The History of Infinity[1]

What is it?
Where did it come from?
How do we use it?
Who are the inventors?

1   The Beginning

As there is no record of earlier civilizations regarding, conceptualizing, or discussing infinity, we will begin the story of infinity with the ancient Greeks. According Anaximander of Miletus (mid-6th century BCE) and student of Thales, the word *apeiron* meant unbounded, infinite, indefinite, or undefined. Originally it was used to reference the *unlimited* mass possessing no specific qualities. This was the universal state preceeding the *separation* into contrasting qualities, such as hot and cold, wet and dry. Thus *apeiron* represented the primitive unity of all phenomena. For the Greeks, the original chaos out of which the world was formed was *apeiron*.

Eventually, it became a negative, even pejorative word. Aristotle thought being infinite was a privation not perfection. It was the absence of limit. Pythagoreans had no traffic with infinity. Everything in their world was number. Indeed, the Pythagoreans associated good and evil with finite and infinite. Though it was not well understood at the time, the Pythagorean discovery of incommensurables, for example $\sqrt{2}$, would require a clear concept and understanding of infinity.

Yet, to the Greeks, the concept of infinity was forced upon them from the physical world by three traditional observations.

- Time seems without end.

- Space and time can be unendingly subdivided.

- Space is without bound.

That time appears to have no end is not too curious. Perhaps, owing to the non-observability of world-ending events as in our temporal world of life and death, this seems to be the way the universe is. The second, the apparent conceivability of unending subdivisions of both space and time, introduces the ideas of the infinitesimal and the infinite process. In this spirit, the circle can be viewed as the result of a limit of inscribed regular polygons with increasing numbers of sides.[2] These two have had a lasting impact, requiring the notion of infinity to be clarified. Zeno, of course, formulated his paradoxes by mixing finite reasoning with infinite and limiting processes. The third was possibly not an issue with the Greeks as they believed that the universe was bounded. Curiously, the prospect of time having no beginning did not perplex the Greeks, nor other cultures to this time.

With theorems such that the number of primes is without bound and thus the need for numbers of indefinite magnitude, the Greeks were faced with the prospect of infinity. Aristotle avoided the actuality of infinity by defining a *minimal* infinity, just enough to allow these theorems, while not introducing a whole new *number* that is, as we will see, fraught with difficulties. This definition of potential, not actual, infinity worked and satisfied mathematicians and philosophers for two millenia. So, the integers are potentially infinite because we can always add one to get a larger number, but the infinite set (of numbers) as such does not exist.

Aristotle argues that most magnitudes cannot be even potentially infinite because by adding successive magnitudes it is possible to exceed the bounds of the universe. But the universe is potentially infinite in that it can be repeatedly subdivided. Time is potentially infinite in both ways. Reflecting the Greek thinking, Aristotle says the infinite is imperfect, unfinished and unthinkable, and that is about the end of the Greek contributions. In geometry, Aristotle admits that points are on lines but points do not comprise the line and the continuous cannot be made of the discrete. Correspondingly, the definitions in Euclid's *The Elements* reflect the less than clear image of these basic concepts. In Book I the definitions of point and line are given thusly:

---

[2]It was this observation that first inspired mathematicians to believe that the circle can be squared with a compass and straight edge. It is a simple matter to square a polygon. Can the limiting case be more difficult, it was thought? It was, and would not be resolved until 1881 by Lindemann.

> Definition 1. A *point* is that which has not part.
> Definition 4. A *straight line* is a line which lies evenly with
> the points on itself.

The attempts were consistent with other Greek definitions of primitive concepts, particularly when involving the infinitesimal and the infinite (e.g. the continuum). The Greek inability to assimilate infinity beyond the potential-counting infinity had a deep and limiting impact on their mathematics.

Nonetheless, infinity, which is needed in some guise, can be avoided by inventive wording. In Euclid's *The Elements*, the very definition of a point, *A point is that which has no part*, invokes ideas of the infinite divisibility of space. In another situation, Euclid avoids the infinite in defining a line by saying it can be extended as far as necessary. The parallel lines axiom requires lines to be extended indefinitely, as well. The proof of the relation between the area of a circle and its diameter is a limiting process in the cloak of a finite argument via the method of exhaustion. Archimedes proved other results that today would be better proved using calculus.

These theorems were proved using the method of exhaustion, which in turn is based on the notion of "same ratio", as formulated by Eudoxus. We say

$$\frac{a}{b} = \frac{c}{d} \quad \text{if for every positive integers } m, n \text{ it follows that}$$

$ma < nb$ implies $mc < nd$ and likewise for $>$ and $=$.

This definition requires an infinity of tests to validate the equality of the two ratios, though it is never mentioned explicitly. With this definition it becomes possible to prove the Method of Exhaustion. It is

> By successively removing half or more from an object, it's
> size can be made indefinitely small.

The Greeks were reluctant to use the incommensurables to any great degree. One of the last of the great Greek mathematicians, Diophantus, developed a new field of mathematics being that of solving algebraic equations for integer or rational solutions. This attempt could be considered in some way a denial of the true and incommensurable nature of the solutions of such equations.

Following the Greeks, the Arabs became the custodians of Greek heritage and advanced mathematical knowledge in general, particularly in algebra. They worked freely with irrationals as objects but did not examine closely their nature. This would have to wait another thousand years.

---

2   Ideas become clearer

Following the Arabs, European mathematicians worked with irrationals as well, though there was some confusion with infinity itself. St. Augustine adopted the Platonic view that God was infinite and could have infinite thoughts. St Thomas Aquinas allowed the unlimitness of God but denied he made unlimited things. Nicolas of Cusa (1401 - 1464) was a circle-squarer[3] that used infinity and infinite process as analogy to achieving truth and heavenly Grace. A type of paradox arose in medieval thinking. It was understood that a larger circle should have more points than a smaller circle, but that they are in one-to-one correspondence. (See below.)

