A Brief History of Infinity

The infinite has always been a slippery concept. Even the commonly accepted mathematical view, developed by Georg Cantor, may not have truly placed infinity on a rigorous foundation

by A. W. Moore

For more than two millennia, mathematicians, like most people, were unsure what to make of the infinite. Several paradoxes devised by Greek and medieval thinkers had convinced them that the infinite could not be pondered with impunity. Then, in the 1870s, the German mathematician Georg Cantor unveiled transfinite mathematics, a branch of mathematics that seemingly resolved all the puzzles the infinite had posed. In his work Cantor showed that infinite numbers existed, that they came in different sizes and that they could be used to measure the extent of infinite sets. But did he really dispel all doubt about mathematical dealings with infinity? Most people now believe he did, but I shall suggest that in fact he may have reinforced that doubt.

The hostility of mathematicians toward infinity began in the fifth century B.C., when Zeno of Elea, a student of Parmenides, formulated the well-known paradox of Achilles and the tortoise [see "Resolving Zeno's Paradoxes," by William I. McLaughlin; SCIENTIFIC AMERICAN, November 1994]. In this conundrum the swift demigod challenges the slow tortoise to a race and grants her a head start. Before he can overtake her, he must reach the point at which she began, by which time she will have advanced a little. Achilles must now make up the new distance separating them, but by the time he does so, she will have advanced again. And so on, ad infinitum. It seems that Achilles can never overtake the tortoise. In like manner Zeno argued that it is impossible to complete a racecourse. To do so, it is necessary to reach the halfway point, then the three-quarters point, then the seven-eighths point, and so on. Zeno concluded not only that motion is impossible but that we do best not to think in terms of the infinite.

The mathematician Eudoxus, similarly wary of the infinite, developed the so-called method of exhaustion to circumvent it in certain geometric contexts. Archimedes exploited that method some 100 years later to find the exact area of a circle. How did he proceed? In the box on page 114, I present not his actual derivation but a corruption of it. Part of Archimedes' own procedure was to consider the formula for the area of a polygon with n equal sides—call it P_n—inscribed inside a circle C. According to the distortion of his argument, this formula can be applied to the circle itself, which is just a polygon with infinitely many, infinitely small sides.

The perversion of Archimedes' argument has some intuitive appeal, but it would not have satisfied Archimedes. We cannot uncritically make use of the infinite as though it were just some unusually big integer. Part of what is going on here is that the larger n is, the more nearly P_n matches C. But it is also true that the larger n is, the more nearly P_n approximates a circle with a bulge—call it C^*. The key point intuitively is that C, unlike its deformed counterpart C^*, is the limit of the polygons—or what they are tending toward.

Still, it is very hard to see any way of capturing this intuition without, once again, thinking of C as an "infinigon." Archimedes provided a way. He pinpointed the crucial difference between C and C^* by proving the following point: no matter how small an area you consider, call it ε (the Greek letter epsilon), there exists an integer n that is large enough for the area of P_n to be within ε of the area of C. The same is not true of C^*. This fact, combined with a similar result for circumscribed polygons and supplemented with a refined version of the logic contained in that argument, finally enabled Archimedes to show, without ever invoking the infinite, that the area of a circle equals πr^2.

The Actual and Potential Infinite

Although Archimedes successfully ducked the infinite in this particular exercise, the Pythagoreans (a religious society founded by Pythagoras) happened on a case in which the infinite was truly inescapable. This find shattered their belief in two fundamental cosmological principles: Peras (the limit), which subsumed all that was good, and Apeiron (the unlimited or infinite), which encompassed all that was bad. They had insisted that the whole of creation could be understood in terms of, and indeed was ultimately constituted by, the positive integers, each of which is finite. This reduction was made possible, they maintained, by the fact that Peras was ever subjugating Apeiron.

Pythagoras had discovered, however, that the square of the hypotenuse (the longest side) of a right-angled triangle is equal to the sum of the squares of the other two sides. Given this theorem, the ratio of a square's diagonal to each side is $\sqrt{2}$ to 1, since $1^2 + 1^2 = (\sqrt{2})^2$. Were Peras impervious, this ratio should be expressible in the form p to q, where p and q are both positive integers. Yet this is impossible. Imagine two positive integers, p and q, such that the ratio of p to q, or p divided by q, is equivalent to $\sqrt{2}$. We can assume that p and q have no common factor greater than 1 (we could, if necessary,

A. W. MOORE is a tutorial fellow in philosophy at St. Hugh's College of the University of Oxford. He studied for his Ph.D. in the philosophy of language at Balliol College at Oxford. His main academic interests are in logic, metaphysics and the philosophies of Immanuel Kant and Ludwig Wittgenstein, all of which have informed his work on the infinite. He is at work on a book about the metaphysics of objectivity and subjectivity.

