This is page 54
Printer: Opaque this

CHAPTER 2

# Set Theory: Taming the Infinite

## 2.1 Introduction

"I see it, but I don't believe it!" This disbelief of Georg Cantor in his own creations exemplifies the great skepticism that his work on infinite sets inspired in the mathematical community of the late nineteenth century. With his discoveries he single-handedly set in motion a tremendous mathematical earthquake that shook the whole discipline to its core, enriched it immeasurably, and transformed it forever. Besides disbelief, Cantor encountered fierce opposition among a considerable number of his peers, who rejected his discoveries about infinite sets on philosophical as well as mathematical grounds.

Beginning with Aristotle (384–322 B.C.E.), two thousand years of Western doctrine had decreed that actually existing collections of infinitely many objects of any kind were not to be part of our reasoning in philosophy and mathematics, since they would lead directly into a quagmire of logical contradictions and absurd conclusions. Aristotle's thinking on the infinite was in part inspired by the paradoxes of Zeno of Elea during the fifth century B.C.E. The most famous of these asserts that Achilles, the fastest runner in ancient Greece, would be unable to surpass a much slower runner, provided that the slower runner got a bit of a head start. Namely, Achilles would then first have to cover the distance between the starting positions, during which time the slower runner could advance a certain distance. Then Achilles would have to cover that distance, while the slower runner would again advance, and so on. Even though the distances would be getting very small, there are infinitely many of them, so that it would take Achilles infinitely long to cover all of them [4, p. 179]. Aristotle deals with this and Zeno's other paradoxes (Exercise 2.1) at great length, concluding that the way to resolve them is to deny the possibility of collecting infinitely many objects into a complete and actually existing whole. The only allowable concept is the so-called potential infinite. While it is inadmissible to consider the complete collection of all natural numbers, it is allowed to consider a finite collection of such numbers that can be enlarged as much as one wishes. As an analogy, when we study a function $f(x)$ of a real variable $x$, then we may be interested in the behavior of $f(x)$ as $x$ becomes arbitrarily large. But we are not allowed simply to replace $x$ by $\infty$ to find out, because we might get a nonsensical result.

A paradox of a more overtly mathematical nature, given by Galileo Galilei (1564–1642) in the seventeenth century, shows that there are just as many perfect squares as there are natural numbers, by pairing off each number with its square:

$$\begin{array}{ccccc} 1 & 2 & 3 & 4 & \cdots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ 1 & 4 & 9 & 16 & \cdots \end{array} .$$

But on the other hand, since not all natural numbers are perfect squares, there are clearly more natural numbers than perfect squares. Galileo even observes that the perfect squares become ever sparser as one progresses through the natural numbers, making the pairing above even more paradoxical. He concludes that "the attributes 'larger,' 'smaller,' and 'equal' have no place either in comparing infinite quantities with each other or comparing infinite with finite quantities" [65, p. 33]. A comprehensive account of the philosophical struggle with the concept of the infinite from early Greek thought to the present can be found in [123] (see also [145]).

It was left to a theologian from Prague in the early nineteenth century to make a systematic study of mathematical paradoxes involving infinity. After studying philosophy, physics, and mathematics at the University of Prague, Bernard Bolzano (1781–1848) decided that his calling was to be a theologian, even though he had been offered a chair in mathematics. As a professor of theology at the University of Prague, beginning in 1805, Bolzano nonetheless spent part of his time pursuing mathematical research. After being dismissed from his position for expressing allegedly heretical opinions in his sermons, and being prohibited from ever teaching or publishing again, he used his enforced leisure to work almost exclusively on mathematics. His philosophical interests had always drawn him to questions about the foundations of mathematics, its definitions, methods of proof, and the nature of its concepts. Thus, he naturally was led to the philosophy of the infinite. Not only did he conclude that mathematics was well equipped to deal with infinite sets in a systematic manner, free of contradictions, he even went as far as arguing that mathematics was the proper realm in which to discuss and resolve *all* paradoxes involving the infinite.[1]

Bolzano made it clear that he considered the property of a *one-to-one correspondence* between a set and a proper subset of itself, as in Galileo's example above, fundamental to the nature of infinite sets, to be exploited in a mathematical investigation, rather than to be used as justification to avoid all discussion of the matter. By a one-to-one correspondence between two sets we mean a pairing of the elements of the sets in such a way that each element of one gets paired with a unique element of the other, without any elements in either set being left over (Exercises 2.2–2.3). The first text in this chapter is a collection of excerpts from Bolzano's book *Paradoxien des Unendlichen* (Paradoxes of the Infinite) [16], published after his death.

Bolzano worked for the most part in isolation, with little contact to the rest of the mathematical world. *Paradoxes of the Infinite* was published in 1851 in Germany, thanks to the efforts of Bolzano's friend F. Prihonsky. Unfortunately, this work, as well as his other manuscripts in analysis and geometry, failed to draw the attention of the mathematical centers in the rest of Europe until quite some time later. When the dramatic events in the theory of infinite sets started to unfold with the work of Georg Cantor (1845–1918) more than twenty years later, the motivation came from the foundations of analysis (see the analysis chapter), one of the major subjects that occupied the mathematical world during much of the nineteenth century, rather than Bolzano's pioneering insights. Since Cantor's work received a fair amount of criticism from theologians, however, the views of the theologian Bolzano were of great importance to him.