---

[3]Though many ancient Greeks believed that the circle could not be squared by a compass and straight-edge, Cusa thought he could do so. One key was his belief that a circle is a polygon with the greatest possible number of sides.
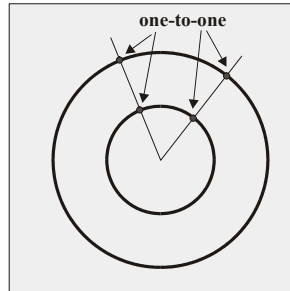
In 1600 Galileo (1564 - 1642) suggested the inclusion of an infinite number of infinitely small gaps. But he understood the problem was using finite reasoning on infinite things. He said, "It is wrong to speak of infinite quantities as being the one greater or less than or equal to the other." With the insight of genius, he claimed infinity is not an inconsistent notion, but rather it obeys different rules.

In a more practical direction, Leonardo of Pisa (ca. 1170 - ca. 1250), known better as *Fibonacci*, demonstrated a cubic equation that could not be solved within the context of any of the numbers discussed in Euclid. (That is those numbers of the form $\sqrt{\sqrt{a} \pm \sqrt{b}}$, where a and b are rational.) Moreover, confusion was evident in understanding the nature of irrationals and its ultimate link with infinity. In his book *Arithmetica Integra* of 1544, Michael Stifel (1487-1567) makes the following observations about irrationals. There are irrationals because they work in proving geometrical figures. But how can they *be* because when you try to give a decimal representation they flee away. We can't get our hands on them. Thus, an irrational is not a true number, but lies hidden in a cloud of infinity. This typifies the confused, uncertain feeling of professional mathematicians, while clearly illustrating the connection to infinity.

The nature of infinity was not clarified until 1874, with a fundamental paper by Georg Cantor. In the interim, calculus and analysis was born and fully developed into a prominent area of mathematics.

Steven Simon (1548-1620), an engineer by trade, was one of the earliest mathematicians to abandon the *double reductio ad absurdum* argument of antiquity and adopt a limit process without the "official"

trappings of the Greeks, the *double reductio ad absurdum* argument. This was the acceptance of limits as an infinite process not requiring metrization. In one result, Simon proves that the median of a triangle divides it into two triangles of equal area.
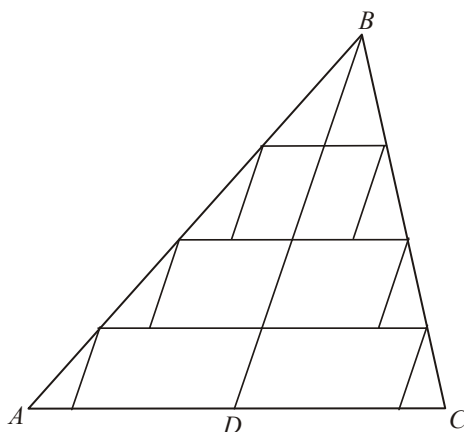
He accomplished this by a successive subdivision argument into rectangles and estimating the excess. was a practical mathematician/engineer who desired to establish results in an understandable way and to spread the new decimal methods. The limiting part of his argument, that $\frac{1}{2^n}$ tends to zero as $n \to \infty$, he took as self-evident.

Fermat took limiting processes in another direction in proving quadrature formulas for power functions x^p. His arguments appear in many ways modern, though again, his limiting process involves an essential step not unlike Simon's.

At this point the following arguments seem certain. There can be no theory of irrationals without a working facility and definition of infinity. Without a theory of irrationals, there can be no analysis, and without analysis, mathematics would be without a major branch. Even still, the understanding of polynomials can never be complete without a thorough understanding of irrationals, though not perhaps in the same way.

3   The Emergence of Calculus

John Wallis (1616-1703), arguably the most important mathematician in 17th century England except Newton, was Savilian professor of geometry at Oxford, having originally studied theology. In his work *Arithmetica Infinitorum* he extends the work of Torricelli (1608 - 1647)and Cavalieri (1598 - 1647) on *indivisibles*[4] and establishes, by a great leap of induction that

$$\frac{4}{\pi} = \frac{3 \cdot 3 \cdot 5 \cdot 5 \cdot 7 \cdot 7 \cdot 9 \cdots}{2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \cdot 8 \cdot 8 \cdots}$$

This infinite expansion for π, though not the first, clearly illustrates an infinite process without justification. In 1657 Wallis gives the symbol, ∞, which indicates an unending curve. It caught on immediately. He also introduced fractional power notation.

Like Wallis, Newton, Leibniz, the Bernoulli's, Euler, and others that invented and then pursued the new calculus, there was little serious regard for proof and for any theory of limits and the infinite. An ∞ appearing in a computation would be attributed to be a paradox. The mathematical legitimacy of the calculation of derivatives by Newton, based on *moments* was faulted by George Berkeley, Anglican bishop of Cloyne, in his book *The Analyst*. Let's review the argument for computing the derivative of x^2 alá Issac Newton (1642-1727). We compute the difference

$$\begin{aligned} (x+o)^2 - x^2 &= x^2 + 2xo + o^2 - x^2 \\ &= 2x\,o + o^2 \end{aligned}$$

Divide by the moment o to get

$$2x + o$$

Now *drop* the term o to get the derivative $2x$. This was exactly what Bishop Berkeley objected to. How, he argued, can this mathematics be legitimate when on the one hand one computes with the term o as if it is a true number and then simply eliminates it when needed.

---

[4]The theory of indivisibles is one of the more curious "false starts" on the road to calculus. With it one assumes that an area consists of a continuum of vertical lines extending from the bottom of the region to the top. Requiring subtle interpretations of area, it was difficult to apply. Indeed, though quadratures for powers were obtained, quadratures of other functions were not obtained in this way.

Berkeley did not object to the spectacular results this new analysis was achieving, but his objection struck at the heart of what had not yet been mathematically articulated as a legitimate process. He writes,

> And what are these fluxion? The velocities of evanescent increments? And what are these evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?