UNLIMITED EXPANSE is conveyed by the baked floor of California's Death Valley. To grasp the infinite, mathematicians have confronted several paradoxes, concluding that infinities come in different types and that some are bigger than others.

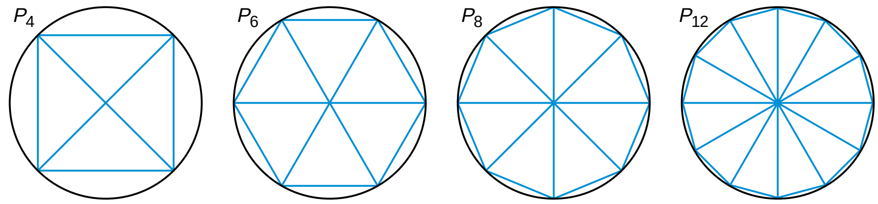
GARY YEOWELL *Tony Stone Images*

divide by that factor). Now, p^2 is twice q^2. So p^2 is even, which means that p itself is even. Hence, q must be odd, otherwise 2 would be a common factor. But consider: if p is even, there must be a positive integer r that is exactly half of p. Therefore, $(2r)^2$ equals $2q^2$, or $2r^2$ equals q^2, which means that q^2 is even, and so q itself is also even, contrary to what was proved above.

For the Pythagoreans, this result was nothing short of catastrophic. (According to legend, one of them was shipwrecked at sea for revealing the discovery to their enemies.) They had come on an "irrational" number. In doing so, they had seen the limitations of the positive integers, and they had been forced

Archimedes and the Area of a Circle

How did Archimedes use the method of exhaustion to find the area of a circle? Here is the corruption of his argument. Imagine a circle C that has a radius r. For each integer n greater than 2, we can construct a regular polygon with n sides and inscribe it inside C. This n-sided polygon—call it P_n—can be divided into n congruent triangles. Label the base of each triangle b_n and its height h_n. Then the area of each triangle is $1/2\ b_n h_n$. Thus, the area of P_n as a whole is $n(1/2\ b_n h_n)$, or $1/2\ nb_n h_n$. But C itself is a polygon with infinitely many, infinitely small sides. In other words, C results when we extend the original definition of P_n and allow n to be infinite. In this case, nb_n is the circumference of C, which equals $2\pi r$ (which follows from the definition of π), and h_n is the radius r. So the area of C is $1/2(2\pi rr)$, or simply πr^2.



JARED SCHNEIDMAN DESIGN

to acknowledge the presence of the infinite in their very midst. Indeed, a modern mathematician would say that $\sqrt{2}$ is a kind of "infinite object." Not only is its decimal expansion infinite, but this expansion never adopts a recurring finite pattern.

In the fourth century B.C. Aristotle recognized a more general problem. On the one hand, we are under pressure to acknowledge the infinite. Quite apart from what we may have to say about $\sqrt{2}$, time appears to continue indefinitely, numbers seem to go on endlessly, and space, time and matter seem to be forever divisible. On the other hand, we are under pressure from various sources, including Zeno's paradoxes, to repudiate the infinite.

Aristotle's solution to this dilemma was masterful. He distinguished between two different kinds of infinity. The actual infinite is that whose infinitude exists at some point in time. In contrast, the potential infinite is that whose infinitude is spread over time. All the objections to the infinite, Aristotle insisted, are objections to the actual infinite. The potential infinite, on the other hand, is a fundamental feature of reality. It deserves recognition in any process that can never end, including counting, the division of matter and the passage of time itself. This distinction between the two types of infinity provided a solution to Zeno's paradoxes. Traversing a region of space does not involve moving across an actual infinity of subregions, which would be impossible. But it does mean crossing a potential infinity of subregions, in the sense that there can be no end to the process of dividing the space. This conclusion, fortunately, is harmless.