Cantor began his mathematical career in number theory, but soon became interested in analysis. During his student days in Berlin, one of his professors was the great Karl Weierstrass (1815–1897), a major figure in the theory of functions during the second half of the nineteenth century, and Cantor began working in this area. An important problem he turned to was the question whether it was possible to represent a function of a real variable by two different so-called trigonometric series, that is, certain infinite series whose terms involve trigonometric functions. Given $f(x)$, can one find real numbers $a_n, b_n$ for $n = 0, 1, 2, \dots$ such that

$$f(x) = \sum_{n=0}^{\infty} a_n \sin(nx) + b_n \cos(nx)?$$

[1] In India, on the other hand, the Jaina mathematics of the middle of the first millenium B.C.E. not only entertained the idea that many different kinds of infinity could exist, but actually developed the beginnings of a system of infinite sets and infinite numbers [91, pp. 18, 218–219, 249–253].

Furthermore, for what values of $x$ is this equation valid, and what conditions on $f(x)$ and the series are necessary? (See the appendix in the analysis chapter for a brief review of infinite series.) Cantor set out to show that whenever such a representation existed it was unique, under quite general conditions. In 1870, he succeeded in showing that such uniqueness followed if the series was convergent for every $x$ with limit $f(x)$. Not satisfied with this result, he worked out several extensions during the following year. After showing first that the result held true if the hypothesis was true for all but finitely many values of $x$, he pushed on to show that certain infinite sets $x_1, x_2, x_3, \ldots$ of exceptional values were allowed. (For details on Cantor's work on trigonometric series see [36, Ch. 2].)

As Cantor was trying to generalize his result yet further, it became clear to him that the sorts of questions he needed to answer about infinite sets of points on the real number line required a much deeper understanding of the nature of the real numbers than was possible with the essentially geometric concept of the number line. What was needed was an "arithmetization" of the real number concept. Besides Cantor, several other mathematicians had recognized the same need and offered their own constructions of the real numbers, depending on what they perceived to be the essential property of the number line (see the analysis chapter). Richard Dedekind (1831–1916) based his construction on the insight that each real number is in some sense determined by all the rational numbers to the left and right of it; it is a "Dedekind cut," in modern terminology. Thus, the real number then could be *defined* as that pair of sets of rational numbers [38]. Dedekind succeeded in defining the arithmetic operations of addition and multiplication on such pairs and proving that they behaved just as one would expect from actual real numbers. Cantor based his definition on the idea that each real number could be defined by a sequence of rational numbers converging to it, and proposed a definition based on such sequences [28, pp. 92–102]. However, neither Dedekind nor Cantor succeeded in proving that their "real numbers" were in fact just like the geometric number line, in the sense that each point on the number line corresponded to a "number" in their system and vice versa. It is considered an *axiom*, a basic truth accepted without proof.

Equipped with his new definition, Cantor went back to study the nature of the exceptional sets of values in his uniqueness theorem, which seemed to be related to the structure of the real numbers themselves. Soon he was on his way to discoveries that would forever change his life and the nature of mathematics. In order to compare different kinds of infinite sets of real numbers, Cantor used the notion of one-to-one correspondence, like Bolzano before him. Two sets $X$ and $Y$ are considered to have the same *power*, in Cantor's terminology, or cardinality, if there is a one-to-one correspondence between $X$ and $Y$. For instance, the set of natural numbers and the set of perfect squares considered above have the same power, since one has the one-to-one correspondence $n \leftrightarrow n^2$. The first important result Cantor stumbled across was a rather amazing fact about the power of the continuum, as the set of real numbers was called. This happened in a roundabout way, whose description requires a bit of terminology. (We follow [36, Ch. 3].)

A real number $r$ is called *algebraic* if it satisfies an equation of the form

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0,$$

where $a_n, \ldots, a_0$ are integers. Otherwise, $r$ is called *transcendental*. For instance, $\sqrt[3]{2}$ is algebraic, since it satisfies the equation $x^3 - 2 = 0$. Cantor had found a proof that the set of algebraic numbers had the same power as the natural numbers, that is, it was *countable*. Likewise, he had shown that the set of all rational numbers was countable as well, as we shall see below. Now he was looking for an application of this result. A good candidate was a theorem proven by the French mathematician Joseph Liouville (1809–1882) in 1851, to the effect that each interval of real numbers contained infinitely many transcendental numbers, and he explained how to produce examples. (See [112, Ch. 12] for details on Liouville's work on transcendental numbers.) Cantor gave a new proof of this result by showing that every interval of real numbers had strictly *larger* cardinality than the natural numbers, that is, it was *uncountable*. Hence, every interval had to contain infinitely many numbers that were not algebraic. In fact, virtually any number that one might randomly pick out of an interval was bound to be transcendental. When Cantor published his result in 1874 [28, pp. 115–118], it had just been proven the year before by Charles Hermite that the Euler constant $e$ is transcendental (see [114]). (It was not until 1882 that F. Lindemann proved that $\pi$ is also transcendental [10].)