Newton did give a definition of the derivative similar in appearance to the modern definition but sufficiently far off the mark not to satisfy objections. There resulted, on the basis of Berkeley's objection, a strong effort to place calculus on a theoretical foundation, but this was not to be achieved for another two centuries. In *A Defense of Free-thinking in Mathematics* of 1735, which was a response to a rejoinder to *The Analyst* Berkeley devastates the new analysis:

> Some fly to proportions between nothings. Some reject quantities because [they are] infinitesimal. Others allow only finite quantities and reject them because inconsiderable. Others place the method of fluxions on a foot with that of exhaustions, and admit nothing new therein. Others hold they can demonstrate about things incomprehensible. Some would prove the algorithm of fluxions by *reductio ad absurdum*; others *a priori*. Some hold the evanescent increments to be real quantities, some to be nothings, some to be limits. As many men, so many minds... Lastly several ... frankly owned the objections to be unanswerable.

The great Leonard Euler (1707-1783) did not improve the theoretical state of affairs at all. He pursued the new analysis with an abandon that would have cautioned even Newton and Leibniz. Consider these two series studied by Euler.

$$\frac{1}{(x+1)^2} = 1 - 2x + 3x^2 - 4x^3 \cdots \qquad (*)$$

$$\frac{1}{1-x} = 1 + x + x^2 + x^2 + \cdots \qquad (**)$$

Put $x = -1$ into (*) and there results

$$\infty = 1 + 2 + 3 + \cdots$$

Put $x = 2$ into (**) and there results

$$-1 = 1 + 2 + 4 + \cdots$$

The series for -1 is term-for-term greater than the series for ∞. Therefore,

$$-1 > \infty$$

These sort of computations were prevalent in the analysis of the day and were called paradoxes. By substituting $x = -1$ into (**), Euler also noted

$$\frac{1}{2} = 1 - 1 + 1 - 1 + \cdots$$

Euler freely allowed $\frac{0}{0}$ to have a definite value, and thereby was influential in advancing the proportion between nothings. Such was the state of affairs, a field exploding with knowledge and profound results that still impact modern mathematics, with intrinsic inconsistencies neither understood nor for which there was anything resembling a theory. The Greek model of rigorous, axiomatic geometry had been forgotten.

---

4   The Roots of Infinity

What finally forced the issue were consequences owing to trigonometric series. Jean d'Alembert (1717-1783) derived essentially the modern wave equation for the vibrating string, and showed that trigonometric series could be used to solve it. This was a considerable departure from power series, for which most mathematicians understood the limits of validity. On the other hand, trigonometric series were new and more difficult to analyze. d'Alembert limited himself to initial conditions that were periodic functions, making the analysis easier. Euler, shortly afterward allowed the initial condition to be any function free of jumps as the string was one piece.[5] He insured the periodicity by extending it

---

[5]To the mathematicians of this time, the concept of function was murky. At this time Euler, interpreted a function to be an analytical expression, or a finite number of analytical expressions pieced together.

periodically outside the interval. Daniel Bernoulli (1700-1782) took the ideas further by claiming all the new curves, those defined piecewise by expressions, could be represented by trigonometric series. This was soundly rejected by d'Alembert and Euler. Euler argued that functions cannot be continuous and discontinuous.

The state of affairs remained unresolved for almost a century until Jean Baptiste Joseph de Fourier (1768-1830) applied trigonometric series to the heat problem. The trigonometric series are similar to those for the wave equation, but the requirement of continuity of initial conditions was not demanded, if only on physical grounds, as it was for the vibrating string. His fundamental paper of 1807 was rejected by no less than Legendre, Laplace, and Lagrange, though later his continued work was encouraged. Fourier returned to the interpretation of the coefficients of the Fourier series as areas, as opposed to antiderivatives. Consider the sine series.

$$
\begin{aligned}
f(x) &= \sum_{n=1}^{\infty} b_n \sin nx \\
b_n &= \frac{2}{\pi} \int_0^{\pi} f(s) \sin ns \, ds
\end{aligned}
$$

Fourier held that every function[6] could be represented by a trigonometric series.

The question of the day and for sometime to come was this: **Classify the functions for which the Fourier series converge.** This simple question had a profound impact on the development of analysis and literally forced rigor upon the subject, first for the ideas of continuity, then for the definition of the integral, and finally for the notion of set. This in turn put mathematicians square up against infinity itself. This is one of the more curious *threads* in the history of mathematics. A relatively straight forward problem led to the creation of set theory, functional analysis, and the rigorization of analysis. Below are listed a few steps along the way.

- 1817 — Bolzano tried to prove what is now called the Intermediate Value Theorem, but was stalled because no theory of the reals

---

[6] Bear in mind that Fourier viewed a function in a far narrower sense that we do today. Basically, for Fourier a function was a piecewise continuous function, express by a finite number of pieces of expressible analytic functions.

existed. He needed the theorem that every bounded set has a least upper bound.

- 1821 — Cauchy gives a nearly modern definition of limit and continuity, though he uses uniform continuity when he hypothesizes pointwise continuity. He defined convergence and divergence of series, and produced what is now known as the Cauchy convergence criterion.

- 1823 — Cauchy gives the definition of integral in terms of the limit of sums of rectangles. This makes Fourier's use of integrals more rigorous.

- 1829 — Dirichlet gives a condition on a function to have a convergent Fourier series. (The function must be monotonic with a finite number of jumps.) This condition, that the number of discontinuities the function can have is finite is due to the current state of the theory of integrability. As an example of a function that cannot be integrated, he produces the function

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is rational} \\ 1 & \text{if } x \text{ is irrational} \end{cases}$$

sometimes called the "salt and pepper" function.

- 1831 — Carl Frederich Gauss (1777 - 1855) objected to using infinity in, "I protest against the use of an infinite quantity as an actual entity; this is never allowed in mathematics. The infinite is only a manner of speaking, in which one properly speaks of limits to which certain ratios can come as near as desired, while others are permitted to increase without bound."

- 1850 — Karl Weierstrass gives the modern δ, ϵ definition of continuity, discovers and applies uniform convergence and gave a theory of irrational numbers (1860) as series of rationals. (For example, $\sqrt{2} = 1 + \sum_{n=1}^{\infty} \binom{1/2}{n}/n!$.) Predecessors had defined irrationals, if they did at all, as limits of rationals. However, Cantor observed irrationals must already exist in order for them to be a limit of sequences. Using the method of condensation of singularities, Weierstrass produces a continuous function that is nowhere differentiable. This defeated decades of research to prove that all continuous functions must be differentiable, except at perhaps a set of exceptional points.

