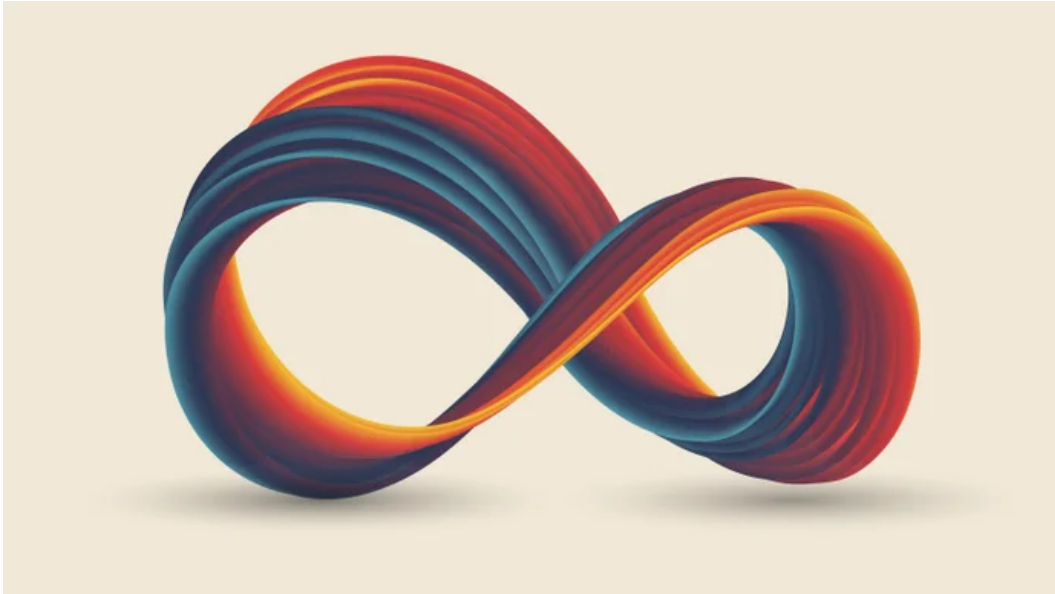


# A Deep Math Dive into Why Some Infinities Are Bigger Than Others

The size of certain infinite sets has been a mystery. Now, it turns out, each one is different than the next, and they can all be ordered by size

By Martin Goldstern, Jakob Kellner on August 16, 2021



Credit: Lidiia Moor *Getty Images*

Simple mathematical concepts such as counting appear to be firmly anchored in the natural process of thinking. Studies have shown that even very young children and animals possess such skills to a certain extent. This is hardly surprising because counting is extremely useful in terms of evolution. For example, it is required for even very simple forms of trading. And counting helps in estimating the size of a hostile group and, accordingly, whether it is better to attack or retreat.

Over the past millennia, humans have developed a remarkable notion of counting. Originally applied to a handful of objects, it was easily extended to vastly different orders of magnitude. Soon a mathematical framework emerged that could be used to describe huge quantities, such as the distance between galaxies or the number of elementary particles in the universe, as well as barely conceivable distances in the microcosm, between atoms or quarks.

We can even work with numbers that go beyond anything currently known to be relevant in describing the universe. For example, the number  $10^{10^{100}}$  (one followed by  $10^{100}$  zeros, with  $10^{100}$  representing one followed by 100 zeros) can be written down and used in all kinds of calculations. Writing this number in ordinary decimal notation, however, would require more elementary particles than are probably contained in the universe, even employing just one particle per digit. Physicists estimate that our cosmos contains fewer than  $10^{100}$  particles.

Yet even such unimaginably large numbers are vanishingly small, compared with infinite sets, which have played an important role in mathematics for more than 100 years. Simply counting objects gives rise to the set of natural numbers,  $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ , which many of us

encounter in school. Yet even this seemingly simple concept poses a challenge: there is no largest natural number. If you keep counting, you will always be able to find a larger number.

Can there actually be such a thing as an infinite set? In the 19th century, this question was very controversial. In philosophy, this may still be the case. But in modern mathematics, the existence of infinite sets is simply assumed to be true—postulated as an axiom that does not require proof.