Very quickly Cantor realized that the significance of his theorem went far beyond a new proof for Liouville's theorem. He had hit upon an essential characteristic of the *continuous* set of real numbers versus the *discrete* set of integers. More startling results were waiting to be discovered. To his own disbelief, he succeeded in showing that it was possible to put the points of the plane into one-to-one correspondence with the points on the line, and, more generally, the points of a space of arbitrarily high dimension could be brought into such a one-to-one correspondence with the line. This seemed to be a direct contradiction to the notion of dimension, which was thought to be a unique characteristic of a space, not to mention the fact that it went against all common sense. (The notion of dimension was later shown to be indeed characteristic of the space, because it was impossible to find a one-to-one correspondence between sets of different dimensions that was *continuous* [118].) It was this result that prompted Cantor's exclamation, "I see it but I don't believe it!" in a letter to his colleague and friend Dedekind [27]. (There is an interesting story attached to the Cantor–Dedekind correspondence. See [76].) At the end of the 1878 paper [28, pp. 119–133] in which he published these results, Cantor concludes that in order to study general infinite sets in higher-dimensional spaces, it is therefore sufficient to study *linear* sets, that is, infinite subsets of the real line. He then divides these subsets into equivalence classes, with two subsets belonging to the same class if they have the same cardinality. Thus, one such class contains the subset of natural numbers, as well as the integers (Exercise 2.2), the rational numbers, and the algebraic numbers, and all other countable subsets of the real line. There is a second class containing, among others, each interval on the real line, as well as the real numbers themselves. Cantor now makes the following rather amazing pronouncement:

> Through a method of induction, which we shall not describe here, one is led to the theorem that the number of classes resulting from this division of linear sets is finite, namely equal to *two*.
>
> Thus, the linear sets would consist of two classes, of which the first contains all sets that can be given the form of a function $\nu$, where $\nu$ runs through all positive integers, while the second class contains all those sets that can be given the form of a function $x$, where $x$ can assume all values $\geq 0$ and $\leq 1$. According to these two classes, there are only two possible powers of linear sets. We postpone a more detailed investigation of this question until a later occasion [28, p. 132 f.].

That is, Cantor believed that any infinite subset of the continuum, as the real line was called, is either countable or is in one-to-one correspondence with the points of the closed interval $[0, 1]$. The thread that runs through the rest of our story traces the fate of Cantor's unsubstantiated claim, which will lead us right up to the present. In a tremendous eruption of creativity, Cantor now follows up with a series of six papers on the theory of infinite linear sets, published between 1878 and 1884 [28, pp. 139–246]. These papers represent the pinnacle of his mathematical achievements, and he will not produce anything comparable ever again. He begins the fifth of these papers, published in 1883, as follows [28, p. 165]:

> The presentation of my investigations to date in the theory of sets has reached a point where its progress depends on an extension of the concept of whole number beyond its present limits. As far as I know, this extension points in a direction that has not yet been investigated by anybody.
>
> My dependence on this extension of the number concept is so great that without it I would hardly be able to make the least step forward in the theory of sets. This circumstance may serve as an explanation, or as an excuse, if necessary, why I am introducing apparently foreign ideas into my investigations. Namely, I am concerned with an extension, respectively continuation, of the sequence of whole num-

> bers beyond the infinite. As daring as this may seem, I can proclaim not only the hope, but the firm conviction, that this extension will, in time, appear rather simple, appropriate, and natural. I am quite aware that with this undertaking I am in opposition to widely held views about the mathematical infinite and frequently voiced beliefs about the essence of numbers.

As mentioned earlier, the key ingredient in the construction of his new theory of "transfinite numbers" is the tool of one-to-one correspondence and the resulting equivalence relation (Exercise 2.4). To understand the thought behind Cantor's transfinite numbers, we need to think first about finite numbers. The number 5, for instance, can be thought of as the result of abstracting from the elements of the set $\{a, b, c, d, e\}$ by ignoring their particular identity and the order in which they are given. Similarly, Cantor defines infinite "cardinal numbers" as abstractions of infinite sets, by ignoring the particular nature and order of their elements, that is, by retaining only an unordered collection of "placeholders," which are distinct, but otherwise indistinguishable. Thus, given an infinite set, such as the set of natural numbers, the associated cardinal number can be thought of as the equivalence class, under the above equivalence relation, of all sets that are in one-to-one correspondence with the set of natural numbers, that is, the collection of all those sets that are countable. Cantor called this "number" $\aleph_0$ (pronounced aleph naught). (The symbol $\aleph$ is the first letter of the Hebrew alphabet. The reason for the subscript will become clear later on.) Another way to think of $\aleph_0$ is as the collection of properties that the set of natural numbers shares with every other countable set, and only with those. Since he had shown in his 1874 paper that the set of real numbers is uncountable, there is at least one more cardinal number different from $\aleph_0$. Since the natural numbers are a proper subset of the real numbers, one would like to say that this second number is larger than $\aleph_0$.

