Cantor's infinities

by Patrick Dehornoy, Nicolas Oresme mathematics laboratory,
Caen University

In 1874, Georg Cantor, then aged 29 and a young professor at Halle University, published a four-page note in *Crelle's Journal*, demonstrating that the set of algebraic numbers is countable, and the set of real numbers uncountable. The article was revolutionary because, for the first time, infinity was no longer considered an unattainable limit but rather a potential object of investigation. The legacy of this work was extraordinary: not only did it herald the dawn of set theory – in fact, a theory of infinity – it also contained the embryonic beginnings of the continuum problem, which would dominate Cantor's later years and remains the driving force behind the development of this theory. Although once the object of rather overblown fascination, itself fuelled by a misunderstanding, this theory is now largely unknown, at the very moment when early signs of a possible solution to Cantor's continuum problem are beginning to emerge.

This text outlines the context and content of Cantor's article, before moving on to discuss the two main developments it entailed, namely the construction of transinfinite ordinals, including the rather amusing application to Goodstein sequences, and the continuum problem, including the frequently encountered misinterpretation of the significance of Gödel and Cohen's results, as well as Woodin's recent results, which hint at what could be a future solution.

A SHORT NOTE AND TWO SIMPLE RESULTS

1. The author

George Cantor was born in St Petersburg in 1845, to a Russian mother and a German businessman father who was of Jewish origin but had converted to Protestantism. He spent his early years in Russia. The family returned to Germany when Georg was 11 years old, first to Wiesbaden and then Frankfurt. Cantor attended the Realschule in Darmstadt, where his mathematical talents were remarked upon, followed by the Zurich Polytechnic in 1862, and, from 1863 onwards, Berlin University, when he was awarded the equivalent of a master's in 1867.

In 1869, at the age of 24, he defended his thesis on number theory, received his accreditation to supervise research, and obtained a post at Halle University (Saxony-Anhalt). There, under the influence of his colleague Eduard Heine (1821-1881), he turned to analysis, focusing on the problem of the uniqueness of the representation of a function by trigonometric series, which he solved positively in 1870. The question would continue to play an important background role in Cantor's thinking on the development of a theory of sets, in particular through the study of what are known as sets of uniqueness.

***Figure 1*: The young Georg Cantor** *(at the time of his 1874 article?).*

From 1872 onwards, Cantor corresponded with Richard Dedekind (1831-1916), who was 14 years his senior and had just put forward the definition of Dedekind cuts of real numbers.[1] It was in this context that Cantor became interested in issues relating to what is now known as countability, i.e. the possibility of enumerating the elements of a set. The fundamental result we will discuss below, namely the uncountability of the set of real numbers, was announced for the first time in a letter to Dedekind dated 7 December 1873. It was published the following year in *Crelle's Journal*, under the title "Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen".[2] This short article

1. A Dedekind cut is a partition of the set of rational numbers into two subsets *A* and *B*, such that any element of *A* is less than any element of *B*; Dedekind demonstrated that the set of cuts behaves exactly as one would expect the set of real numbers to behave, with the cut (*A*,*B*) representing the unique real number between *A* and *B*. Dedekind cuts can thus be used to construct real numbers.
2. "On a Property of the Collection of All Real Algebraic Numbers".

contains two results relating to the possibility, or not, of numbering real numbers.

2. A positive result …

Real numbers are the coordinates of the points on a straight line. In particular, they include the integers 0, 1, 2, etc. and rational numbers in the form *p/q*, where *p,q* are integers and *q* is non-zero. They also include many irrational numbers. A real (or complex) α is termed algebraic if there exists at least one algebraic equation with integer coefficients for which α is the solution. All integers are algebraic since the integer *n* is the (unique) solution to the equation $x - n = 0$. All rational numbers are algebraic since the rational *p/q* is the (unique) solution to the equation $qx - p = 0$. A typical example of a non-rational algebraic number is $\sqrt{2}$, which is the solution to the equation $x^2 - 2 = 0$. There are a great many algebraic numbers: any real number that can be written in integers using the operations $+,-,\times,/,\sqrt{\ },\sqrt[3]{\ },\sqrt[5]{\ },\ldots$ is algebraic. There are also a great many more besides: since Abel, it has been known that algebraic equations exist whose solutions cannot be expressed using the aforementioned operations.

And yet in his text Cantor demonstrates that:

Theorem 1 *Algebraic numbers can be counted.*

In other words: there are not more algebraic numbers that there are natural integers. Cantor's demonstration is not hard to follow:

Demonstration. For any algebraic equation *E*

$$a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n = 0 \quad a_0 > 0,$$

let us call *height* of *E* the integer

$$a_0 + |a_1| + \ldots + |a_{n-1}| + |a_n| + n,$$

and say that an algebraic number α admits *height N* if α is the root of at least one equation of height *N*. Note that a given algebraic number certainly admits an infinity of heights.

By construction, the height of an equation is at least 2. There is only one equation of height 2, namely $x = 0$, and, as a consequence, only one real admitting height 2, namely 0.

In the same way, there are four equations of height 3, namely

$$2x = 0, \quad x+1=0, \quad x-1=0, \quad x^2=0,$$

and, as a consequence, three reals admitting height 3, namely -1, 0, 1.

Thus, for any value of the integer N, there exists only a finite number of equations of height N, with an upper limit of $(2N)^N$, and, it follows, a finite number of algebraic reals admitting height N, with an upper limit of $(2N)^N \cdot N$, since an nth-degree equation has at the most n roots.