- 1854 — Georg Friedrich Bernhard Riemann (1826-1866), Dirichlet's student, gives a more general definition of integral, the Riemann integral, thought by many to be the most general possible. He gave an example of a function with an infinite number of discontinuities that has an integral. Riemann posed many problems about Fourier series including those that led to set theory.

- 1858 — Dedekind gives a theory of irrational numbers based on *cuts*, now called Dedekind cuts.

- Cauchy and Weierstrass eliminate infinitesimals and infinite values and replaced them by infinite processes and conditions — very Eudoxian to say the least.

---

5   Infinity and Georg Cantor

Georg Cantor (1845 - 1918) was a student of Dedekind and inherited from him the problem of establishing the class of functions which has a converging Fourier series. Following his teacher, he began to study families of functions having convergence Fourier series as classified by their exceptional points. That is, following even the first ideas of convergence, Cantor expanded the number of exceptional points a function may have and still have a converging Fourier series — except at those points. His first attempt in 1872 allowed for an infinite number of exceptional points answering a question of Riemann.

Here are the details. Given an infinite set of points S. Define the *derived* of S ,S', to be the set of limit points of S. Define S'' to be the derived set of S', also called the *second derived* set of S, and so on. Cantor was able to show that if the trigonometric series

$$0 = \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx$$

converges to zero except at a set of points which has a finite k^{th} derived set, for some (finite) k, then $a_n = b_n = 0,\ n = 1, 2, \ldots$. In this paper he also showed the existence of such sets for every n.

Cantor most certainly was aware that the process of derivations could be carried out indefinitely. Use the notation $S^{(n)}$ to be the nth derived set of S. Then $S^{(n+1)} = (S^{(n)})'$, the derived set of S^n. Defining in this way $S^{(\infty)}$ to be those points in $S^{(n)}$ for every finite n, we can continue to apply the derive operation. Thus we get the following sets of points:

$$S^{(0)}, \ S^{(1)}, \ldots, \ S^{(\infty)}, \ S^{(\infty+1)}, \ldots, \ S^{(\infty \cdot 2)}, \ldots$$
$$S^{(\infty \cdot 4)}, \ldots, S^{(\infty^2)}, \ldots, \ S^{(\infty^\infty)}, \ldots$$

The number ∞ appears naturally in this context. So also do *numbers* $\infty + 1$, $\infty + 2$ and so on. The root of these infinite numbers was the attempt to solve a problem of analysis.

However, Cantor now devoted his time to the set theoretic aspects of his new endeavor, abandoning somewhat the underlying Fourier series problems. He first devoted his time to distinguishing the sets of rationals and reals. In 1874, he established that the set of algebraic numbers[7] can be put into one-to-one correspondence with the natural numbers.[8] But the set of real numbers cannot be put into such a correspondence. We show the simpler

Theorem. The set of rationals is one-to-one correspondence with the natural numbers.

Proof #1. Let $r_{m,n} = \frac{m}{n}$ be a rational number represented in reduced form. Define the relation
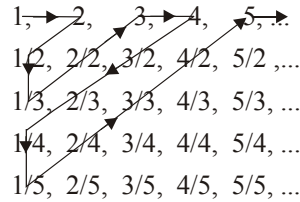
$$r_{m,n} \rightarrow 2^m \, 3^n$$

This gives the correspondence of the rationals to a subset of the natural numbers, and hence to the natural numbers. ∎

Proof #2.[9] (Arrange all the rationals in a table as shown below. Now count the

---

[7]For our context, the *algebraic* numbers are all the real zeros of finite polynomials having integer coefficients.

[8]Such sets are nowadays called *denumerable* or *countable*.

[9]This proof is the more common and simpler to understand

$$1, \; 2, \; 3, \; 4, \; 5, \; .$$
$$1/2, \; 2/2, \; 3/2, \; 4/2, \; 5/2 \; ,...$$
$$1/3, \; 2/3, \; 3/3, \; 4/3, \; 5/3, \; ...$$
$$1/4, \; 2/4, \; 3/4, \; 4/4, \; 5/4, \; ...$$
$$1/5, \; 2/5, \; 3/5, \; 4/5, \; 5/5, \; ...$$

numbers as shown by the arrows. This puts the rationals into correspondence with the natural numbers. As you may note, there is some duplication of the rationals. So to finish, simply remove the duplicates. Alternatively, build the table with the rationals already in lowest order. ∎

The proof for algebraic numbers is only slightly more complicated.

The proof of the other result, that the real numbers cannot be put into such a correspondence invoked a new and clever argument. Called *Cantor's diagonal method*, it has been successfully applied to many ends.

Theorem. The set of reals cannot be put into one-to-one correspondence with the natural numbers.

First Proof. We give here the 1891 proof. Restrict to the subset of reals in the interval $(0, 1)$. Supposing they are denumerable as the set $\{a_n\}_{n=1}^{\infty}$, we write their decimal expansions as follows:

$$
\begin{aligned}
a_1 &= 0 \,.\, d_{1,1} \, d_{1,2} \, d_{1,3} \, \ldots \\
a_2 &= 0 \,.\, d_{2,1} \, d_{2,2} \, d_{2,3} \, \ldots \\
a_3 &= 0 \,.\, d_{3,1} \, d_{3,2} \, d_{3,3} \, \ldots \\
&\;\;\vdots
\end{aligned}
$$

where the d's are digits 0 - 9. Now define the number

$$a = 0 \,.\, d_1 \, d_2 \, d_3 \, \ldots$$

by selecting $d_1 \neq d_{1,1}, \; d_2 \neq d_{2,2}, \; d_3 \neq d_{3,3}, \ldots$. This gives a number not in the set $\{a_n\}_{n=1}^{\infty}$, and the result is proved. ∎