And indeed, Cantor defines an order relation among cardinal numbers as follows. If $M$ and $N$ are two cardinal numbers, then we say that $M < N$ if there is a one-to-one correspondence between $M$ and a proper subset of $N$ but there is no subset of $M$ that is in one-to-one correspondence with $N$. In trying to understand the nature of this relation, an obvious question presents itself. Are every two cardinal numbers comparable? That is, given cardinal numbers $M$ and $N$, is it always true that either $N < M$, $M < N$, or $M = N$? The affirmative statement is known as the *trichotomy principle*. Or are there two sets such that neither one is equivalent to a subset of the other? While Cantor could show that if $N$ was equivalent to a subset of $M$ and $M$ was equivalent to a subset of $N$, then $M$ and $N$ had to be equivalent themselves, it was going to be a long time before the full trichotomy principle could be established. Since we would expect this principle to hold for a number system, we should actually not have been so hasty to bestow this status on the collection of cardinalities.

But before expending any effort on such questions, one should make sure that there are in fact more than the two infinite cardinal numbers that we know about so far, since otherwise the question is already answered. Cantor found an absolutely ingenious and elementary proof that the power set (the set of all subsets) of any set has a cardinal number strictly larger than the cardinal number of the given set. In particular, the set of real numbers has the cardinality of the power set of the natural numbers (Exercise 2.5). With one stroke of the pen, he showed that there was in fact a whole universe of cardinal numbers out there, waiting to be explored. For these new creations to deserve being called numbers, one should, of course, be able to do arithmetic with them, that is, add them and multiply them. It turned out that union of sets made a good candidate for "addition," with the Cartesian product of sets as "multiplication." One can see that with these definitions, the usual rules of arithmetic hold, such as the commutative, associative, and distributive laws. In addition, when applied to finite cardinal numbers, these two operations reduce to ordinary addition and multiplication. Naturally, one would not expect these new numbers to behave like the finite cardinal numbers in all respects. For instance, the easily verified identity $\aleph_0 + 1 = \aleph_0$ makes for decidedly different calculations. Certain cardinal numbers simply obliterate others under addition. Similar peculiarities happen with multiplication. In fact, we already observed a rather striking one. If we call the cardinal number of the continuum $\mathfrak{c}$, then Cantor's result that the plane has the same cardinality as the line says precisely that $\mathfrak{c} \cdot \mathfrak{c} = \mathfrak{c}$, since the plane can be thought of as the Cartesian product of the line with itself.

One more arithmetic operation is needed, that of exponentiation. Define $M^N$ to be the cardinal number of the set of all functions from $N$ to $M$. As an example, let $N = 2$, and represent 2 by the set $\{0, 1\}$. A function

$$f : N \longrightarrow \{0, 1\}$$

assigns to each element of $N$ either 0 or 1, and defines in this way a subset of $N$, namely that of all elements that are assigned 1. For instance, the function

$$f : \{a, b, c\} \longrightarrow \{0, 1\},$$

which assigns the value 0 to $a$ and 1 to $b$ and $c$, corresponds to the subset $\{b, c\}$. Two distinct functions correspond to different subsets, and each subset corresponds to a function. We therefore obtain a one-to-one correspondence between the set of functions from $N$ to $\{0, 1\}$ and the power set of $N$. Hence, $2^N$ is just the cardinal number of the power set of $N$, just as in the case when $N$ is finite (Exercise 2.21 in the Cantor section). In particular, as mentioned earlier, the power of the continuum is equal to $2^{\aleph_0}$ (Exercises 2.6–2.7). We can now rephrase the claim Cantor made at the end of his 1878 paper about infinite linear sets. Clearly, the cardinal number of each infinite linear set will be greater or equal to $\aleph_0$ and less than or equal to the power of the continuum $\mathfrak{c}$. Furthermore,

$$\aleph_0 < 2^{\aleph_0} = \mathfrak{c}.$$

Cantor's assertion that there are exactly two cardinalities among the infinite subsets of the continuum now is equivalent to saying that $2^{\aleph_0}$ is exactly the next largest cardinal number after $\aleph_0$. This claim quickly became known as the *Continuum Hypothesis*, and its proof was of foremost importance to Cantor, who continued to work on it as long as he was mathematically active. The second and main source in this chapter is an excerpt from a long work that Cantor published in two installments in 1895 and 1897, entitled *Beiträge zur Begründung der Transfiniten Mengenlehre* (Contributions to the Founding of Transfinite Set Theory) [28, pp. 282–351]. In it he gives a grand summary and generalization of his earlier theory.