Algebraic numbers can therefore be enumerated as follows: first, all the algebraic numbers admitting height 2 are enumerated, followed by all the algebraic numbers admitting height 3, and then by all the algebraic numbers admitting height 4, etc. As every algebraic number admits a height, the list that this creates – which is actually redundant – contains all the algebraic numbers.

3. … and a negative result …

On the other hand, the situation is different if one considers the collection of all real numbers, and this is the second result demonstrated by Cantor:

Theorem 2 *Real numbers cannot be counted.*

Demonstration. Let α_0, α_1, etc. be a given sequence of real numbers. We will demonstrate a real number α that is different from each of the numbers α_n, which shows that no enumeration of real numbers can be exhaustive. Without loss of generality, we assume that $\alpha_0 = 0$ and $\alpha_1 = 1$.

We will attempt to extract a sub-sequence α_{n_0}, α_{n_1}, etc. from the sequence of α_ns, verifying

(1) $$\alpha_{n_0} < \alpha_{n_2} < \alpha_{n_4} < \ldots < \alpha_{n_5} < \alpha_{n_3} < \alpha_{n_1}.$$

We start from $n_0 = 0$ and $n_1 = 1$, and thus from $\alpha_{n_0} = 0$ and $\alpha_{n_1} = 1$. We proceed by induction. Suppose $i > 0$ and n_i constructed. One of the two following statements must be true.

→ Either no integer n verifies

(2) $n > n_i$ and α_n is between $\alpha_{n_{i-1}}$ and α_{n_i} (strictly)

in which case we set down $\alpha = (\alpha_{n_{i-1}} + \alpha_{n_i})/2$, and α is thus distinct from α_n for any n.

→ Or there exists n verifying (2) and thus we define n_{i+1} as being the smallest such integer. Note that, in this case, for each integer n verifying $n_i < n < n_{i+1}$, the real α_n is *not* between α_{n_i} and $\alpha_{n_{i+1}}$ (else this integer n would have been chosen for n_{i+1}).

If the construction is not aborted, we have obtained real numbers α_{ni}, verifying (1). The completeness of $\mathbb{R}$ implies that there exists at least one real number α between the two half-sequences, i.e. verifying

$$(3) \qquad \alpha_{n_0} < \alpha_{n_2} < \alpha_{n_4} < \ldots < \alpha < \ldots < \alpha_{n_5} < \alpha_{n_3} < \alpha_{n_1}.$$

Thus α cannot be equal to any of the reals α_n. By (3), this is clear when n is in the form n_i. Otherwise, there exists i such that n is between n_i and n_{i+1}. In this case however, α, by construction, is between α_{n_i} and $\alpha_{n_{i+1}}$, whereas, as we noted above, α_n is not. Once again then, this gives us $\alpha \neq \alpha_n$.

Cantor notes that, taken together, Theorems 1 and 2 allow for the re-demonstration of the existence of non-algebraic real numbers, which had been established for the first time by Liouville in 1851.

4. Why these results are remarkable

The infinite has appeared in mathematical texts since Antiquity, but it appears in counter-relief, as a negative property (the infinite is that which is not finished) and an unattainable limit, not as an object of study in itself.

Towards the middle of the 19th century, the thinking in this area gained in maturity and the notion of the infinite began to be considered in more mathematical terms. For example, in a posthumous text entitled *Paradoxien des Unendlichen* ("Paradoxes of the Infinite"), published in 1851, Bernhard Bolzano (1781–1848) observed that there are as many elements in the real interval [0,5] as in the interval [0,12] and that, therefore, in an infinite collection, a proper part can be as great as the whole. But this was no more than what Thâbit bin Qurrâ al-Harrânî (836–901) had done a thousand years earlier. Nevertheless, infinity would remain a *terra incognita* and the object of neither result nor demonstration, or even definition; it was only around 1888 that Richard Dedekind explicitly put forward the property mentioned above as the definition of the infinite.

What is profoundly innovative in Cantor's article is the fact that it demonstrates the properties of the infinite. What Cantor does is to demonstrate the first theorem on infinity – in the event, that there exists not one infinity, but at least two: the infinity of algebraic numbers is the same as that of integers, but it is not the same as that of real numbers. Aside from the statement of the result, which is perhaps not so important in itself, what is innovative is the possibility of its very existence: with Cantor, infinity becomes an object of study. The letter to Dedekind of December 1873 is the point of departure for a completely new mathematical theory – the theory of the infinite, which would come to be known as set theory. It is rare that the point of departure for what would become such an important current of thought can be dated with such precision.

One point is remarkable. Cantor entitled his article "On a Property of the Collection of All Real Algebraic Numbers", which corresponds to the first theorem but not to the second, and yet it is the latter which to us seems the most innovative result. As Dauben suggests,[3] it is worth asking whether the emphasis placed on the positive result (it is possible to enumerate …), as opposed to the negative result (it is not possible to enumerate …), might not be a precaution Cantor takes to avoid seeing his article rejected by Leopold Kronecker (1823–1891), the then editor of *Crelle's Journal*, who had the greatest contempt for the notion of the infinite and for all speculations that we would now call non-effective.

5. The diagonal argument

In 1891, eighteen years after the article of 1874, Cantor published a new and even simpler and more striking demonstration of the second theorem. It was this that went down in history as the authoritative demonstration. The so-called *diagonal* argument[4] that is the basis of this demonstration has elements in common with a construction developed in 1875 by Paul du Bois-Reymond (1831-1889), but the combination of self-reference and a negation, which is the decisive point, seems to be used for the first time in Cantor's text. As we now know, this argument gave rise to great things, since it is the fundamental technical ingredient in several of the great results of logic produced in the 20th century, in particular Russell's paradox, Gödel's incompleteness theorems, Turing's construction of undecidable sets, and hierarchy theorems in complexity theory.