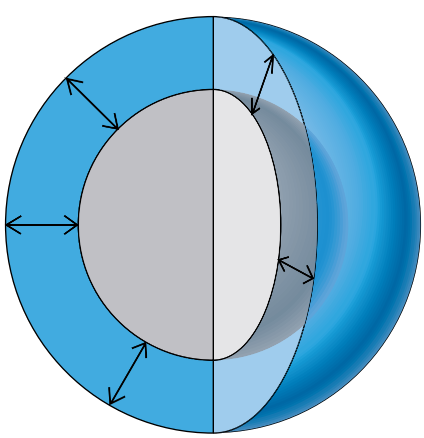
Aristotle's parting of the actual and the potential infinite long stood as orthodoxy. Nevertheless, scholars usually interpreted his reference to time as a metaphor for something deeper and more abstract. Existing "in time" or existing "all at once" assumed much broader meanings. To take exception to the actual infinite was to object to the very idea that some entity could have a property that surpassed all finite measure. It was also to deny that the infinite was itself a legitimate object of study.

Some 2,000 years later the infinite, both actual and potential, exercised mathematicians once more as they developed the calculus. Early work on the calculus, ushered in by Isaac Newton and Gottfried Wilhelm Leibniz in the late 17th century, fell far short of Greek standards of rigor. Indeed, mathematicians had made extensive, uncritical use of infinitesimals, items taken to be too small for measure. Sometimes these quantities were considered equal to zero. For example, when they were added to another number, the value of the original number remained the same. At other times, they were taken to be different from zero and used in division. Guillaume François Antoine de l'Hôpital wrote: "A curve may be regarded as the totality of an infinity of straight segments, each infinitely small: or ... as a polygon with an infinite number of sides." Only in the 19th century did French mathematician Augustin-Louis Cauchy and German mathematician Karl Weierstrass resuscitate the method of exhaustion and give the calculus a secure foundation.

The Infinite and Equinumerosity

As a result of Cauchy's and Weierstrass's work, most mathematicians felt less threatened by Zeno's paradoxes. Of more concern by then was a family of paradoxes born in the Middle Ages dealing with equinumerosity. These puzzles derive from the principle that if it is possible to pair off all the members of one set with all those of another, the two sets must have equally many members. For example, in a nonpolygamous society there must be just as many husbands as wives. This principle looks incontestable. Applied to infinite sets, however, it seems to flout a basic notion first articulated by Euclid: the whole is always greater than any of its parts. For instance, it is possible to pair off all the positive integers with those that are even: 1 with 2, 2 with 4, 3 with 6 and so on—despite the fact that positive integers also include odd numbers.

The medievals proffered many simi-



JARED SCHNEIDMAN DESIGN

1	2	3	4	. . .	n	. . .
↕	↕	↕	↕		↕	
1	4	9	16	. . .	n^2	. . .

SETS ARE THE SAME SIZE if all their members can be paired with one another. But this principle seems to be violated in infinite sets. All the squared integers can be matched with every single positive integer (*above*), even though the set of squares seems smaller. Similarly, all the points on the smaller sphere can be paired off with those on the larger one (*left*).

SUBSETS	POSITIVE INTEGERS 1	2	3	4	5 . . .
EVEN INTEGERS	NO	YES	NO	YES	NO
ODD INTEGERS	YES	NO	YES	NO	YES
PRIMES	NO	YES	YES	NO	YES
SQUARES	YES	NO	NO	YES	NO
MULTIPLES OF 4	NO	NO	NO	YES	NO . . .
⋮					
DIAGONALIZED SET	YES	YES	NO	NO	YES . . .

JARED SCHNEIDMAN DESIGN

CANTOR'S THEOREM—that no set has as many members as it has subsets—is proved by diagonalization, which creates an extra subset. Each subset of the set of positive integers is represented as a series of yeses and noes. A yes indicates that the integer belongs to the subset; a no that it does not. Replacing each yes with a no, and vice versa, down the diagonal (*shaded area*) creates another subset.

lar examples, some of which were geometric. In the 13th century the Scottish mathematician John Duns Scotus puzzled over the case of two concentric circles: all the points on the shorter circumference of the smaller circle can be paired off with all the points on the longer circumference of the bigger circle. The same result applies to two spheres [*see bottom illustration on opposite page*]. Some 350 years later Galileo discussed a variation of the pairing example of the even integers, based instead on squared integers. Particularly striking is the fact that as increasingly larger segments of the sequence of positive integers are considered, the proportion of these integers that are squares tends toward zero. Nevertheless, the pairing still proceeds indefinitely.

It is certainly tempting, in view of these difficulties, to eschew infinite sets entirely. More generally, it is tempting to deny, as did Aristotle, that infinitely many things can be gathered together all at once. Eventually, though, Cantor challenged the Aristotelian view. In work of great brilliance he took the paradoxes in his stride and formulated a coherent, systematic and precise theory of the actual infinite, ready for any skeptical gaze. Cantor accepted the "pairing off" principle and its converse, namely, that no two sets are equinumerous unless their members can be paired off. Accordingly, he accepted that there are just as many even positive integers as there are positive integers altogether (and likewise in the other paradoxical cases).