Another of Cantor's creations was to provide a new perspective on this problem, namely his theory of *ordinal numbers*. To create the cardinal number of a set, we are supposed to abstract from the nature of its elements as well as from the order in which they are given. If we now abstract only from the former, we obtain the *order type* of the set. Call two sets $A$ and $B$, each equipped with a specific ordering, equivalent if there exists a one-to-one correspondence $\phi$ between their elements, which respects the order relation. Let us call such an equivalence an *order equivalence*. That is, if $a < a'$ in $A$, then $\phi(a) < \phi(a')$ in $B$. Naturally, a given set can usually be ordered in several different ways. For instance, the set $\mathbf{N}$ of natural numbers comes equipped with its natural ordering, but one could also order it by reversing the ordering, that is, $1 > 2 > 3 > \cdots$. Then one can show that there is no order-preserving one-to-one correspondence between these two ordered sets.

As another example, consider the set $Q$ of rational numbers between 0 and 1. It, too, is ordered by its natural ordering. Cantor gave another, very interesting, ordering for $Q$ [28, pp. 296 f.]. Let $p/q$ and $r/s$ be elements of $Q$, and suppose that they are written in reduced form, that is, the numerator and denominator have no common factors. Define

$$p/q \prec r/s$$

if either $p+q < r+s$ in the natural ordering of the integers, or $p+q = r+s$ and $p/q < r/s$ in the natural ordering of the rational numbers. Then one can show that with this new ordering, $Q$ is order equivalent to $\mathbf{N}$ (Exercise 2.8). Furthermore, no such order equivalence can exist by using the natural ordering on $Q$. And, incidentally, this order equivalence shows that $Q$ is countable. (Can this equivalence be extended to one between the set $\mathbf{Q}$ of all rational numbers and $\mathbf{N}$?)

Among ordered sets there is a special, distinguished kind, namely the so-called *well-ordered* sets. They have the property that any nonempty subset of a well-ordered set has a least element. The first infinite such set is, of course, **N**. The set $Q$ above with the ordering $\prec$ is well-ordered, even though it is not well-ordered with its natural ordering. (Why?) Thus, a given ordered set may not be well-ordered, but sometimes one can find a different ordering that is a well-ordering. One of the questions that will very soon enter our discussion asks whether this is *always* possible. Can every set be well-ordered? Well-ordered sets should be thought of as particularly simple and reminiscent of the ordered set **N** of natural numbers. They were singled out by Cantor to be the basis of his theory of ordinal numbers. As a result, this theory appears much more familiar and like the theory of natural numbers to us than that of the cardinal numbers.

Cantor defines the ordinal numbers simply to be all the order types of well-ordered sets. We add two ordinal numbers $A$ and $B$ by defining $A+B$ to be the disjoint union $A \uplus B$ with the following ordering. For $x, y \in A \uplus B$, let $x < y$ if either both $x$ and $y$ are in $A$, resp. $B$, and $x < y$ in the ordering on $A$, resp. $B$, or if $x \in A$ and $y \in B$. That is, every element of $B$ is larger than every element of $A$. For example, if $A = \mathbf{N}$, and $B = \{a < b < c\}$, then $A + B$ is the order type

$$1 < 2 < 3 < \cdots < a < b < c.$$

Note that this order type is clearly different from the order type of **N**, which is commonly denoted by $\omega$ (Exercise 2.9). Thus, we see that

$$\omega < \omega + 3.$$

If, on the other hand, we form $B + A$, then we obtain the order type of

$$a < b < c < 1 < 2 < \cdots ,$$

which is easily seen to be the same order type as $\omega$, that is,

$$3 + \omega = \omega.$$

Addition of ordinal numbers is therefore not commutative. Cantor also defined multiplication of ordinal numbers, which we shall not discuss here.

In one essential way, ordinal numbers are much easier to understand than cardinal numbers. Namely, it is not hard to prove that any two ordinal numbers can be compared [63, p. 254]. In other words, the trichotomy principle encountered earlier holds for ordinal numbers. In this way, they form a natural extension of the whole numbers:

$$1, 2, 3, \ldots , \omega, \omega + 1, \omega + 2, \ldots , \omega \cdot 2, \omega \cdot 2 + 1, \ldots .$$

It is time to return to the cardinal numbers, our main topic of interest. To every ordinal number $\alpha$, we can now of course associate its cardinal number by taking any set of ordinal type $\alpha$, which then has a cardinal number associated to it. Since any two sets of the same ordinal type also have the same cardinal type, it does not matter which set we take. This association is not unique (that is, not one-to-one), since different ordinal numbers may well have the same cardinal number, such as $\omega$ and $\omega+1$, for instance. In this way we obtain a special class of cardinal numbers, namely those of well-ordered sets. And we can use certain ordinal numbers to label these cardinal numbers as follows. If we begin with the first infinite ordinal number $\omega$, then its associated cardinal number is $\aleph_0$, which we have already encountered. Moving along in the series of ordinal numbers, at some point we encounter an ordinal that has a cardinal number larger than $\aleph_0$. (It is true, but not immediately obvious, that we will encounter a first such ordinal.) Call its cardinal number $\aleph_1$. Continuing in this way, we obtain a sequence of cardinal numbers

$$\aleph_0, \aleph_1, \aleph_2, \ldots, \aleph_\omega, \aleph_{\omega+1}, \ldots .$$

In particular, to every $\aleph_\alpha$, there is a unique next aleph, namely $\aleph_{\alpha+1}$. In fact, one can show that $\aleph_{\alpha+1}$ is the cardinal number of the well-ordered set of all ordinal numbers whose cardinal number is less than or equal to $\aleph_\alpha$. In particular, this set has larger cardinality than $\aleph_\alpha$. For details see [63, Ch. III, §11].