> *Demonstration of Theorem 2 by the diagonal argument.* Let α_0, α_1, etc. be a given sequence of real numbers. We will once again demonstrate a real number α that is different from each of the α_ns. This time, we will not use

3. J. W. Dauben, *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Cambridge, Mass., 1979; reed. 1990.

the order of the real numbers, but rather the existence of a decimal development for each real number. For each integer *n*, there exists an infinite sequence $(a_{n,1}, a_{n,2}, ...)$ of digits between 0 and 9, such that we have

$$\alpha_n - E(\alpha_n) = \overline{0, a_{n,1} a_{n,2} \cdots},$$

where $E(x)$ designates the integer part of x, and this sequence is unique if there is a requirement that the digits $\alpha_{n,i}$ must not all be equal to 9 from a certain rank onwards. We thus set down $0^* = 1, 1^* = 2^* = ... = 9^* = 0$, and let α be the real whose decimal development is

$$\overline{0, a^*_{1,1} a^*_{2,2} \cdots}.$$

Thus, regardless of *n*, the real α is different from α_n, since the *n*th digit of the development of α_n is $\alpha_{n,n}$, whereas that of α is $a^*_{n,n}$, which is different from $\alpha_{n,n}$ by construction.

THE LEGACY (1): ORDINALS

The reach of Cantor's article was quite extraordinary: set theory as a whole, and, from there, a significant proportion of 20th-century mathematics, can be traced back to it. This legacy can be described using Cantor's later work as a point of departure. Still in Halle, where he became a professor in 1879 at the age of 34, Cantor was becoming increasingly interested in what would become set theory. Between 1879 and 1884 he published a series of six articles in *Mathematische Annalen*; these would form the basis of the theory.

***Figure 2*: Georg Cantor, probably in the 1880s.**

We will distinguish two main themes within this legacy, the first being the possibility to count beyond the finite, which leads to the notion of the transfinite ordinal.

6. A theory of transfinite numbers

What Cantor showed was that once the conceptual barrier that made the infinite inaccessible had been crossed, there was nothing to prevent the development of an arithmetic of infinite numbers (or rather, transinfinite numbers, i.e. beyond finite), one very similar to the arithmetic of integers, and which could be used in particular for inductive demonstrations.

The idea is to extend the sequence of integers, i.e. to count beyond 0, 1, 2, etc. To do this, the principle that Cantor uses to underpin his construction is a well-known property for integers and one that is found at the heart of demonstrations by induction, namely that any non-empty set has a smaller element. What Cantor observes is that, if we maintain this principle, there is only one way to extend the sequence of integers. For example, there must exist a smaller transfinite number that is greater than all the integers, and Cantor calls it ω. Next, there must exist a smaller transfinite number that is greater than ω, and Cantor calls it $\omega+1$. Next, of course, come $\omega+2$, $\omega+3$, etc., followed by a smaller transfinite number that is greater than all the $\omega+n$s and which is called $\omega+\omega$, or $\omega\times2$. We continue with $\omega\times2+1$, and then a little later on, with $\omega\times3$, and so on. There exists a smaller transfinite number that comes after all the $\omega\times n$s, and it is written as $\omega\times\omega$, or ω^2. There's no reason to stop now, and so further down the line we come to ω^3 and ω^4, then ω^ω and ω^{ω^2}, and eventually even to ω^{ω^ω} … and beyond that, to many more transfinite numbers besides.

What Cantor demonstrates – without however convincing the very reticent Kronecker – is that the above description is not simply a gratuitous and uncertain extrapolation, but rather a coherent system that can be used in demonstrations. For example, at the same time as the Swedish mathematician Ivar Bendixson (1861–1935), Cantor himself used it in 1883 in his study of sets of uniqueness to provide a demonstration by induction on transfinite numbers. This produced what has remained a famous result on the structure of the closed subsets of a real straight line, namely that any such set can be written as the meeting of a countable set and a set in which all the points are accumulation points.

7. An amusing application of ordinals to a mathematical demonstration

Transfinite numbers – now known as *ordinals* – are a powerful means to demonstrate mathematical results. What is interesting – and can appear paradoxical – is that the use of infinite ordinals sometimes makes it possible to demonstrate properties of finite objects that would otherwise remain inaccessible. A spectacular example is provided by the convergence of Goodstein sequences in arithmetic. These are sequences of integers defined by a simple induction from the notion of development in iterated base *p*. Developing an integer *n* in base *p* consists in decomposing *n* in the form of a decreasing sum

$$n = p^{n_1} \times c_1 + \ldots + p^{n_k} \times c_k$$

where the figures c_i range between 1 and p – 1, and the exponents n_i are integers, necessarily strictly smaller than *n*. One can then express the exponents n_i themselves in base *p*, and iterate the process. The expression thus obtained can be called the iterated development of *n* in base *p*. For example, the development of 26 in base 2 is $2^4 + 2^3 + 2^1$: the development of 4 is 2^2, that of 3 is 2^1+1, and, finally, the iterated development of 26 in base 2 is $2^{2^{2^1}} + 2^{2^1+1} + 2^1$.