Let us for the sake of argument, and contemporary mathematical convention for that matter, follow suit. If this principle means that the whole is no greater than its parts, so be it. We can in fact use this idea to define the infinite, at least in its application to sets: a set is infinite if it is no bigger than one of its parts. More precisely, a set is infinite if it has as many members as does one of its proper subsets.

What remains an open question, once things have been clarified in this way, is whether all infinite sets are equinumerous. Much of the impact of Cantor's work came in his demonstration that they are not. There are different infinite sizes. This proposition results from what is known as Cantor's theorem: no set, and in particular no infinite set, has as many members as it has subsets. In other words, no set is as big as the set of its subsets. Why not? Because if a set were, it would be possible to pair off all its members with all its subsets. Some members would then be paired off with subsets that contained them, others not. So what of the set of those members that were not included in the set with which they had been paired? No member could be paired off with this subset without contradiction.

The argument can be recast in a diagram [*see illustration above*]. For convenience, I will focus on the set of positive integers. I can represent any subset of the set of positive integers by an infinite sequence of yeses and noes, registering whether successive positive integers do or do not belong to the set. For example, the set of even integers can be represented by the sequence <no, yes, no, yes, no…>, corresponding to 1, 2, 3, 4, 5, and so forth. We can do the same for the set of odd integers <yes, no, yes, no, yes…>, the set of prime numbers <no, yes, yes, no, yes…> and the set of squares <yes, no, no, yes, no…>. Generally, then, any assignment of different subsets to individual positive integers (such as the purely arbitrary example illustrated) can be represented as an infinite square of yeses and noes.

To show that at least one subset is nowhere on this list of subsets, we make a new subset by moving down the "square's diagonal," replacing each yes with a no, and vice versa. In the case illustrated, we write <yes, yes, no, no…>. What results represents the subset in question. For by construction it differs from the first subset listed with respect to whether 1 belongs to it, from the second with respect to whether 2 belongs to it, from the third with respect to whether 3 belongs to it, and so on. There is a pleasant historical quirk here: just as studying a diagonal had led the Pythagoreans to acknowledge an infinitude beyond the grasp of the positive integers, the same was true in a different way in Cantor's case.

Cantor later devised infinite cardinals—numbers that can be used to measure the size of infinite sets. He invented a kind of arithmetic for them as well. Having defined his terms, he explored what happens when one infinite cardinal is added to another, when it is multiplied by another, when it is raised to a power, and so forth. His work showed mathematical craftsmanship of the highest caliber. But even in his own terms, difficulties remained. The continuum problem is perhaps the best known of these troubles. The set of positive integers, we have seen, is smaller than the set of sets of positive integers. But how much smaller? Specifically, are there any sets of intermediate size?

Cantor's Continuum Hypothesis

Cantor's own hypothesis, his famous "continuum hypothesis," was that there are not. But he never successfully proved this idea, nor did he disprove it. Subsequent work has shown that the situation is far graver than he had imagined. Using all the accepted methods of modern mathematics, the issue cannot be settled. This problem raises philosophical questions about the determinacy of Cantor's conception. Asking whether the continuum hypothesis is true may be like asking whether Hamlet was left-handed. It may be that not enough is known to form an answer. If so, then we should rethink how well Cantor's work tames the actual infinite.

Of even more significance are questions surrounding the set of all sets. Given Cantor's theorem, this collection must be smaller than the set of sets of sets. But wait! Sets of sets are themselves sets, so it follows that the set of sets must be smaller than one of its own proper subsets. That, however, is

Diagonalization and Gödel's Theorem

The diagonalization used in establishing Cantor's theorem also lies at the heart of Austrian mathematician Kurt Gödel's celebrated 1931 theorem. Seeing how offers a particularly perspicuous view of Gödel's result.

Gödel's theorem deals with formal systems of arithmetic. By arithmetic I mean the theory of positive integers and the basic operations that apply to them, such as addition and multiplication. The theorem states that no single system of laws (axioms and rules) can be strong enough to prove all true statements of arithmetic without at the same time being so strong that it "proves" false ones, too. Equivalently, there is no single algorithm for distinguishing true arithmetical statements from false ones. Two definitions and two lemmas, or propositions, are needed to prove Gödel's theorem. Proof of the lemmas is not possible within these confines, although each is fairly plausible.