Set theory is about more than describing sets. Just as, in arithmetic, you learn to apply arithmetical operations to numbers—for example, addition or multiplication—you can also define set-theoretical operations that generate new sets from given ones. You can take unions— $\{1, 2\}$  and  $\{2, 3, 4\}$  becomes  $\{1, 2, 3, 4\}$ —or intersections— $\{1, 2\}$  and  $\{2, 3, 4\}$  becomes  $\{2\}$ . More excitingly, you can form power sets—the family of all subsets of a set.

## COMPARING SET SIZES

---

The power set  $P(X)$  of a set  $X$  can be easily calculated for small  $X$ . For instance,  $\{1, 2\}$  gives you  $P(\{1, 2\}) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$ . But  $P(X)$  grows rapidly for larger  $X$ . For example, every 10-element set has  $2^{10} = 1,024$  subsets. If you really want to challenge your imagination, try forming the power set of an infinite set. For example, the power set of the natural numbers,  $P(\mathbb{N})$ , contains the empty set,  $\mathbb{N}$  itself, the set of all even numbers, the prime numbers, the set of all numbers with the sum of digits totaling 2021,  $\{12, 17\}$ , and much, much more. As it turns out, the number of elements of this power set exceeds the number of elements in the set of natural numbers.

To understand what that means, you first have to understand how the size of sets is defined. For the finite case, you can count the respective elements. For instance,  $\{1, 2, 3\}$  and  $\{\text{Cantor}, \text{Gödel}, \text{Cohen}\}$  are of the same size. If you wish to compare sets with numerous (but finitely many) elements, there are two well-established methods. One possibility is to count the objects contained in each set and compare the numbers. Sometimes, however, it is easier to match the elements of one set to another. Then two sets are of the same size if and only if each element of one set can be uniquely paired with an element of the other set (in our example:  $1 \rightarrow \text{Cantor}$ ,  $2 \rightarrow \text{Gödel}$ ,  $3 \rightarrow \text{Cohen}$ ).

This pairing method also works for infinite sets. Here, instead of first counting and then deriving concepts such as “greater than” or “equal to,” you follow a reverse strategy. You start with defining what it means that two sets,  $A$  and  $B$ , are of the same size—namely, there is a mapping that pairs each element of  $A$  with exactly one element of  $B$  (so that no element of  $B$  is left over). Such a mapping is called bijection.

Similarly,  $A$  is defined to be less than or equal to  $B$  if there is a mapping from  $A$  to  $B$  that uses each element of  $B$  once at most.

After we have these notions, the size of sets is denoted by cardinal numbers, or cardinals. For finite sets, these are the usual natural numbers. But for infinite sets, they are abstract quantities that just capture the notion of “size.” For example, “countable” is the cardinal

number of the natural numbers (and therefore of every set that has the same size as the natural numbers). It turns out that there are different cardinals. That is, there are infinite sets  $A$  and  $B$  with no bijection between them.

At first sight, this definition of size seems to lead to contradictions, which were elaborated by the Bohemian mathematician Bernard Bolzano in *Paradoxes of the Infinite*, published posthumously in 1851. For example, Euclid's "The whole is greater than the part" appears self-evident. That means if a set  $A$  is a proper subset of  $B$  (that is, every element of  $A$  is in  $B$ , but  $B$  contains additional elements), then  $A$  must be smaller than  $B$ . This assertion is not true for infinite sets, however! This curious property is one reason some scholars rejected the concept of infinite sets more than 100 years ago.

For example, the set of even numbers  $E = \{0, 2, 4, 6, \dots\}$  is a proper subset of the natural numbers  $\mathbb{N} = \{0, 1, 2, \dots\}$ . Intuitively, you might think that the set  $E$  is half the size of  $\mathbb{N}$ . But in fact, based on our definition, the sets have the same size because each number  $n$  in  $E$  can be assigned to exactly one number in  $\mathbb{N}$  ( $0 \rightarrow 0, 2 \rightarrow 1, 4 \rightarrow 2, \dots, n \rightarrow n/2, \dots$ ).

Consequently, the concept of "size" for sets could be dismissed as nonsensical. Alternatively, it could be termed something else: *cardinality*, for example. For the sake of simplicity, we will stick to the conventional terminology, even though it has unexpected consequences at infinity.