Definition 1 *(i) For* $q \geq 2$, $T_{p,q} : \mathbb{N} \to \mathbb{N}$ *is defined as follows:* $T_{p,q}(n)$ *is the integer obtained by replacing every p by q in the iterated development of n in base p and by evaluating the result.*
(ii) For each integer d, the Goodstein sequence in base d is defined as the sequence of integers g_2 , g_3 , …, *defined by* $g_2 = d$, *then, inductively,* $g_{P+1} = T_{p,p+1}(g_p) - 1$ *if* g_p *is not zero, and* $g_{P+1} = 0$ *if* g_p *is zero.*

For example, starting with $g_2 = 26$ one finds:

$$T_{2,3}(26) = T_{2,3}(2^{2^{2^1}} + 2^{2^1+1} + 2^1) = 3^{3^{3^1}} + 3^{3^1+1} + 3^1 = 3^{27} + 3^4 + 3 = 7625597485071$$

hence $g_3 = 7625597485070$. Once then does the same thing again, replacing 3 with 4, and so on and so forth. It seems clear that the sequence thereby obtained very rapidly tends towards infinity. And yet in 1942 Reuben Goodstein (1912–1985) demonstrated the following result:

Theorem 3 *For any integer d, the Goodstein sequence in base* d *converges towards 0: there exists an integer* p *verifying* $g_p = 0$.

Demonstration. The argument is extremely simple if we use ordinal arithmetic. To do this, we introduce a function $T_{p,\omega}$, analogous to $T_{p,q}$, for each integer *p*, but which goes from $\mathbb{N}$ into the ordinals: $T_{p,\omega}(n)$ is the ordinal obtained by replacing *p* by ω in the iterated development of *n* in base *p*. Thus, for example, we have

$$T_{2,\omega}(26) = T_{2,\omega}(2^{2^{2^1}} + 2^{2^1+1} + 2^1) = \omega^{\omega^{\omega^1}} + \omega^{\omega^1+1} + \omega^1.$$

The properties of the arithmetic of ordinals entail that each of the functions $T_{p,\omega}$ is a strictly increasing function. Thus, for $p \geqslant 2$, we set down $\widetilde{g}_p \doteq T_{p,\omega}(g_p)$. For each integer *d*, we thus have a sequence of ordinals $\widetilde{g}_2, \widetilde{g}_3$, etc. Yet, by construction, for each *p* such as g_p being non-zero, we have

$$\begin{aligned} \widetilde{g}_{p+1} = T_{p+1,\omega}(g_{p+1}) &= T_{p+1,\omega}(T_{p,p+1}(g_p) - 1) \\ &< T_{p+1,\omega}(T_{p,p+1}(g_p)) = T_{p,\omega}(g_p) = \widetilde{g}_p. \end{aligned}$$

By construction, then, regardless of *p*, *q*, *r* verifying $2 \leqslant p \leqslant q \leqslant r \leqslant \omega$,, and regardless of *n*, we have $T_{q,r}(T_{p,q}(n)) = T_{p,r}(n)$, and, in particular, $T_{p+1,\omega}(T_{p,p+1}(n)) = T_{p,\omega}(n)$. The fundamental property of the sequence of ordinals, namely that any non-empty set has a smaller element, entails that any strictly decreasing sequence of ordinals must be finite. There thus necessarily exists an integer *p* such that $\widetilde{g}_p$ is zero, and therefore g_p is also zero (Figure 5).

$$\begin{tikzcd}
\widetilde{g}_2 \arrow[r, equal] & \widetilde{g}_2 \arrow[r, ">"] & \widetilde{g}_3 \arrow[r, equal] & \widetilde{g}_3 \arrow[r, ">"] & \widetilde{g}_4 \arrow[r, equal] & \cdots \\
g_2 \arrow[u, "T_{2,\omega}"'] \arrow[r, "T_{2,3}"'] & \bullet \arrow[u, "T_{3,\omega}"'] \arrow[r, "-1"'] & g_3 \arrow[u, "T_{3,\omega}"'] \arrow[r, "T_{3,4}"'] & \bullet \arrow[u, "T_{4,\omega}"'] \arrow[r, "-1"'] & g_4 \arrow[u, "T_{4,\omega}"'] \arrow[r] & \cdots
\end{tikzcd}$$

***Figure 3**: Demonstration of Goodstein's theorem: below, the integers, above, their images among the infinite ordinals, which erase the changes of base; this leaves only the -1s, which force the decrease so long as 1 is not yet reached.*

The essential point in the preceding demonstration is the existence of the ordinal ω, that is to say, the existence of a transfinite number that dominates all the integers in the same manner as ω does, *i.e.* the distance between 3 and ω is the same as that between 2 and ω.

What is remarkable is that Theorem 3 – which is a pure result of arithmetic, in the sense that it is a stated in such a way as to bring into play only integers and their elementary operations, and which was demonstrated so simply using infinite ordinals – cannot be demonstrated without reaching for such a tool. Indeed, drawing on a method developed by Paris and Harrington in 1978, in 1981 Kirby and Paris demonstrated that Goodstein's theorem cannot be demonstrated using Peano axioms alone, in other words, from within the framework of standard arithmetic. In one sense, this result legitimates Kronecker's suspicion of Cantor's methods;[5] in another, it illustrates their visionary scope.

8. Ordinals today

Over a century later, tensions have cooled, the fundamental questions have been elucidated, and ordinals and transfinite induction are both part of mathematicians' palette of tools. However, with the exception of mathematical logic and some aspects of theoretical computing (the termination of rewriting systems), which often exploit these tools, it has to be said that these elegant, powerful tools remain little used in core mathematics – with a few notable exceptions such as Martin's Borel determinacy theorem. This is not very surprising to that extent that, in the end, mathematics makes only limited use of current infinity, that is to say, an infinity that is not simply the indefinite continuation of the sequence of integers.