Second Proof. This proof, which appeared in 1874, is not as well known. We show that for any sequence $v_1, v_2, \ldots$ of reals there is a

number that is not in the sequence in any interval of real numbers (a, b). First, let a_1 and b_1 be the first members of the sequence in (a, b) with $a_1 < b_1$. Let a_2 and b_2 be the first members of the sequence in (a_1, b_1) with $a_2 < b_2$, and so on. Thus $a_1, a_2, \ldots$ is an increasing sequence, and $b_1, b_2, \ldots$ is a decreasing sequence. There are three cases. If the sequences are finite, then any number inside the last chosen interval satisfies the requirement. Suppose now the sequences are infinite and they converge to limits, a_∞ and b_∞, respectively. If they are equal, then this value satisfies the requirement. If not, any value in the open interval (a_∞, b_∞) does so. ∎

Seeking undenumerable sets, Cantor considered topological notions for his derived sets. We say a set $S \subset (a, b)$ is *dense* if $S' \supset (a, b)$. We say S is *closed* if $S' \cap S = S'$. We say S is *isolated* if $S' = \emptyset$. Finally, we say S is *perfect* if $S' = S$. Remarkably, Cantor showed that perfect sets must be uncountable. One of the most famous perfect sets is so-called the *middle thirds* set defined as the residual of the open interval $(0, 1)$ by first removing the middle third (i.e. $(\frac{1}{3}, \frac{2}{3})$). Next remove the middle thirds of the two subintervals remaining and the middle thirds of the four remaining subintervals after that, and so on. This set is one of the first examples of an uncountable Lebesgue measurable set of measure zero that mathematics graduate students learn.

At this point he was in possession of two orders of infinity, countable and uncountable infinity. Being unable to determine an infinity in between, he gave a proof that every set of points on the line could be put in one-to-one correspondence with either the natural numbers or reals. His proof was incorrect, but his quest is known today and is called the *continuum hypothesis*. The problem is open today and is complicated. In 1938, Kurt Gödel proved that the continuum hypothesis cannot be disproved on the basis of the set-theoretic principles we accept today. Moreover, in 1963, Paul Cohen established that it cannot be proved within these principles. This means that the continuum hypothesis is undecidable.

Cantor was not without detractors. Though his methods were enthusiastically received by some mathematicians, his former teacher Leopold Kronecker believed that all of mathematics should be based on the natural numbers. This may be called *finitism*. He also believed that mathematics should be constructed, and this is called *constructivism*. He soundly rejected Cantor's new methods, privately and publicly. As a

journal editor, Kronecker may have delayed the publication of Cantor's work.

By 1879 Cantor was in possession of powers of infinity, defining two sets to be of the same *power* if they can be placed into one-to-one correspondence. Using his diagonalization method, he was able to demonstrate orders or powers of infinity of every order. Here is how to exhibit a set of higher power than that of the reals. Let $\mathcal{F}$ be the set of real-valued functions defined on the reals. Assume that this class of functions has the same power as the reals. Then they can be counted as $f_y(x)$, where both x and y range over the reals. Define a new function $f(x)$ such that

$$f(x) \neq f_x(x)$$

for each real x. This function cannot be in the original set $\mathcal{F}$. In turn, this method can be applied recursively to obtain higher and higher powers of infinity. There is another connection with subsets of sets. Indeed, in the argument above the subset of $\mathcal{F}$ consisting of functions assuming only the values 0 and 1 could have been used. In such a way it is possible to see that we are looking at the set of all subsets of the reals. A subset corresponding to a particular function is the set of values for which it has the value 1. Conversely, any subset generates a function according to the same rule.

In all this, infinity is now a number in its own right, though it is linked with counting ideas and relations to sets of sets. The term *power* gave us the expression power set, or set of subsets of a given set. For a finite set with n elements, the set of all subsets has size 2^n. However, the *power* of a set is an attribute of a set akin to the *cardinality* of a set. Two sets have the *same power* if they can be put in one-to-one correspondence.

In about 1882, Cantor introduced a new infinity, distinguishing cardinality from order, cardinal numbers from ordinal numbers. (i.e. one, two three from first, second, third). He would say that $(a_1, a_2, \ldots)$ and $(b_2, b_3, \ldots, b_1)$ have the same cardinality or power, but their order is different. The first has order ω while the second has order $\omega + 1$. For finite sets, there is only one order that can be given, even though elements can be transposed. Therefore, ordinal and cardinal numbers can be identified.

Using a method similar to the second proof above, Cantor showed how

to produce a set with power greater than the natural numbers, namely, the set of all ordinal numbers of the power of the natural numbers. From this, he went on to construct the power set of the set of ordinals, and so on generating higher and higher powers. Now, to make contact with the power of the real numbers, Cantor made the assumption that the reals were *well-ordered*, which is defined below. From this, he established that the power (cardinality) of the real numbers is less than, equal to or greater than each of the new powers, but not which of them it is.

Notation: By 1895 Cantor defined cardinal exponentiation. Using the term $\aleph_0$ (aleph-null) to denote the cardinality of the natural numbers, he defined $2^{\aleph_0}$ for the cardinality of the reals. With $\aleph_1$ (and more generally $\aleph_\omega$ denoting the ωth cardinal) the next larger cardinal than $\aleph_0$, the continuum hypothesis is written as $2^{\aleph_0} = \aleph_1$.

Cantor and others produced similar examples of a special category of nowhere-dense sets as an application arose of these ideas. First, a *nowhere dense* set S is a set for which the complement of its closure is dense, i.e. $\sim\bar{S}$ is dense. The set of binary fractions $\{\frac{1}{2^n}\}_1^\infty$ and the Cantor middle thirds set are nowhere dense, but the rationals are dense. The special new category consists of those that are "fat" in the following way: Every finite covering of the set by intervals should have total length greater than some given number, say 1. It becomes natural to say that such sets have *content*, and the content of the particular nowhere dense set under consideration is the infimum of the total length of all finite coverings. The idea of content was to play a major role in the development of the modern integral, notably the Jordan completion to the Riemann-Cauchy integral and ultimately the Lebesgue integral. So, we see here, sets and infinity now giving rise to new ideas for analysis. And note that the Fourier series problem that served as the root of these investigations would find its ultimate solution within the context of the modern integral.