In the late 1800s, German logician Georg Cantor, founder of modern set theory, discovered that not all infinite sets are equal. According to his proof, the power set  $P(X)$  of a (finite or infinite) set  $X$  is always larger than  $X$  itself. Among other things, it follows that there is no largest infinity and thus no "set of all sets."

## AN UNRESOLVED HYPOTHESIS

---

There is, however, something akin to a smallest infinity: all infinite sets are greater than or equal to the natural numbers. Sets  $X$  that have the same size as  $\mathbb{N}$  (with a bijection between  $\mathbb{N}$  and  $X$ ) are called countable; their cardinality is denoted  $\aleph_0$ , or aleph null. For every infinite cardinal  $\aleph_\alpha$ , there is a next larger cardinal number  $\aleph_{\alpha+1}$ . Thus, the smallest infinite cardinal  $\aleph_0$  is followed by  $\aleph_1$ , then  $\aleph_2$  and so on. The set  $\mathbb{R}$  of real numbers (also called the real line) is as large as the power set of  $\mathbb{N}$ , and this cardinality is denoted  $2^{\aleph_0}$ , or "continuum."

In the 1870s, Cantor ruminated over whether the size of  $\mathbb{R}$  was the smallest possible cardinal above  $\aleph_0$ —in other words, whether  $\aleph_1 = 2^{\aleph_0}$ . Previously, every infinite subset of  $\mathbb{R}$  that had been studied had turned out to be either as large as  $\mathbb{N}$  or  $\mathbb{R}$  itself. This led Cantor to what is known as the continuum hypothesis (CH): the assertion that the size of  $\mathbb{R}$  is the smallest possible uncountable cardinal. For decades, CH kept mathematicians busy, but a proof eluded them. Later, it became clear their efforts had been doomed from the start.

Set theory is extremely powerful. It can describe virtually all mathematical concepts. But it also has limitations. The field is based on the axiomatic system formulated more than 100 years ago by German logician Ernst Zermelo and elaborated by his German-Israeli colleague

Abraham Fraenkel. Called ZFC, or Zermelo-Fraenkel set theory (C stands for “axiom of choice”), the system is a collection of basic assumptions sufficient to carry out almost all of mathematics. Very few problems require additional assumptions. But in 1931 Austrian mathematician Kurt Gödel recognized that the system has a fundamental defect: it is incomplete. That is, it is possible to formulate mathematical statements that can neither be refuted nor proved using ZFC. Among other things, it is impossible for a system to prove its own consistency.

The most famous example of undecidability in set theory is CH. In a paper published in 1938, Gödel proved that CH cannot be disproved within ZFC. Neither can it be proved, as Paul Cohen showed 25 years later. It is thus impossible to solve CH using the usual axioms of set theory. Consequently, it remains unclear whether sets exist that are both larger than the natural numbers and smaller than the real numbers.

Cardinality is not the only notion to describe the size of a set. For example, from the point of view of geometry, subsets of the real line  $\mathbb{R}$ , the two-dimensional plane (sometimes called the  $x$ - $y$  plane) or the three-dimensional space can be assigned length, area or volume. A set of points in the plane forming a rectangle with side lengths  $a$  and  $b$  has an area of  $a \cdot b$ . Calculating the area of more complicated subsets of the plane sometimes requires other tools, such as the integral calculus taught in school. This method does not suffice for certain complex sets. But many can still be quantified using the Lebesgue measure, a function that assigns length, area or volume to extremely complicated objects. Even so, it is possible to define subsets of  $\mathbb{R}$ , or the plane, that are so frayed that they cannot be measured at all.

In two-dimensional space, a line (such as the circumference of a circle, a finite segment or a straight line) is always measurable, and its area is zero. It is therefore called a null set. Null sets can also be defined in one dimension. On the real line, the set with two elements—for example  $\{3, 5\}$ —has a measure zero, whereas an interval such as  $[3, 5]$ —that is, the real numbers between three and five—has a measure two.