THE LEGACY (2): THE CONTINUUM PROBLEM

The other direct legacy of the article of 1874 is the theory of cardinals and its central problem, the continuum problem, which entails determining the cardinal of the set of real numbers. Cantor sought the solution to the continuum problem for the rest of his life. Still in Halle – Kronecker's opposition prevented him from finding a post in Berlin – he continued to develop his theory of sets, with notable results such as the diagonal argument of 1891 and the theory of the comparability of cardinalities in 1897. However, he never solved the continuum problem. His final years were rather sad. Although the scientific scope of his work had been widely recognised by his peers, his life from 1884 until his death in 1918 was overshadowed by scientific polemics and above all by increasingly severe episodes of depression that led to him being committed to care institutions on several occasions.

5. Kronecker would certainly not have been reassured to learn that the smallest integer *p* for which the *p*-th term in Goodstein's sequence in base 4 equals zero is $3 \times 2^{402653211} - 2$.

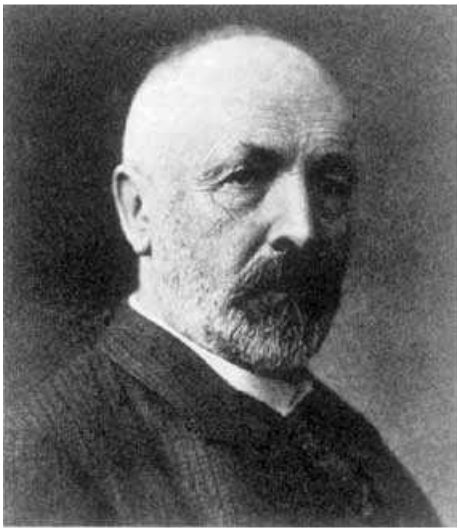

Figure 4: George Cantor, probably around 1900.

The continuum problem for its part remained at the heart of set theory throughout the 20th century. This is more than ever the case today, given that hopes of a solution are beginning to emerge.

9. An infinity of infinities

Theorem 2 showed that there are at least two infinities that cannot be placed in bijective correspondence, namely the infinity of the set *N* of natural integers and the infinity of set *R* of real numbers. Intuitively, then, these infinities are not of the same size, and Theorem 2 introduces a new field of enquiry – the comparison of the size of infinities.

As early as 1878, Cantor suggested formalising the comparison of sizes in the terms we still use today, namely the existence of bijections and injections: we say that a set *A*, whether finite or infinite, is the same size (or has the same cardinality) as a set *B* if there exists a bijection of *A* onto *B*; similarly, we say that A is at the most the same size (or has the same cardinality) as *B* if there is an injection of *A* into *B*. It should be noted that, in the case of finite sets, these definitions correspond closely to the standard comparison of numbers of elements. In 1897, Cantor – and, at the same time, Felix Bernstein (1878–1956) and Ernst Schröder (1841–1902) – would show that this comparison of cardinalities is a total order: if there is an injection of *A* into *B* and an injection of *B* in *A*, then there is also a bijection of *A* onto *B*.

It should also be noted that the theory of infinite cardinalities soon became separate from that of transfinite ordinals: while counting and ordering are equivalent tasks for finite sets, the same is not true of infinite sets. More precisely, there exists only one type of total order on a finite set of a given cardinality, whereas there exist multiple two-to-two non-isomorphic total orders on an infinite set. For example, from the point of view of size, $\mathbb{N}$ and $\mathbb{R}$ are equivalent, whereas they are not when provided with their usual orders.

In this context, Theorem 2 states that there are at least two distinct infinite cardinalities. Cantor himself would demonstrate a much stronger result using a form of the diagonal argument.

Theorem 4 *The set of integers, its set of parts, the set of parts thereof, and so on, are two-to-two of different sizes.*

Demonstration. We start by demonstrating that, regardless of set *E*, there is no surjection, and certainly no bijection, of *E* onto the set of parts $\mathfrak{P}(E)$. Let *f* be a given application of *E* in $\mathfrak{P}(E)$. We will show that *f* is not surjective by demonstrating a part of *E* that does not belong to the image of *f*. To this end, we set down

$$A = \{x \in E \mid x \notin f(x)\}. \tag{4}$$

Let α be a given element of *E*. One of the two following statements must be true. Either α is in *A*, which signifies that α does not belong to $f(\alpha)$. As α is in *A*, $f(\alpha)$ is not equal to *A*. Or α is the complement of *A*, which signifies that α belongs to $f(\alpha)$. As α is not in *A*, $f(\alpha)$ once again is not equal to *A*. Therefore *A* cannot belong to the image of the application *f*, yet this application is not surjective.

This being demonstrated, set down $E_0 = \mathbb{N}$, $E_1 = \mathfrak{P}(E_0)$, $E_2 = \mathfrak{P}(E_1)$, etc. According to the above, regardless of *i*, there cannot be a bijection of E_i onto E_{i+1}. Furthermore, for any non-empty set *E*, the application that sends *X* onto *x* if *X* is the singleton $\{x\}$, and onto a fixed element x_0 , if this is not the case, is a surjection of $\mathfrak{P}(E)$ onto *E*. Consequently, for each *i* and *j* verifying $j \geqslant i + 2$, there exists a surjection of E_j onto E_{i+1}. Thus, if there existed a bijection of E_i onto E_j, we would deduce by composition a surjection of E_i onto E_{i+1}, in contrast to what we have seen above.