For the alephs, we have therefore verified the trichotomy principle. But what about the rest of the cardinal numbers? Cantor firmly believed that every cardinal number was in fact an aleph. Equivalently, he believed that every set could be well-ordered. In particular, he believed that the set of real numbers could be well-ordered. The Continuum Hypothesis can then be restated as the claim that

$$2^{\aleph_0} = \aleph_1.$$

In fact, one can now formulate a Generalized Continuum Hypothesis, to claim that if $\alpha$ is any ordinal number, then

$$2^{\aleph_\alpha} = \aleph_{\alpha+1}.$$

At the end of this Herculean construction effort in 1884, two results still eluded Cantor. First, a proof of the trichotomy principle was needed, or, equivalently, a proof that every cardinal number is an aleph, so that consequently every set can be well-ordered. This would provide the final validation of his cardinal numbers as a bona fide number system. Secondly, despite enormous effort, he was unable to prove the Continuum Hypothesis.

While a number of prominent mathematicians realized the magnitude of Cantor's creation, a large segment of the mathematical community viewed his transfinite numbers with skepticism or even outright hostility. This opposition adversely affected his professional and personal life, and might have contributed to a series of nervous breakdowns, which began in 1884 and were to continue for the rest of his life. He repeatedly decided to end his mathematical career, only to return to research, but with diminished power and creativity. Gradually, however, his work gained increased acceptance, and by the turn of the century Cantor's work was widely recognized as revolutionary and of extreme importance. During the Second International Congress of Mathematicians in Paris in 1900, David Hilbert (1862–1943), one of the great mathematicians of both the nineteenth and twentieth centuries, gave a celebrated lecture in which he outlined twenty-three problems, which in his opinion were the major unsolved problems for the new century [42]. As the very first one he listed the Continuum Hypothesis, together with the search for an explicit well-ordering of the real numbers. Anybody who solves one of Hilbert's problems today can expect immediate fame.

By the time the Third International Congress was to be held in Heidelberg, Germany, in 1904, there was growing concern, however, over several antinomies[2] that had been discovered and that threatened the very foundations on which set theory was built. Cantor had in fact observed certain paradoxical phenomena himself as early as 1895. In his attempts to show that every cardinal number was an aleph, he was led to consider the set of all sets. He had shown earlier that the power set of any set had to have larger cardinality than the set itself. This forced the conclusion that the set of all sets had to contain a set of larger cardinality, namely its power set. But this clearly contradicted the most basic results about cardinal numbers that Cantor had proved. The only possible conclusion was that the "set of all sets" could not be regarded as a completed, self-contained object. It was an "absolute infinite" that was beyond intellectual contemplation. This discovery did not worry Cantor particularly; he felt that it simply illuminated the limits of his theory, and indeed the limits of what was subject to rational investigation. Shortly thereafter, similar paradoxes were discovered by other mathematicians, the most startling of which was found by the British philosopher and logician Bertrand Russell (1872–1970). (See [87, pp. 124 f.] for the letter in which Russell communicates the paradox to the German mathematician Gottlob Frege (1848–1925).) His paradox pointed to an even more basic problem inherent in the very fundamentals of set theory. His construction was essentially a modern version of the classical "barber paradox," according to which there is a town with a (male) barber who shaves precisely all those men in the town who do not shave themselves. Then the question is posed whether the barber shaves himself. Either a "yes" or a "no" answer to the question quickly leads to a contradiction. (Try it!) Russell's version of this paradox was the set whose elements were exactly all those sets that do not contain themselves as an element. Here, too, either answer to the question whether this set contains itself as an element leads to a contradiction (Exercise 2.10). Set theory clearly had an inherent problem, and everyone was getting increasingly worried.

But things were about to get worse for Cantor. During the Heidelberg Congress, he attended a talk of the Hungarian mathematician Julius König

[2] Statements that contradict themselves.

PHOTO 2.1. Russell's letter to Frege.

(1849–1914), who claimed to have a proof that the cardinal number of the continuum was not an aleph. Consequently, the real numbers could not be well-ordered. A central pillar of Cantor's belief system appeared to be shattered. Cantor was deeply humiliated and furious, his work and reputation challenged before the assembled mathematical community. Less than twenty-four hours later, the young German Ernst Zermelo (1871–1953) had found a flaw in König's proof. While an immediate catastrophe had been averted, the seeds of doubt were irreversibly sown in Cantor and everybody else. He felt that there was a real possibility now that somebody could find a way to seriously undermine the plausibility of the Continuum Hypothesis. After the Congress, a group of mathematicians, including Hilbert, gathered to discuss the implications of König's paper. Even though it was clear by the end of the Congress that König's proof did not hold water, nobody could be sure that it was not possible to patch it up.