## NEGLECTIBLE SETS

---

The concept of a null set is extremely useful in mathematics. Often, a theorem is not true for all real numbers but can be proved for all real numbers outside of a null set. This is usually good enough for most applications. Yet null sets may seem quite large. For example, the rational numbers within the real line are a null set even though there are infinitely many of them. This is because any countable—or finite—set is a null set. The converse is not true: a subset of the  $x$ - $y$  plane with a large cardinality need be neither measurable nor of large measure. For example, the entire plane with its  $2^{\aleph_0}$  elements has an infinite measure. But the  $x$  axis with the same cardinality has a two-dimensional measure (or “area”) zero and thus is a null set of the plane.

Such “negligible” sets led to fundamental questions about the size of  $\aleph_1$  infinite cardinals, which remained unanswered for a long time. For example, mathematicians wished to know the minimum size a set must have for it not to be a null set. The family of all null sets is denoted by  $\mathcal{N}$ , and the smallest cardinality of a non-null set is denoted by  $\text{non}(\mathcal{N})$ . It follows

that  $\aleph_0 < \text{non}(\mathcal{N}) \leq 2^{\aleph_0}$ , because any set of size  $\aleph_0$  is a null set, and the whole plane has size  $2^{\aleph_0}$  and is not a null set. Thus,  $\aleph_1 \leq \text{non}(\mathcal{N}) \leq 2^{\aleph_0}$ , because  $\aleph_1$  is the smallest uncountable cardinal. If we assume CH, then  $\text{non}(\mathcal{N}) = 2^{\aleph_0}$ , because, in that case,  $\aleph_1 = 2^{\aleph_0}$ .

We can define another cardinal number,  $\text{add}(\mathcal{N})$ , to answer the question, What is the minimal number of null sets whose union is a non-null set? This number is less than or equal to  $\text{non}(\mathcal{N})$ : if  $A$  is a non-null set containing  $\text{non}(\mathcal{N})$  many elements, the union of all the  $\text{non}(\mathcal{N})$  many one-element subsets of  $A$  is the non-null set  $A$ . But a smaller number of null sets (though they would not be one-element sets) could also satisfy the requirements. Therefore,  $\text{add}(\mathcal{N}) \leq \text{non}(\mathcal{N})$  holds.

The cardinal  $\text{cov}(\mathcal{N})$  is the smallest number of null sets whose union yields the whole plane. It is also easy to see that  $\text{add}(\mathcal{N})$  is smaller than or equal to  $\text{cov}(\mathcal{N})$  because, as already mentioned, the plane is a non-null set.

We can also consider  $\text{cof}(\mathcal{N})$ , the smallest possible size for a basis  $X$  of  $\mathcal{N}$ . That is, a set  $X$  of null sets that contains a superset  $B$  of every null set  $A$ . (That means  $A$  is a subset of  $B$ .) These infinite cardinals— $\text{add}(\mathcal{N})$ ,  $\text{cov}(\mathcal{N})$ ,  $\text{non}(\mathcal{N})$  and  $\text{cof}(\mathcal{N})$ —are important characteristics of the family of null sets.

For each of these four cardinal characteristics, an analogous characteristic can be defined using a different concept of small, or negligible, sets. This other notion of smallness is “meager.” A meager set is a set contained in the countable union of nowhere dense sets, such as the circumference of a circle in the plane, or finitely or countably many such circumferences. In one dimension, the normal numbers form a meager set on the real line, while the remaining reals, the non-normal numbers, constitute a null set.

Accordingly, the corresponding cardinal characteristics can be defined for the family of meager sets:  $\text{add}(\mathcal{M})$ ,  $\text{non}(\mathcal{M})$ ,  $\text{cov}(\mathcal{M})$  and  $\text{cof}(\mathcal{M})$ . Under CH, all characteristics are the same, namely  $\aleph_1$ , for both null and meager sets. On the other hand, using the method of “forcing,” developed by Cohen, mathematicians Kenneth Kunen and Arnold Miller were able to show in 1981 that it is impossible to prove the statement  $\text{add}(\mathcal{N}) = \text{add}(\mathcal{M})$  within ZFC. In other words, the numbers of null and meager sets that must be combined to produce a non-negligible set are not provably equal.

Forcing is a method to construct mathematical universes. A mathematical universe is a model that satisfies the ZFC axioms. To show that a statement  $X$  is not refutable in ZFC, it is enough to find a universe in which both ZFC and  $X$  are valid. Similarly, to show that  $X$  is not provable from ZFC, it is enough to find a universe where ZFC holds but  $X$  fails.