The two ingredients of the diagonal argument can both be seen in Theorem 4, namely the combination of self-reference (here, the simultaneous use of *x* and *f*(*x*), as with the diagonal digits $a_{i,i}$ in section 1), and a negation ($x \notin f(x)$ here and the use of $a_{i,i}$ * in section 1).

10. The continuum hypothesis

As soon as there is an infinity of different infinite cardinalities, one obvious question is to determine the position of the cardinalities of the most common sets, $\mathbb{N}$ and $\mathbb{R}$. For the cardinality of $\mathbb{N}$, one can easily see that it is the smallest of the infinite cardinals: $\mathbb{N}$ injects into any infinite set.[6] According to Theorem 2, the cardinality of $\mathbb{R}$ is strictly greater than that of $\mathbb{N}$. This is what we call the continuum problem:[7] determining which infinity is the cardinality of $\mathbb{R}$.

In 1877, before he had even established the existence of an infinity of infinities and their comparability, Cantor had predicted a solution to the continuum problem with the continuum hypothesis:

Any infinite part of $\mathbb{R}$ that is not in bijection with $\mathbb{N}$ is in bijection with $\mathbb{R}$.

The continuum hypothesis signifies that there is no set of a strictly intermediary size between those of $\mathbb{N}$ and $\mathbb{R}$, i.e. in terms of cardinalities, that the cardinality of $\mathbb{R}$ (the continuum) is an immediate successor to that of $\mathbb{N}$ (the countable).

Cantor never managed to demonstrate (or refute) the continuum hypothesis. The only notable result he would demonstrate with regard to the continuum problem is the so-called Cantor–Bendixson theorem on the structure of closed sets (mentioned above). This straightforwardly entails that any closed subset of $\mathbb{R}$ that is not in bijection with $\mathbb{N}$ is in bijection with the total $\mathbb{R}$: it can thus be said that closed sets fulfil the continuum hypothesis. Alas, Cantor was never able to obtain an analogous result for the more complicated subsets of $\mathbb{R}$ – and the developments in set theory in the 20th century show that mathematicians of that time certainly lacked the technical means to do so.

Nevertheless, Cantor's later work, though it in no way solves the question, make it possible to formulate the continuum problem in the concise, symbolic form in which it is often stated today. The point of departure is another fundamentally important result from Cantor, adding considerable precision to Theorem 4 and to the Cantor–Bernstein–Schröder theorem.

6. In modern terms, there needs to be at least a weak form of the axiom of choice in order to confirm this statement; these on the whole minor questions would only arise later and do not really affect Cantorian theory of cardinals, which is, in the main, a theory of well-orderable sets, i.e. a theory where the axiom of choice is valid.
7. In Cantor's time, the set of real numbers was called the continuum.

Theorem 5 (Cantor) *There exists a sequence of cardinalities indexed by the ordinals* $\aleph_0 < \aleph_1 < \aleph_2 < \ldots < \aleph_\omega < \aleph_{\omega+1} < \ldots$ *such that any*[8] *infinite set admits one (and only one) of the alephs as its cardinality.*[9]

Thus, not only is there an infinity of infinities, we also have a complete description of the structure of this family of infinities, i.e. a well-organised sequence indexed by the ordinals. Theorem 5 in particular shows that, for each cardinality κ, there exists a smaller cardinality that is strictly greater than κ, which is not at all obvious: on the face of it, it would have been quite possible for the order of cardinalities to include dense intervals, like the order of rational numbers.

From this point onwards, $\aleph_0$ is the cardinality of $\mathbb{N}$, and the continuum problem becomes a case of determining which aleph is the cardinality of $\mathbb{R}$. The continuum hypothesis thus assumes the simple form card($\mathbb{R}$)= $\aleph_1$, since $\aleph_1$ is, by definition, the immediate successor to $\aleph_0$ in the sequence of alephs.

In addition, it is easy to define a bijection[10] between $\mathbb{R}$ and $\wp(\mathbb{N})$, and therefore between $\mathbb{R}$ and the set of applications of $\mathbb{N}$ in {0,1}. Since the cardinality of $\mathbb{N}$ is $\aleph_0$, it is natural to note the cardinality of $\{0,1\}^{\mathbb{N}}$ as $2^{\aleph_0}$, which is also therefore the cardinality of $\mathbb{R}$. With this formalism, the continuum hypothesis corresponds to the equality

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

11. The development of a new discipline

In 1900, David Hilbert (1862–1943) presented his famous list of 23 problems for 20th-century mathematics to the International Congress of Mathematicians in Paris, with determining the truth or falsity of the continuum hypothesis ranked in first place. This is a sign that the reticence surrounding the use of infinity other than as an unattainable limit had been dispelled and that the fundamental nature of Cantor's work had been recognised. Hilbert described the transfinite arithmetic of Cantor as "the most astonishing product of mathematical

8. Here again, it is necessary to specify "any well-orderable set" to take into account problems of choice.
9. $\aleph$ is the first letter, *aleph*, of the alphabet.
10. For example, it is possible to associate each real between 0 and 1 with the sequence d_i of 0 and 1 in its development in base 2, and to then associate this sequence with the part of N composed of indexes where this sequence equals 1 (if d_i = 1, i is in the part of N; if d_i = 0 is not in the part of N).

thought, one of the most beautiful realisations of human activity in the domain of the purely intelligible".

Direct progress on the continuum problem was slow, since it could be achieved only once a considerable foundational base had been developed. Following on from the Cantor–Bendixson theorem demonstrating that closed sets fulfil the continuum hypothesis, one early result was the theorem demonstrated in 1916 by the 20-year-old Pavel Alexandroff (1896–1982): Borel sets fulfil the continuum hypothesis.[11] We now know that this result was the best that could be produced at that time; indeed, this avenue of research could be pursued only from the 1970s onwards, with the development of what is known as the descriptive theory of modern sets, i.e. the close study of the subsets of $\mathbb{R}$.