At this point we have come full circle. The problem created the solution. In 1873, the French mathematician Paul du Bois Reymond (1831 - 1889) discovered a continuous function for which its Fourier series diverged at a single point, solving a long standing open problem. That this was the tip of the iceberg on divergence of Fourier series is illustrated below by three theorems. These results are essentially the current best possible pointwise results for Fourier series. We first need

the definition: A set $E \subset R$ is said to have *measure* zero if for every $\epsilon > 0$ there exists a finite set of intervals I_1, I_2, ..., I_k on for which

1. $E \subset \cup_{i=1}^{k} I_i$

2. $\sum_{i=1}^{k} |I_i| < \epsilon$, where for any interval I, $|I|$ is the length of I.

(Of course, k and the intervals I_1, I_2, ..., I_k depend on the set E and on ϵ.)

Theorem. (L. Carleson, *Acta Mathematica*, 116, p.135-157, 1964.) If $f(x)$ is continuous on $[0, \pi]$ (or even Riemann integrable) then $S_m(f, x) \to f(x)$ for all $t \notin E$, where E is some set of measure zero.

Here

$$S_m(f, x) = \frac{a_0}{2} + \sum_{n=1}^{m} a_n \cos nx + b_n \sin nx$$

In relation to Cantor's theorem, it is easy to show that sets E for which the kth derived set is finite must have measure zero. Corresponding to Carleson's theorem, we have the

Theorem.(Kahane and Katznelson.) If E is a set of measure zero then there exists a continuous function $f(x)$ on $[0, \pi]$ for which

$$\lim_{m \to \infty} \sup S_m(f, x) = \infty$$

for all $x \in E$.

These results compliment other counterintuitive results such as

Theorem. (A. 2N. Kolmogorov) There exists a Lebesgue integrable function $F(x)$ whose Fourier series diverges at *every* point.

See Zygmund [1959] and Katznelson [1976] for further details.

---

6   The Theory of Sets

According to Cantor, a set M is "a collection into a whole, of definite, well distinguished objects (called elements) of M of our perception

and thought." For example the numbers $\{1, 2, , \ldots, 10\}$ constitute a set. So also does the set of primes between 1 and 1000. The order of the elements of the set is unimportant. Thus, the sets $\{1, 2, , 3\}$ and $\{3, 1, , 2\}$ are the same. Therefore, two sets M and N are the same if they have the same number of elements.

This view was emphasized by Gottlob Frege (1848 - 1926), in his development of set theory, who took the approach that infinite collections cannot be counted. He sought a theory that is independent of counting. Thus, he took one-to-one correspondences to be basic, not well-orderings. Intrinsic to this is the notion of *cardinality*.

Definition. A set M is said to be *equivalent* to a set N, in symbols: $M \sim N$, if it is possible to make the elements of N correspond to the elements of M in a one-to-one manner.

This is of course an equivalence relation: (1) $M \sim M$; (2) $M \sim N$ implies $N \sim M$; (3) if $M \sim N$ and $N \sim P$ then $M \sim P$.

Definition. By a *cardinal number* of a power m we mean an arbitrary representative M of a class of mutually equivalent sets. The cardinal number of the power of a set M will also be denoted by $|M|$.

At this point we have the following cardinals:

$$0, \ 1, \ 2, \ \ldots, \aleph_0, \ \aleph_1, \ \aleph_2, \ldots$$

The latter three are called the *transfinite* cardinals. We also know how to construct more cardinals by taking the power set (the set of all subsets) of any representative of a cardinal. Note that cardinals are ordered by this

Definition. A set M is said to have a *smaller* cardinal number than a set N, in symbols: $|M| < |N|$, if and only if M is equivalent to a subset of N, but N is equivalent to no subset of M.

Of the transfinite cardinals, $\aleph_0$ is the smallest. The continuum hypothesis affirms that $2^{\aleph_0} = \aleph_1$. We have shown that the cardinality of all the functions on any interval (or uncountable set) is $\aleph_1$. Can you show that the cardinality of the *continuous* functions on any interval has cardinality $\aleph_1$? That there are infinitely many transfinite cardinals follows from an argument similar to the diagonal argument above. We include this statement as

Theorem. For every set M, the set $U(M)$ of all its subsets has a greater cardinal number than M.

By assuming there is a largest cardinal, we bump into one of the famous paradoxes of set theory, first formulated by Bertrand Russell (1872 - 1970) in about 1901. Often called Cantor's paradox it goes like this: The class of classes can be no larger than the class of individuals, since it is contained in the class of individuals. But the class of classes is the class of all subclasses of the class of individuals, and so Cantor's diagonal argument shows it to be larger than the class of individuals. Another paradox, formulated by both Russell and Ernst Zermelo (1871 - 1953), is this:

Theorem. A set M which contains each of its subsets m, m', ... as elements, is an inconsistent set. (That is, it leads to contradictions.)

Proof. Consider those subsets m which *do not* contain themselves as elements. Their totality is denoted by M_0. Since $M_0 \subset M$, we can inquire if it contains itself. If so it must be a subset of some m that does not contain itself. But $m \subset M_0$ and this implies m does contain itself, a contradiction. If not it must be a member of the original set of those subsets m which *do not* contain themselves as elements, and therefore is in M_0. ∎

Russell published several versions of this paradox. The *barber* paradox is the simplest: A barber in a certain town has stated that he will cut the hair of all those persons and only those persons in the town who do not cut their own hair. Does the barber cut his own hair? Paradoxes of this type and the above paradox of size threatened early, intuitive set theory. Note that this one does not involve infinity at all; often, it is called a *semantic* paradox.

The other axioms of set theory as given by Zermelo follow:

1. *Axiom of Extensionality:* if, for the sets M and N, $M \subset N$ and $N \subset M$, then $M = N$.

2. *Axiom of Elementary Sets:* There is a set with no elements, called the empty set, and for any objects in M, there exist sets $\{a\}$ and $\{a, b\}$.

3. *Axiom of Separation:* If a propositional function $P(x)$ is definite for a set M, then there is a set N containing precisely those

elements x of M for which $P(x)$ is true.

4. *Power Set Axiom:* If M is a set, then the power set (the set of all subsets) $U(M)$ of a set M is a set.

5. *Axiom of Union:* If M is a set, then the union of M is a set.