## MATHEMATICAL UNIVERSSES WITH SURPRISING PROPERTIES

Kunen and Miller used this method to construct a mathematical universe that satisfies  $\text{add}(\mathcal{N}) < \text{add}(\mathcal{M})$ . In this model, more meager than null sets are required to form a non-negligible set. Accordingly, it is impossible to prove  $\text{add}(\mathcal{N}) = \text{add}(\mathcal{M})$  from ZFC.

In contrast, Tomek Bartoszyński discovered three years later that the converse inequality  $\text{add}(\mathcal{N}) \leq \text{add}(\mathcal{M})$  can be proved using ZFC. This points to an asymmetry between the two notions of smallness. Let us note that this asymmetry is not visible if we assume CH because CH implies  $\aleph_1 = \text{add}(\mathcal{N}) = \text{add}(\mathcal{M})$ .

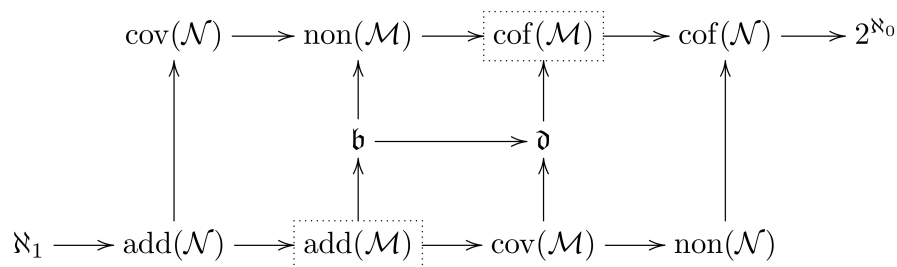
To summarize:  $\text{add}(\mathcal{N}) \leq \text{add}(\mathcal{M})$  is provable, but neither  $\text{add}(\mathcal{N}) = \text{add}(\mathcal{M})$  nor  $\text{add}(\mathcal{N}) < \text{add}(\mathcal{M})$  is provable. This is the same effect as with CH: it is trivial to prove that  $\aleph_1 \leq 2^{\aleph_0}$ , but neither  $\aleph_1 < 2^{\aleph_0}$  nor  $\aleph_1 = 2^{\aleph_0}$  is provable.

In addition to the cardinal numbers defined so far, there are two important cardinal characteristics— $\mathfrak{b}$  and  $\mathfrak{d}$ —that refer to dominating functions of real numbers. For two continuous functions (of which there are  $2^{\aleph_0}$  many)  $f$  and  $g$ ,  $f$  is said to be dominated by  $g$  if the inequality  $f(x) < g(x)$  holds for all sufficiently large  $x$ . For example, a quadratic function such as  $g(x) = x^2$  always dominates a linear function, say  $f(x) = 100x + 30$ .

The cardinal number  $\mathfrak{d}$  is defined as the smallest possible size of a set of continuous functions sufficient to dominate every possible continuous function.

A variant of this definition gives the cardinal number  $\mathfrak{b}$ , namely the smallest size of a family  $B$  with the property that there is no continuous function that dominates all functions of  $B$ . It can be shown that  $\aleph_1 \leq \mathfrak{b} \leq \mathfrak{d} \leq 2^{\aleph_0}$  holds.

Several additional inequalities have been shown to hold between the 12 infinite cardinals we just defined. All these inequalities are summarized in Cichoń’s diagram, introduced by British mathematician David Fremlin in 1984 and named after his Polish colleague Jacek Cichoń. For typographical reasons, the less-or-equal signs are replaced by arrows.



Credit: Jakob Kellner

There are two additional relations:  $\text{Add}(\mathcal{M})$  is the smaller one of  $\mathfrak{b}$  and  $\text{cov}(\mathcal{M})$ . Likewise,  $\text{cof}(\mathcal{M})$  is the larger of  $\mathfrak{d}$  and  $\text{non}(\mathcal{M})$ . These two “dependent” cardinals are marked with a frame in the Cichoń diagram. The diagram thus comprises 12 uncountable cardinalities of which no more than 10 can be simultaneously different.

