbers, and lent itself to a vast range of generalisations.

He began with just two distinct elements, $m$ and $w$, and considered the set $M$ of all possible sequences $E = (x_i)_{i\geq 1}$ such that each $x_i$ is either $m$ or $w$. (In these binary days of computer science, we would immediately translate these into sequences using only 0 and 1.) If the set $M$ of these sequences were countable, we could write it as a sequence, so its elements could be listed as

\[
E_1 = (a_{11}, a_{12}, \ldots, a_{1n}, \ldots) \\
E_2 = (a_{21}, a_{22}, \ldots, a_{2n}, \ldots) \\
\vdots \\
E_n = (a_{n1}, a_{n2}, \ldots, a_{nn}, \ldots) \\
\vdots
\]

The sequence $(b_1, b_2, \ldots, b_n, \ldots)$, where, for each $n \geq 1$, $b_n$ is either $m$ or $w$, but where we insist that $b_n \neq a_{nn}$, is obviously a member of $M$ but does not equal any of the $E_i$. Thus $M$ cannot be countable.\textsuperscript{11}

Cantor showed that the diagonal argument can be applied to any set $M$ to show that the cardinality of a set is always less than that of its so-called power set, $P(M)$, defined as the set of all subsets of $M$ (including $\emptyset$ and $M$ itself). To see why, let us begin by counting the subsets of small sets:

\textsuperscript{10}The DMV remains the premier professional organisation for German mathematicians today.

\textsuperscript{11}Cantor added that the same technique can be used to prove the uncountability of $\mathbb{R}$. As Zermelo remarked in a footnote when editing Cantor’s Collected Works in 1932, this claim needs a minor amendment: binary expansions do not represent rational numbers uniquely, since expansions of the form $0.a_1a_2\ldots a_n01111\ldots$ and $0.a_1a_2\ldots a_n10000\ldots$ represent the same rational in $[0, 1]$, for example. But we can always decide in advance which representation to use throughout – exactly as we did for decimal expansions in Chapter 7, Section 6.3.
∅ has only one subset, namely itself,
the singleton set \( \{m\} \) has two subsets, \( \emptyset \) and \( \{m\} \),
the set \( \{a, b\} \) has four, namely \( \emptyset \), \( \{a\} \), \( \{b\} \), \( \{a, b\} \),
the set \( \{a, b, c\} \) has eight: \( \emptyset \), \( \{a\} \), \( \{b\} \), \( \{c\} \), \( \{a, b\} \), \( \{a, c\} \), \( \{b, c\} \), \( \{a, b, c\} \).

In general, a set with \( n \) elements has \( 2^n \) subsets. This follows from the binomial theorem: \((a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}\) (see Chapter 5), taking \( a = b = 1 \); a set with \( n \) elements has \( \binom{n}{k} = \frac{n!}{k!(n-k)!} \) distinct subsets with \( k \) elements.

So what about infinite cardinals? Write the power set of \( M \) as \( \mathcal{P}(M) \). The above suggests the notation \( |\mathcal{P}(M)| = 2^{\lvert M \rvert} \). Cantor’s argument showed that \( \lvert M \rvert < |\mathcal{P}(M)| \), and this immediately provides an infinite, strictly increasing sequence of infinite cardinal numbers. (The proof is given in MM.)

In terms of Cantor’s aleph notation for infinite cardinals, this means that the power set \( \mathcal{P}(\mathbb{N}) \) has a higher cardinal number greater than \( \aleph_0 \). By analogy with a set with \( n \) elements, whose power set has \( 2^n \) elements, we may adopt the notation \( 2^{\aleph_0} \) for the cardinality of the power set \( \mathcal{P}(\mathbb{N}) \).

3.2. Unsolved problems and paradoxes. Despite wrestling with it for many years, Cantor remained unable to resolve a fundamental question that

\[12\]https://commons.wikimedia.org/wiki/File:Henri_Poincaré_sitting.jpg
had occupied him since the late 1870s: given that the real number system \( \mathbb{R} \) is uncountable, is its cardinal number the next greatest after that of \( \mathbb{N} \)? Although he could not prove this, Cantor remained convinced that it is, and this claim became known as his Continuum Hypothesis (CH).

With the notation developed above, Cantor’s Continuum Hypothesis (CH) can now be framed succinctly. He knew that the real numbers can be placed in a one-one correspondence with the power set \( \mathcal{P}(\mathbb{N}) \) of the natural numbers.\(^\text{13}\) Denoting the cardinality of the real number system \( \mathbb{R} \) by \( c \), this means that \( c = 2^{\aleph_0} \).

Cantor’s claim is that there is no set with cardinal number strictly between \( \aleph_0 \) and \( c \). In other words, his Continuum Hypothesis asserts that any subset \( X \) of \( \mathbb{R} \) is either countable or has \( |X| = c \). In the well-ordered sequence of transfinite cardinals \( \aleph_1 \) is the next greatest cardinal after \( \aleph_0 \). Thus the Continuum Hypothesis takes the form: \( \aleph_1 = 2^{\aleph_0} \).