Then, a month later, Zermelo sent an excited letter to Hilbert outlining a proof that *every* set could be well-ordered, and therefore König's proof was impossible to fix. Zermelo's result implied also, of course, a proof that every cardinal number was an aleph, hence the trichotomy principle for general cardinal numbers was proven, since alephs had already been shown to be comparable. In particular, the cardinal number of the continuum was an aleph. Now it was just a matter of deciding which one. All was well, or so it seemed. After the speedy publication of Zermelo's proof and careful scrutiny by his peers all over Europe, criticism focused on one particular step. For a given set $S$, Zermelo assumed that it was possible to simultaneously choose an element from every nonempty subset of $S$. The assumption that this is always possible is known as the *Axiom of Choice*. And while it was pointed out that this axiom was being used tacitly all over mathematics without any objections, it now seemed dubious especially to those who doubted the conclusions one could draw with its help, such as Zermelo's Well-Ordering Theorem, as it became known. Maybe the continuum was too large a set for the Axiom of Choice to be applicable. Zermelo's proof was purely existential; it gave no indication of how to carry out such a choice for a set such as the continuum. The debate was touching the very fundamentals of mathematical reasoning. Was the Axiom of Choice to be accepted as an irreducible mathematical principle, or did it require proof? This issue, together with the paradoxes, made it clear that the foundations of set theory had to be examined very carefully.[3]

Zermelo took on this task. He decided to pursue a strategy that had already led to great success in geometry. While Euclid's *Elements* had been considered the pinnacle of mathematical achievement for two thousand years, it became clear upon closer inspection that much of it could not stand up to the standards of nineteenth century mathematical rigor [96, pp. 188 ff.]. (See the geometry chapter.) A new axiomatic development of geometry had been given by Hilbert in 1899 [88]. Euclid had founded his geometry on a collection of definitions of basic concepts and five axioms, which were to be taken as fundamental, unprovable truths. The whole structure of Euclidean geometry was to unfold from this basis by logical deduction. One problem with Euclid's definitions is, of course, that in order to define a concept, other concepts are needed, which would then need to be defined in turn, leading to an infinite chain of definitions. Hilbert solved this problem by simply leaving notions like “point” and “line” undefined, specifying only their relationships.

Similarly, Zermelo did not attempt to define the basic notion of “set” and “element of,” which seemed to lie at the heart of most of the paradoxes. By 1907 Zermelo had worked out a list of seven axioms, which were to allow the rigorous development of Cantor's set theory, including Zermelo's Well-Ordering Theorem, while at the same time excluding all known paradoxes. Basically, the axioms specified the existence of certain sets and a number of well-defined procedures to build new sets from old ones. These procedures were to be sufficiently powerful to construct all sets needed in the Cantorian theory, but not powerful enough to construct sets large enough

---

[3]While the Axiom of Choice might appear quite reasonable and harmless at first sight, one can actually prove rather amazing facts with it, beyond the well-ordering of sets. For instance, it was proven by the two Polish mathematicians Stefan Banach and Alfred Tarski that one can take a ball in three-dimensional space, cut it up into finitely many congruent pieces, and then reassemble the pieces to obtain a ball of *arbitrarily large* volume. One can furthermore show that this result cannot be proven without the use of the Axiom of Choice. (See [176].)

to lead to trouble, like Russell's paradox. Zermelo published his article *Untersuchungen über die Grundlagen der Mengenlehre, I* (Investigations on the Foundations of Set Theory, I) in the journal *Mathematische Annalen* in 1908. The last original source is an excerpt from this paper including his list of axioms.

While his paper represented a big step in the right direction, Zermelo had not entirely succeeded in addressing all important issues. He could not provide a proof that the axioms were consistent, that is, did not allow the proof of contradictory theorems, and he could not show that they were independent, that is, that none of them could be proven from the others. However, with some subsequent modifications by the German Adolf Fraenkel (1891–1965) and others, the system known as Zermelo–Fraenkel set theory is now the most commonly accepted model for set theory. We shall subsequently refer to this system as **ZFC** to emphasize the special role of the Axiom of Choice, while **ZF** shall denote the system without the Axiom of Choice.