## HOW DIFFERENT CAN INFINITIES BE?

If CH holds, however,  $\aleph_1$  (the smallest number in the diagram) is equal to  $2^{\aleph_0}$  (the largest number in the diagram), and thus all entries are equal. If, on the other hand, we assume CH

to be false, then they could be quite different.

For several decades, mathematicians tried to show that none of the less-or-equal relations in Cichoń's diagram can be strengthened to equalities. To do that, they constructed many different universes in which they assigned the two smallest uncountable cardinals,  $\aleph_1$  and  $\aleph_2$ , to the entries of the diagram in various ways. For example, they created a universe for which  $\aleph_1 = \text{add}(\mathcal{N}) = \text{cov}(\mathcal{N})$  and  $\aleph_2 = \text{non}(\mathcal{M}) = \text{cof}(\mathcal{M})$ .

This work enabled researchers in the 1980s to confirm that for all pairs of cardinals, only the relationships indicated in the diagram can be proved in ZFC. More precisely, for every labeling of the (independent) Cichoń diagram entries with the values  $\aleph_1$  and  $\aleph_2$  that honors the inequalities of the diagram, there is a universe that realizes the given labeling.

So we have known for nearly four decades that all assignments of  $\aleph_1$  and  $\aleph_2$  to the diagram are possible. But what can we say for more than two values? Could, for example, all the independent entries be simultaneously different? Some cases with three characteristics have been known for 50 years, and in the 2010s, more universes were discovered (or constructed) in which up to seven different cardinals appeared in the Cichoń diagram.

In a 2019 paper we constructed with Israeli mathematician Saharon Shelah of the Hebrew University of Jerusalem, a universe in which the maximum possible number of different infinite values—10, that is—appears in Cichoń's diagram. In doing so, however, we used a stronger system of axioms than ZFC, one that assumes the existence of “large cardinals,” infinities whose existence is not provable in ZFC alone.

While we were very pleased with this result, we were not entirely satisfied. We worked for two more years to find a solution using only the ZFC axioms. Together with Shelah and Colombian mathematician Diego Mejía of Shizuoka University in Japan, we finally succeeded in proving the result without these additional assumptions.

We have thus shown that the 10 characteristics of the real numbers can all be different. Let us note that we did not show that there can be at least, at most or precisely 10 infinite cardinals between  $\aleph_1$  and the continuum. This was already proved by Robert Solovay in 1963. In fact, the size of the set of real numbers can vary greatly: there could be eight, 27 or infinitely many cardinal numbers between  $\aleph_1$  and  $2^{\aleph_0}$ —even uncountably many. Rather our result proves that there are mathematical universes in which the 10 specific cardinal numbers between  $\aleph_1$  and  $2^{\aleph_0}$  turn out to be different.

This is not the end of the story. As is usual for mathematics, many questions remain open, and new ones arise. For example, in addition to the cardinal numbers described here, many other infinite cardinalities lying between  $\aleph_1$  and the continuum have been discovered since the 1940s. Their precise relationships to one another are unknown. To distinguish some of these characteristics in addition to those in Cichoń's diagram is one of the upcoming challenges. Another one is to show that other orderings of 10 different values are possible. Unlike in the case for the two values  $\aleph_1$  and  $\aleph_2$ , where we know that all possible orders are

consistent, in the case of all 10 values, we could only show the consistency of two different orderings. So, who knows, there may still be hitherto undiscovered equalities—involving more than two characteristics—hidden in the diagram.

*This article originally appeared in Spektrum der Wissenschaft and was reproduced with permission.*

---

#### **ABOUT THE AUTHOR(S)**

**Martin Goldstern** is a mathematician at the Technical University of Vienna.

---

**Jakob Kellner** is a mathematician at the Technical University of Vienna.

---

Scientific American is part of Springer Nature, which owns or has commercial relations with thousands of scientific publications (many of them can be found at [www.springernature.com/us](http://www.springernature.com/us)). Scientific American maintains a strict policy of editorial independence in reporting developments in science to our readers.

© 2021 SCIENTIFIC AMERICAN, A DIVISION OF SPRINGER NATURE AMERICA, INC.

ALL RIGHTS RESERVED.