In this form, Cantor’s hypothesis can be generalised in terms of an arbitrary infinite cardinal \( \lambda \). The Generalised Continuum Hypothesis (GCH) states that there can be no infinite cardinal lying between \( \lambda \) and \( 2^\lambda \). In terms of ordinals and alephs it then reads: for any ordinal \( \alpha \), \( \aleph_{\alpha+1} = 2^{\aleph_\alpha} \).

In two critical aspects, therefore, Cantor’s hopes to provide a secure basis for all of set theory were not realised:

(a) He had made no real progress on the question whether every set can be well-ordered, although this claim remained fundamental to his theory.

(b) He had not been able to prove his Continuum Hypothesis.

Moreover, he had become aware that the concept of the set all cardinals, or that of all ordinals, appeared self-contradictory if these were also to be considered as ‘sets’—in other words, his set theory contained paradoxes.

The earliest paradoxes arose when basic questions were asked about the nature of the collection of ‘all’ objects of a particular kind.

(i) The simplest paradox, named after Cantor, questions whether the set of all sets, \( S \), can be a set. If so, it must equal its power set \( \mathcal{P}(S) \): if \( S \) is a set, then \( \mathcal{P}(S) \) is also a set, and by definition it is both contained in (the set of all sets) \( S \) and contains \( \{S\} \) as an element. This, however, yields the contradiction \( |\mathcal{P}(S)| = |S| < |\mathcal{P}(S)| \) by the above diagonal argument, and therefore shows that \( S \) cannot be a set. In other words, the process of set formation

\[^{13}\text{Essentially, map } S \subset \mathbb{N} \text{ to an infinite binary sequence } 0.a_1a_2...a_n..., \text{ using } a_n = 1 \text{ if } n \in S \text{ and } 0 \text{ otherwise. This maps subsets of } \mathbb{N} \text{ injectively into binary representations of real numbers in } [0,1]. \text{ This will imply that } c \geq 2^{\aleph_0}. \text{ On the other hand, treating real numbers as Dedekind cuts (i.e. subsets of } \mathbb{Q} \text{) means that } c \text{ is no greater than } |\mathcal{P}(\mathbb{Q})| = |\mathcal{P}(\mathbb{N})| = 2^{\aleph_0}. \text{ So by the Schroeder-Bernstein theorem, } c = 2^{\aleph_0}.\]
3. COMPARISON OF CARDINALS

without any limitations appears to be highly problematic. Cantor was probably aware of this paradox in 1895, and certainly before he published Part II of his Beiträge in 1897.

(ii) Cantor was aware of a similar result announced in 1897 by Peano’s former student Cesare Burali-Forti (1861-1931). This was the first paradox of set theory to be published. It arises when we consider the set $\Omega$ of all ordinals (recall that ordinals are themselves sets). Now, if $\Omega$ is a set, then we can, according to Cantor’s prescription, form its successor ordinal, which we would denote by $\Omega + 1$. But, as before, we would obtain the nonsensical inequalities $\Omega < \Omega + 1 \leq \Omega$. So, the ‘set of all ordinals’ is also a meaningless concept. Burali-Forti’s paper did not arouse much interest at first, nor was much concern expressed when similar arguments showed that the set of all cardinal numbers, or indeed, the set of all alephs, led to similar contradictions.

To deal with these questions, Cantor sought to distinguish between what he called ‘consistent’ and ‘inconsistent’ concepts. He wished to treat the former as sets, but exclude the latter as ‘absolutely infinite’, which, he argued, ‘can never be conceived complete and actually existing’. To describe this distinction he began to formulate axioms that the process of set-formation would need to satisfy.

There was by now a wider recognition that it was Cantor’s very general definition of what constitutes a set (as given in the Beiträge) that would lead to paradoxes (logicians prefer to call them antinomies, i.e. real contradictions that can be deduced by applying specified logical rules to an apparently true claim). Dedekind’s notion of infinite systems, which he espoused in *Was sind und was sollen die Zahlen?* as an alternative way of describing sets in general, would lead to similar conclusions.

The task of avoiding antinomies was later taken up by Bertrand Russell who argued that, instead of Cantor’s ‘inconsistent’ entities, one should consider properties which do not determine a set (that is, there is no set consisting exactly of the objects that have the property). This conceptual shift, towards describing mathematical entities by means of logical concepts, as well as the search for an axiomatic basis of set theory, was to become a key element of research for several decades, and led to much of the modern subject of mathematical logic.

One particular intervention by Russell was soon to complicate matters further. The catalyst was a letter (dated 16 June 1902) from Bertrand Russell to the German logician and mathematician Gottlob Frege, who had just completed the second volume of his major work *Grundgesetze der Arithmetik* [Basic Laws of Arithmetic]. Russell’s letter led Frege to the conviction that the edifice he had built over a lifetime contained a fundamental flaw. Later
he said that the paradox that Russell had discovered had destroyed set theory! To understand why, we need to outline the background and nature of Frege’s own investigations.