This state of affairs persisted for more than twenty years, despite great efforts by some of the best mathematical minds to resolve the issues of consistency and independence. Then, like a bombshell, news of a most disturbing result hit the mathematics community. The Austrian mathematician Kurt Gödel (1906–1978), destined to become the greatest logician since Aristotle, announced a result at a conference in 1930, now widely known as his *Incompleteness Theorem*, that implied that it was impossible to prove the consistency of the axioms of set theory within that theory. That could only be done in a larger, higher-order theory, which was then in turn subject to questions about its own consistency. With this result, Gödel, who had just received his Ph.D. the year before, doomed all efforts to provide absolute proof of the consistency of set theory, and any other axiomatic theory that contained number theory within it (see the analysis chapter). This negative result was followed in 1935 by a proof that the Axiom of Choice is relatively consistent with the other axioms of **ZFC**. That is, if one could prove a contradiction in **ZFC**, then it could already be proved in **ZF**. And then, in 1939, he supplied a proof that the Continuum Hypothesis, even its generalized form, was consistent with **ZFC**; that is, if one could prove a contradiction using the axioms of **ZFC** together with the Continuum Hypothesis, then this contradiction could already be proved using the axioms alone. Gödel had been in possession of these results since 1937, but had delayed publication in the hope of finding a proof that the Continuum Hypothesis was in fact *independent* of **ZFC**, that is, that the axioms of **ZFC** were not strong enough either to prove or disprove it [42]. But his persistent efforts to substantiate his belief were unsuccessful, even though he continued to work on it for another decade. (A rather unusual view of Gödel's work is contained in [89].)

Such a proof had to wait until 1963, when Stanford mathematician Paul Cohen proved that if **ZFC** is consistent, then it remains consistent if one adds the negation of the Continuum Hypothesis. He also showed that the

PHOTO 2.2. Gödel.

Axiom of Choice is independent of the **ZF** axioms. While this celebrated result provided a temporary ending to almost a century of efforts to prove the Continuum Hypothesis, it is of course a most unsatisfactory one. Most mathematicians are Platonists, in the sense that they believe the objects of mathematics to be real, rather than just the pieces in a formal game with symbols. And if the objects of set theory are considered real, then the Continuum Hypothesis is either true or false. Consequently, the appropriate way to continue is to search for other axioms considered *true*, which, when added to **ZFC**, allow either a proof or a disproof of the Hypothesis. This was Gödel's point of view, and he believed that the Continuum Hypothesis would turn out to be false. Many set theorists today agree. (See [124, 125] for a history of the Continuum Hypothesis and the Axiom of Choice. A good account of the paradoxes of early set theory can be found in [62], and a very good description of Cantor's set theory is in [63].)

Whatever the outcome, the work of Cantor in general and the Continuum Hypothesis in particular has forever changed the face of mathematics. Set theory is at the basis of all mathematics; the fields of mathematical logic, so essential in the design of computers, and topology, widely used in analysis and geometry, are direct outgrowths of the tools developed to solve the problems posed by Cantor's work.

**Exercise 2.1:** Look up the other three paradoxes of Zeno on the impossibility of motion and explain his reasoning. Can you refute any of them?

**Exercise 2.2:** Find a one-to-one correspondence between the set of all integers and that of all positive integers.

**Exercise 2.3:** Show that the points on any two line segments of arbitrary finite length are in one-to-one correspondence.

**Exercise 2.4:** Show that the notion of one-to-one correspondence of sets satisfies the properties of an equivalence relation, namely that it is reflexive, transitive, and symmetric.

**Exercise 2.5:** Show that there is a one-to-one correspondence between the subsets of the set $\mathbf{N}$ of natural numbers and the set of functions from $\mathbf{N}$ into $\{0, 1\}$. Use the binary expansion of real numbers to show that there is a one-to-one correspondence between the set $\mathbf{R}$ of real numbers and the power set of $\mathbf{N}$.

**Exercise 2.6:** Show that if $N$ is a cardinal number, then $N^1 = N$, and $1^N = 1$. If $\mathfrak{c}$ is the cardinal number of the continuum, show that $\mathfrak{c}^{\aleph_0} = \mathfrak{c}$.

**Exercise 2.7:** Extend the product of finitely many cardinal numbers to infinitely many factors. Show that $1 \cdot 2 \cdot 3 \cdot 4 \cdots = \mathfrak{c}$.

**Exercise 2.8:** Using the ordering $\prec$ that Cantor gives for the set $Q$ of rational numbers between 0 and 1, give an explicit one-to-one correspondence between $Q$ and the natural numbers $\mathbf{N}$.

**Exercise 2.9:** If $\alpha$ is the order type of an ordered set $S$, let $\alpha^*$ denote the order type of the set $S^*$ that has the same elements as $S$, but with the reverse ordering. For instance, if $S = \{a < b\}$, then $S^* = \{a > b\}$. Give an example of an ordered set whose order type is $\omega^* + \omega$. (Addition of order types is defined in the same way as that of ordinal numbers.)

**Exercise 2.10:** Write out the details of Russell's paradox.

## 2.2 Bolzano's Paradoxes of the Infinite

"My special pleasure in mathematics rested particularly on its purely speculative parts; in other words, I prized only that part of mathematics which was at the same time philosophy" [178, p. 64] (see also [148, p. 46]). This philosophical bent to Bernard Bolzano's mathematical interests quite naturally attracted him to the study of foundational questions and the philosophy of the infinite.