6. *Axiom of Choice:* If M is a disjoint union of nonempty sets, then there is a subset N of the union of M which has exactly one element in common with each member of M.

7. *Axiom of Infinity:* There is a set M containing the empty set and such that for any object x, if $x \in M$, then $\{x\} \in M$.

   Zermelo was never able to prove the consistency of the axioms and was criticized for it.[10] In 1930, he introduced a new system, now called the Zermelo-Fraenkel set theory, by including the axiom to ensure that the set

   $$\{Z,\ U(Z),\ U(U(Z)),\ \ldots\}$$

   exists, where Z is the set of natural numbers. Without that set one cannot prove the existence of $\aleph_\omega$, where, you recall, ω is the first transfinite ordinal. Fraenkel introduced the replacement axiom.

8. *Axiom of Replacement:* The range of a function of a set is itself a set.

This axiom solves the problem of ensuring the existence of $\{Z, U(Z), U(U(Z)), \ldots\}$.

Returning to the paradoxes, there were two types that threatened early intuitive set theory. First there were the paradoxes of size: (1a) We can always construct a set with larger cardinality, and (1b) if there is a largest set (i.e. the set of all sets), we can construct a larger one by the diagonalization argument.

The other type was of the Russellian type. These paradoxes occur when there is a hidden parameter whose value changes during the reasoning. For example, in a conversation between two people, one in New York and one in Texas, the first can state the time is 8:00PM and the other state the time is 7:00PM. Both are correct. So, is this a

---

[10]Indeed, Kurt Gödel (1906 - 1978) proved that the consistency could not be proved. Even further, he proved that within any system that contains the axioms for the natural numbers must have propositions that cannot be decided. He also was criticized for the axioms he chose and didn't choose.

paradox? Certainly not, everyone knows of the two time zones. This is the hidden parameter. We correct this by requiring that time zones should be specified. So, the New Yorker will say it is 8:00PM EST, and the Texan will say it is 7:00PM CST. There is now no possible confusion.

Other Russellian type paradoxes are not as easy to resolve. Consider the barber problem. Can the sentence, "there is a man who shaves all and only the men that do not shave themselves," be true? No, if the barber is in the range of the quantifier, "all the men." Yes, if not. Restricting the quantifier is then the key to resolving such paradoxes. In general, these paradoxes all have a type characterized as

$$\{x \mid x \notin x\}$$

The usual chain of reasoning shows that this set must both <u>be</u> and <u>not be</u> a member of itself; hence, a paradox. The *fix* to this state of affairs is to introduce a type of *comprehension principle*, proposed in 1930 by Zermelo, that gives the following parametric form of the Russell definition: for each set α

$$y_\alpha = \{x \in \alpha \mid x \notin x\}$$

Once this assumption is made the chain of reasoning toward paradox is blocked. You see, we begin from the outset with a given set, well founded or not. What the *faux* paradox now shows is that y_α cannot be a member of α. Moreover, from this we can conclude that there is no "universal set". If there were, then y_α would have to be in it, but at the same time cannot be in it, by the conclusion drawn about the sets y_α. In short, y_α "diagonalizes" out of α. In regards to the barber paradox, note now that the set of shavers must be defined before extracting those that don't shave themselves. The barber is diagonalized out, and the paradox is avoided.

7   The Axiom of Choice

In 1904 Zermelo first formulated the axiom of choice as such in the distinguished journal *Mathematicsche Annalen*, though it had been in use for almost twenty years. Curiously, though it has been used many times previously, it had not been formally stated as such. It was just part

of the proof of the various results that employed it. For example Cantor used it in 1887 to show any infinite set has a subset of cardinality $\aleph_0$. It was also used in topology, algebra, and analysis.[11] In 1890 Giuseppe Peano (1858 - 1932) argued that one cannot apply a law that selects a member of a class from each of many classes an *infinite* number of times. After the appearance of Zermelo's paper, the very next issue contained detractions by no less than Emile Borel (1871 - 1956) and Felix Bernstein (1878 - 1956) in *Mathematicsche Annalen*. Detractions were also submitted to the *Bulletin del la Société Mathématique de France* though out 1905 by Henri Lebesgue (1875 - 1941) and René Baire (1874 - 1932). The kernel of their argument was this: Unless a definite law specified which element was chosen from each set, no real choice has been made and the new set was not really formed. Specifically, E. Borel referred to the Axiom of Choice as a lawless choice which when used is an act of faith, and that is beyond the pale of mathematics. Defenders did not see the need for a law of choice. The choices are determined, they argued, simply because one thinks of them as determined. Jacques Hadamard (1865 - 1963) was Zermelo's staunchest supporter arguing the practicality of its application in making progress.

This is the state of affairs today. The axiom of choice is widely used and with it, wide and varied results have been obtained. It will no doubt continue to be used until such time as contradictions are obtained. However, denying its validity leads to some rather unusual consequences. For example, if one accept only the countable axiom of choice, every (constructible) set of reals is measurable. Conversely, assuming that every set is measurable leads to the denial of the axiom of choice. Similarly, if one denies the continuum hypothesis by assuming that $2^{\aleph_0} = \aleph_2$, then every set of real numbers is measurable.

The *Well Ordering Axiom*, which was used for example by Cantor to bring the reals into the realm of his other ordinals, was also widely used during this time. A set if (linearly) ordered is there is a relation "$<$" for which given a and B in M, it follows that $a < b$, $b < a$, or $a = b$ is true. The set is well ordered if every subset of M, no matter how it is selected, has a *least* element. For example the natural numbers are well ordered under the natural order, but the reals are not. The well

---

[11]For example, it was used to show the following theorem: Every bounded infinite set of real numbers has a converging subsequence.

ordering axiom states that a linear order exists for the reals. Zermelo, to answer some of the criticisms, of the axiom of choice gave a proof of the well-ordering of the reals that used the axiom of choice. Moreover, he proved the two axioms are equivalent.

Question: Can you prove that every infinite set has a countable subset without the axiom of choice? Can you prove it with the axiom of choice (or an equivalent)? Answers: no and yes. Whether you can prove it or not, the point is you probably believe it can be done, or at the very least should be possible. The axiom of choice allows this "natural fact." Call this its good side. That it has another side is indicated below.