The purpose of Frege’s research had been to base arithmetic upon purely logical concepts. This programme to derive all mathematical principles from the laws of logic alone, became known as logicism. It had attracted mathematical philosophers, including Russell, as well as other mathematicians such as Dedekind and Peano.

In philosophical terms the logicist programme opposed the materialism of David Hume and John Stuart Mill, who argued that our mathematical ideas ultimately arise from our senses through observation. At the same time the logicist viewpoint opposed Kant’s notions of our a priori intuitions of space and time. For example, in the Preface to his Was sind und was sollen die Zahlen?, published in 1888, Dedekind had located his concept of number firmly within ‘the laws of thought’; unlike Hamilton, who had earlier attempted to describe number (and algebra) as reflecting our a priori intuition of ‘pure time’.

In his Foundations of Arithmetic [Grundlagen der Arithmetik], published in 1884, Frege had addressed many of the same issues as Dedekind, but as seen from the viewpoint of a logician rather than as a mathematician. He had developed a meticulous language to express logical concepts, rules of inference and logical axioms. This served to clarify the nature of mathematical reasoning and set the stage for what is known as predicate calculus in mathematical logic today. In his setting, a mathematical proof is a finite sequence of statements, each of which is either an axiom or follows from previous statements in the sequence verified by valid rules of inference.

Frege’s notation and mode of argument would take us beyond the scope of this book, but we can indicate why Russell’s letter had such a destructive impact upon Frege’s system. His two-volume Grundgesetze der Arithmetik (1893/1903) sought, as the title suggests, to identify a small number of ‘basic laws’ of arithmetic upon which the whole structure could be erected solely through the use of logical terms and rules of inference. A key logical axiom Frege needed to complete his programme was his Basic Law V. Rather than use Frege’s abstruse terminology and notation, we explain the difficulty in terms of the (implicit) assumptions about set-formation used by both Cantor and Dedekind, which allowed the formation of sets through self-referential concepts.

\[\text{In his logical universe, containing only objects and functions (the latter taking an object to a value), Frege wished to express the notion of the extension (he used the German word Umfang) of an (unspecified) object. The extension of a concept } F \text{ records the objects for which } F \text{ holds. However, Russell realised that, under Basic Law V one can form a self-contradictory concept, by defining } x \text{ as the extension of some concept which does not apply to } x.\]
Cantor had argued that the term set should apply to ‘every gathering together into a whole of definite, distinct objects m of our perception or of our thought’, while Dedekind said that ‘different things...can be considered from some common point of view, can be associated in the mind, and we say that they form a system S’. Frege criticised these statements, which essentially contend that ‘any precisely specified property’ will suffice to define a set by stipulating the conditions for membership of the set. But despite Frege’s careful construction of his logical system, his Basic Law V in effect amounts to making a similar claim, as Russell pointed out. We can see how Russell created a self-contradictory set under these assumptions:

If we denote the collection of all sets that are not members of themselves by \( R \), then, according to Frege’s Basic Law V (reformulated in terms of sets), \( R \) is admissible as a set. But now we cannot answer the question whether \( R \) is a member of itself! For, if we have \( R \in R \) then, by definition of \( R \), we must have \( R \notin R \). On the other hand, if \( R \notin R \) then, again by definition of \( R \), it follows that \( R \in R \). This then was Russell’s Paradox, which sent the whole logicist programme into considerable turmoil.

It seemed that, in order to maintain the freedom obtained by working with sets in general, contradictions could only be avoided if careful limits were placed on the process of set formation. Set theory urgently required a consistent system of axioms – an axiom system free from contradiction – in which such paradoxes would be avoided.

Russell’s paradox resulted in Frege’s eventual abandonment of his ambitious programme. Bertrand Russell himself, however, devoted much effort over several years to dealing with the paradox he had uncovered. He hoped to avoid antinomies by developing a complex ‘theory of types’, creating an elaborate hierarchy of different types of sets where at each level a set could only contain sets of lower types. Collaborating with Alfred North Whitehead (1861-1947), he produced the massive Principia Mathematica, published in several volumes from 1910 onward, in which they famously arrive at a proof of \( 1 + 1 = 2 \) only after 379 pages.

In many ways the Principia represents the culmination of the logicist project to produce a complete set of axioms and rules of inference within...
symbolic logic from which, in principle, all mathematical truths would follow. It harks back to Leibniz’ search for a *characteristica universalis*, a universal symbolic language in which concepts and ideas could be communicated effectively. But Russell never declared himself fully satisfied with his own efforts, and the focus of the debates about the meaning of mathematical statements shifted to debates about the specific *system of axioms* that would deliver a consistent theory of sets.