Definition 1: A set of positive integers is arithmetically definable if it can be defined using standard arithmetical terminology. Examples are the set of squares, the set of primes and the set of positive integers less than, say, 821.

Definition 2: A set of positive integers is decidable if there is an algorithm for determining whether any given positive integer belongs to the set. The same three sets above serve as examples.

Lemma 1: There is an algorithmic way of pairing off positive integers with arithmetically definable sets.

Lemma 2: Every decidable set is arithmetically definable.

Given lemma 1, diagonalization yields a set of positive integers that is not arithmetically definable. Call this set *D*. Now suppose, contrary to Gödel's theorem, there is an algorithm for distinguishing between true arithmetical statements and false ones. Then *D*, by virtue of its construction, is decidable. But given lemma 2, this proposition contradicts the fact that *D* is not arithmetically definable. So Gödel's theorem must hold after all. Q.E.D.

Kurt Gödel, 1906–78

courtesy of INSTITUTE OF ADVANCED STUDY ARCHIVES

impossible. The whole can be the same size as the part, but it cannot be smaller. How did Cantor escape this trap? With wonderful pertinacity, he denied that there is any such thing as the set of sets. His reason lay in the following picture of what sets are like. There are things that are not sets, then there are sets of all these things, then there are sets of all those things, and so on, without end. Each set belongs to some further set, but there never comes a set to which every set belongs.

Cantor's reasoning might seem somewhat ad hoc. But an argument of the sort is required, as revealed by Bertrand Russell's memorable paradox, discovered in 1901. This paradox concerns the set of all sets that do not belong to themselves. Call this set *R*. The set of mice, for example, is a member of *R;* it does not belong to itself because it is a set, not a mouse. Russell's paradox turns on whether *R* can belong to itself. If it does, by definition it does not belong to *R*. If it does not, it satisfies the condition for membership to *R* and so does belong to it. In any view of sets, there is something dubious about *R*. In Cantor's view, according to which no set belongs to itself, *R*, if it existed, would be the set of all sets. This argument makes Cantor's picture, and the rejection of *R* that goes with it, appear more reasonable.

But is the picture not strikingly Aristotelian? Notice the temporal metaphor. Sets are depicted as coming into being "after" their members—in such a way that there are always more to come. Their collective infinitude, as opposed to the infinitude of any one of them, is potential, not actual. Moreover, is it not this collective infinitude that has best claim to the title? People do ordinarily define the infinite as that which is endless, unlimited, unsurveyable and immeasurable. Few would admit that the technical definition of an infinite set expresses their intuitive understanding of the concept. But given Cantor's picture, endlessness, unlimitedness, unsurveyability and immeasurability more properly apply to the entire hierarchy than to any of the particular sets within it.

In some ways, then, Cantor showed that the set of positive integers, for example, is "really" finite and that what is "really" infinite is something way beyond that. (He himself was not averse to talking in these terms.) Ironically, his work seems to have lent considerable substance to the Aristotelian orthodoxy that "real" infinitude can never be actual.

Some scholars have objected to my suggestion that, in Cantor's conception, the set of positive integers is "really" finite. They complain that this assertion is at variance not only with standard mathematical terminology but also, contrary to what I seem to be suggesting, with what most people would say.

Well, certainly most people would say the set of positive integers is "really" infinite. But, then again, most people are unaware of Cantor's results. They would also deny that one infinite set can be bigger than another. My point is not about what most people would say but rather about how they understand their terms—and how that understanding is best able, for any given purpose, to absorb the shock of Cantor's results. Nothing here is forced on us. We could say some infinite sets are bigger than others. We could say the set of positive integers is only finite. We could hold back from saying either and deny that the set of positive integers exists.

If the task at hand is to articulate certain standard mathematical results, I would not advocate using anything other than standard mathematical terminology. But I would urge mathematicians and other scientists to use more caution than usual when assessing how Cantor's results bear on traditional conceptions of infinity. The truly infinite, it seems, remains well beyond our grasp.

FURTHER READING

INFINITY AND THE MIND: THE SCIENCE AND PHILOSOPHY OF THE INFINITE. Rudy Rucker. Harvester, 1982.

TO INFINITY AND BEYOND: A CULTURAL HISTORY OF THE INFINITE. Eli Maor. Birkhauser, 1986.

THE INFINITE. A. W. Moore. Routledge, 1990.

INFINITY. Edited by A. W. Moore. Dartmouth, 1993.

UNDERSTANDING THE INFINITE. Shaughan Lavine. Harvard University Press, 1994.