We would be remiss to leave out one of the colossal paradoxes that can be proved using the axiom of choice. The most remarkable of these, discovered in 1924, is the **Banach-Tarski paradox**.[12] Called a paradox because of its remarkable conclusion, more properly it should be called the Banach-Tarski theorem.

Theorem. (Banach-Tarski) Given two spheres, say, one of diameter one meter (S_A) and the other the size of the earth (S_B), there are decompositions of both into a finite number of pieces, say, $\{A_i\}_{i=1}^n$ and $\{B_i\}_{i=1}^n$ with $\cup A_i = S_A$ and $\cup B_i = S_B$, for which $A_i \cong B_i$, $i = 1, \ldots, n$.

(Congruence here means that one set can be transformed to the other by a rigid rotation and a translation.) The rather technical proof is not difficult, but is a bit long for this article.

Of the more intuition shattering consequences of this counterintuitive theorem is that a bowling ball can be decomposed into a finite number of pieces and reassembled as a sphere larger than and more massive than the earth. You may argue that the principle of *conservation of mass* is violated, and therefore something is definitely amiss. However, it can be shown that these decompositions are not *measurable* in the sense of Lebesgue, and therefore they have no measurable mass. Consequently, one's intuition is shattered further after acknowledging the existence of 'clumps' of matter that cannot have mass. The axiom of choice is

---

[12]Stefan Banach (1892 - 1945) developed the theory of closed linear normed spaces, now called Banach spaces. These spaces form the foundation of much of modern analysis and numerical analysis. Alfred Tarski (1901 - 1983) was an outstanding logician that analyzed the semantic aspects of logic and was instrumental in providing a framework in which many paradoxes can be removed. The foremost of these paradoxes, *The Liar*, was known even to Aristotle and its study is an active part of contemporary modern set theory.'

a marvelous tool, and in this context, could lead to an entirely new cosmology. The only limit is one's imagination.

8 How Big is Infinity?

So far we have seen the cardinal numbers 0, 1, 2, ..., $\aleph_0, \aleph_1$, ..., $\aleph_\omega$. Of course we can construct even larger cardinals by employing the power set construction. Is that the end of the line? Constructing power set after power set. In this very short section we consider whether there may be even larger cardinals, inaccessible from power set realization. It can be shown that for each set M of cardinals there is a smallest cardinal succeeding all members of M. Denote this cardinal by $\sup M$. For example $\aleph_0 = sum\{0, 1, 2, ...\}$ and $\aleph_1 = sup\{\aleph_0\}$. A cardinal A which is not 0 is said to be *inaccessible* if

1. for every set B of cardinals such that $|B| < A$

$$sup\ B < A$$

2. if $C < A$ for $C \in A$ $2^C < A$

Certainly $\aleph_0$ is inaccessible in this definition. But are there others? In fact, it has been shown that the postulate that there are no other such inaccessible cardinals is consistent with the axioms of the standard Zermelo-Fraenkel set theory. Because of this Tarski introduced a very powerful axiom asserting the existence of inaccessible cardinals. Called the *axiom for inaccessible sets* it reads as

For every set N there is a set M with the following properties:

1. N is equivalent to a subset of M;
2. $\{A : A \subset M\ \&\ A < M\}$ is equivalent to M;
3. there is no subset $P \subset M$ with $|P| < |M|$ such that the power set of P is equivalent to M.

Tarski[13] has shown that the cardinal number of a set M is infinite and inaccessible if and only if M satisfies (2) and (3) above. This

---

[13]See Suppes or Rotman

axiom is so strong that its adoption implies that the axiom of the power set, the axiom of selection of subsets, the axiom of infinity, and the axiom of choice can all be omitted from the original system of axioms. Inaccessible cardinals are infinities beyond infinity in every sense of constructibility of the alephs. It is remarkable that language allows its description. Such sets may well be beyond the comprehension of anyone as objects of true consideration.

9   Conclusion

In some ways, the paradoxes and overall lack of agreement on basic principles in set theory can be seen as parallel to the paradoxes and overall lack of agreement on basic principles in the early days of calculus or noneuclidean geometry. Parallel to that, no doubt there were many paradoxes and overall lack of agreement of basic principles in the fledgling subject of geometry more than two thousand years earlier. It seems that by making various decisions about infinity via its "agents", the axiom of choice and the well ordering axiom, different systems of mathematics result. Therefore, the original absolute axiomatic model of Euclidean geometry within which all propositions can be resolved and that all of science has tried to emulate, is gone forever. Infinity and these trappings of set theory so very much needed to advance the early and modern mathematical theories, has served up a second dish, the demise of certainty.

Will the issues of infinity ever be resolved to the satisfaction of logicians and mathematicians? Like the limit, the understanding of which was finally assimilated after two millenia, a working definition of infinity satisfactory to all practitioners will probably percolate out. For most of us that point has already been achieved.

References

1. Barwise, Jon, and John Etchemendy, *The Liar, an Essay on Truth and Circularity*, Oxford University Press, New York, 1987.

2. Dauben, Joseph Warren, Georg Cantor, His Mathematics and Philosophy of the Infinite, Princeton University Press, Princeton, New Jersey, 1979.

3. Kamke, E. *Theory of Sets*, First American Translation, Dover, New York, 1950.

4. Katznelson, Yitzhak, *An Introduction to Harmonic Analysis*, 2nd corrected edition, Dover, New York, 1976.

5. Kline, Morris, *Mathematics, The Loss of Certainty*, Oxford University Press, New York, 1980.

6. Lavine, Shaughan, *Understanding the Infinite*, Harvard University Press, Cambridge, MA, 1994.

7. Rotman, B. and T. T. Kneebone, *The Theory of Sets and Transfinite Numbers*, Oldbourne, London, 1966.

8. Rucker, Rudy, *Infinity and the Mind*, Princeton University Press, Princeton, New Jersey, 1995.

9. Suppes, Patrick, *Axiomatic Set Theory*, van Nostrand, Princeton, 1960.

10. Zygmund, A. *Trigonometric Series*, 2nd ed., Cambridge University Press, Cambridge, 1959.