The great difficultly of the continuum problem, and, more generally, all problems bringing into play infinity – indeed, the one that made a solution practically impossible in Cantor's time – was the absence of a precise and widely accepted conceptual framework to develop a theory and demonstrate results. Cantor did indeed propose what became a standard definition of the set ("any collection into a whole *M* of definite and separate objects *m* of our intuition or our thought; these objects are called the elements of *M*"). But that would not have been enough to define the rules of the game, namely determining where to start to demonstrate the properties of sets. At the same time as other thinkers, such as Cesare Burali-Forti (1861–1931) and Bertrand Russell (1872–1970), Cantor himself recognised the difficulties raised by the imprecision surrounding the notion of a defined object, and it was only at the beginning of the 20th century that these points began to be elucidated: what mattered to mathematicians was not defining what a set was, but simply obtaining a consensus as to the way in which sets function, that is to say, a point of departure from which to demonstrate theorems. In 1908, Ernst Zermelo (1871–1953) put forward an axiomatic system for sets, which was amended in 1922 by Adolf Fraenkel (1891–1965). This system, known as the Zermelo–Fraenkel system or the *ZF* system, rapidly established itself as a standard point of departure for set theory, in the same way that the Euclidean system is a point of departure for geometry and the Peano system a point for departure for arithmetic.

11. But Alexandroff, disappointed at not having solved the continuum problem, became a theatre producer and only returned to mathematics years later.

12. Two major results …

Once it had been established that the *ZF* system should be the point of departure for a theory of sets, the first stage towards a solution to the continuum problem was to determine if the continuum hypothesis is or is not a consequence of the axioms of this system. This is not the case: Kurt Gödel (1906–1978) – followed, 20 years later, by Paul Cohen (1934–2007) – demonstrated two negative results.

> ***Theorem 6* (Gödel, 1938)** *Unless these are contradictory,*[12] *the negation of the continuum hypothesis is not a consequence of the axioms of the ZF system.*
>
> ***Theorem 7* (Cohen, 1963)** *Unless these are contradictory, the continuum hypothesis is not a consequence of the axioms of the ZF system.*

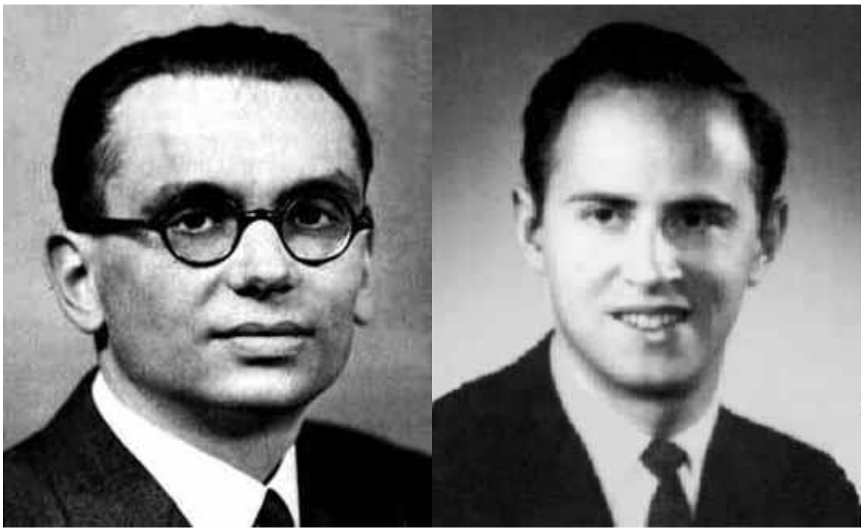

***Figure 5*: Kurt Gödel (left) and Paul Cohen (right)**

Gödel's and Cohen's theorems are rightly considered major steps forward. The demonstration of these theorems required the deployment of completely new means – the method of inner models in Gödel's case, and the method of forcing in Cohen's. Aside from the purely technical difficulties (which, even several decades on, remain substantial), these results necessitated a complete change in perspective on set theory, one analogous in many respects to the Copernican revolution or the discovery of non-Euclidean geometries: it entailed a shift from a vision of a single world of sets to one of a multiplicity of possible worlds.

12. This rhetorical note of caution is necessary because Gödel's second theorem of incompleteness prevents us from establishing the non-contradictory character of the *ZF* system; it is therefore impossible to rule out *a priori* the hypothesis that this system is contradictory.

… and two misunderstandings

In the later decades of the 20th century, the fate of set theory was scarcely happier than the personal fate of its creator, Georg Cantor. Two misunderstandings were at the origin of this situation.

The first misunderstanding was due to the very success of set theory. What Zermolo was probably the first[13] to grasp was the possibility of using sets as the unique basis of the totality of the mathematical edifice. Stated more precisely, it is possible to represent functions (Felix Hausdorff, 1918), integers (John von Neumann, 1923) and, from there, almost all mathematical objects, as sets. Although remarkable, this result – which was systematically deployed in the Bourbaki treatise a few years later – was taken to be much more than it actually is, which is a result of coding (or coordinisation) analogous to the possibility of coding the points of a plane with a couple of real numbers or with a complex number. Little concerned with accuracy, imitators saw an ontological result where there was only a question of coding, positing the dogma of "everything is a set" and having set theory play the role of an all-encompassing whole – a claim it never made for itself: it is difficult for a lay mathematician to believe that the integer 2 is the set $\{\varnothing,\{\varnothing\}\}$, since there is no intuition to support such an identity, and no demonstration of it can be adduced. It was therefore inevitable that the excessive enthusiasm aroused by this approach would be disappointed and that the on the whole minor applications of the theory to wider mathematics would entail a sentiment of rejection to match the initial hopes. The pedagogical excesses of the 1960s, stemming directly from the confusion between "everything can be represented by sets" and "everything is a set", clearly did not restore the image of a misunderstood theory, one often imagined as the manipulation of diagrams of ellipses and crosses that were as abstract as they were devoid of mathematical meaning. The theory of sets is the theory of the infinite and it has very little to do with contemporary mathematicians' – admittedly convenient – everyday use of an elementary overarching vocabulary.

The second misunderstanding is due to the signification of Gödel's and Cohen's theorems. What the intelligent general public, and many mathematicians, have taken from these is that the continuum problem is an insoluble problem, and will remain so for eternity. Some imagine that it is

13. It seems that Cantor did not anticipate this aspect of the development of set theory.

possessed of a mysterious status, being neither true nor false, or else that it is intrinsically unknowable, or devoid of any true meaning. Gödel's and Cohen's results say something quite different, and much more simple: they say, or rather illustrate – because the idea had been known about since Gödel's incompleteness theorems – that Zermelo-Fraenkel's *ZF* system is incomplete, contains lacunae and does not exhaust the properties of sets. What is almost universally accepted is the fact that the *ZF* axioms express properties of sets that our intuition induces us to accept as true. In other words, we judge it opportune to take these axioms as a point of departure and to accept that their consequences are valid. But no one – in any case, no specialist of set theory – has ever claimed that the *ZF* axioms exhaust what we intuitively know about sets. This notion is now being explored, and it is quite possible that in the future a consensus will emerge as to the opportunity of adding new axioms, as and when we recognise new properties of sets and infinity as being pertinent. Gödel's and Cohen's results thus *opened up* the continuum problem far more than they closed it down.

The continuum problem today

Almost 50 years since Cohen's result, the continuum problem has not been solved, but important progress has been made and it is not inconceivable that a solution should emerge in the foreseeable future. The major development in set theory since the 1970s has been the gradual emergence, based on a considerable corpus of convergent results, of a consensus as to the opportunity of adding additional axioms to the ZF system, thereby affirming the existence of ever-greater infinites, which is merely a natural iteration of Cantor's approach. These axioms, known as "large cardinals", have varied technical forms. The most important today is the DP (projective determinacy) axiom: in the 1980s, D. A. Martin, J. Steel and H. Woodin demonstrated that the ZF+DP system provides an equally satisfying description of the world of countable sets as the ZF system provides for the world of finite sets, that is to say, a description which, in practice and heuristically, appears to be complete. Based on the ZF+DP system, the same is also true of the next level of complexity, that of topology and analysis in what are known as projective sets. This type of heuristic completeness has led to a general consensus about adding the DP axiom to the ZF axioms to create a new point of departure for set theory.

***Figure 6*: The mathematician Hugh Woodin (born in 1955)**

The situation for the next level in the scale of complexity – which is level 2 if arithmetic is level 0 and projective analysis level 1 – is less clear. For the moment, there is no consensus as to the axiom(s) which could play the same role for level 2 as the DP axiom plays for level 1. On the other hand, one remarkable result already exists, demonstrated by Hugh Woodin in 2001, namely that **any** axiom which, for level 2, produces the type of description produced by ZF for level 0, and ZF + DP for level 1, entails that the continuum hypothesis is false. This result is not yet the conclusive solution to the continuum problem since, on the one hand, Woodin's theorem remains conditioned by a conjecture that is not as yet fully established, and, on the other, by the fact that there is no consensus as to what a complete description of level 2 might be. Nevertheless, we have here a result that seems to both tip the balance in favour of the falsity of the continuum hypothesis and to show that very important progress is possible with regard to the continuum problem. There is no reason to think that it will not be solved in the future.

CONCLUSION

In one sense, the continuum problem is a minor problem in mathematics; few applications really depend on the continuum hypothesis, and the only statements linked to it bring into play objects that are either very big or very complicated, and for this reason rather removed from the core of contemporary mathematics. That is probably one of the reasons why the continuum problem, which was first on Hilbert's list of problems in 1900, was not mentioned a century later on in the list of Millennium Problems put forward by the Clay Institute;[14] in the category of foundation problems, it was replaced there by the 'P vs NP

14. Clay Mathematical Institute, *Millennium Problems*: http://www.claymath.org/millennium-problems

problem'. On the other hand, this problem remains as fascinating as ever because it is both fundamental and so simply worded, and it remains the driving force behind research in set theory. It is certain that more will be known in a hundred years, and the author of these lines would be most curious to know where the continuum problem and the exploration of infinity will be by then. In any case, if one thing is certain it's that Cantor, with his unassuming note of 1874, opened up a world and has given mathematicians something to chew on for centuries.

■

The author wishes to thank Pierre Ageron and Akihiro Kanamori for the useful information they provided, as well Jean Fichot for pointing out the early existence of a form of the diagonal argument in the work of Paul du Bois-Reymond.

This text is based on the lecture "Un texte, un mathématicien", which was delivered by the author at the Bibliothèque Nationale de France on 18 March 2009, and has been adapted for BibNum *by Alexandre Moatti.*

(publication in French, April 2009)
(English translation by Helen Tomlinson, published November 2016)