The New York Times

A Brief Introduction to Infinity by Edward Frenkel

By Gary Antonick May 30, 2016 12:00 pm

The concept of infinity has perplexed thinkers since the dawn of civilization. Everything we observe in the physical world around us is finite; even the number of atoms in the observable universe, though it is unimaginably large, is still finite. Does infinity really exist? If so, how do we find it?

I recently discussed these questions with Edward Frenkel, Berkeley mathematics professor and author of "Love and Math: The Heart of Hidden Reality." "We have many ways to connect to infinity: through art, through poetry, through love," explained Dr. Frenkel. "But mathematics gives us perhaps the most cerebral and logical way to connect to the infinite. So in this day and age, when we tend to put more trust in rational arguments than in other types of arguments, mathematics becomes our portal to infinity."

The mathematical theory of infinity was created, almost single-handedly, by the German mathematician Georg Cantor at the end of the 19th century. His ideas were so radical that many of his contemporaries flatly refused to accept them. The eminent mathematician Henri Poincaré even called Cantor's theory a "pathology" from which mathematicians needed to be cured. Though such harsh criticism caused Cantor much anguish, he stood his ground. And he was vindicated: Today his theory is a cornerstone of all of mathematics.

"Pursuing his ideas, Cantor showed tremendous courage," continued Dr. Frenkel. "Responding to his critics, he wrote: 'The essence of mathematics lies in its freedom.' In math, we have to follow rigorously the chosen axioms and the rules of logic. But within those rules, we can really let our imagination fly. There is no place in mathematics for dogma or prejudice."

Cantor's idea was that infinity is not a number, but rather a property of a *set*. Informally, a set is a collection of objects (called "elements") that have something in common. For instance, it could be the set of all readers of this blog post.

Given two sets, A and B, we can speak of a "map" from A to B: this is a rule that assigns an element of the set B to each element of the set A. For example, let A be the set of readers of this post, and B the set {0,1,2,3}. Assigning to each reader the number of puzzles he or she solves in this post, we obtain a map from A to B.

The crucial concept introduced by Cantor is that of *one-to-one correspondence* between sets A and B: this is a map from A to B such that each element of B is assigned to one and only one element of A. It follows from this definition that if A has a finite number of elements — say, n — and B is another set, then there is a one-to-one correspondence between A and B if and only if B also has n elements.

Now we can introduce the notion of an *infinite set*: This is a set A such that there is no one-to-one correspondence between A and any finite set B. For example, the set of natural numbers N={1,2,3,...} is an infinite set.

Up to this point, the theory is rather straightforward. But then Cantor made a startling discovery: There are infinite sets that are not in one-to-one correspondence with each other. In other words, infinity comes in different sizes! For example, it turns out that there is no one-to-one correspondence between the set N and the set of real numbers (which includes fractions and irrational numbers like √2 and pi).

Dr. Frenkel hailed Cantor's proof of this fact, often referred to as "Cantor's diagonal argument," as one of the most beautiful proofs in all of mathematics. "At its core is a powerful but simple idea, which can be explained through a sequence of puzzles," he said. These puzzles are presented below.

Before jumping in, part of the challenge this week is deciphering set theory terminology. But I think the journey is worth it.

As a warm-up, let's talk about subsets of a set. Here, by a subset of a given set A, we mean a particular collection of elements of A, in which every element of A either appears once or does not appear at all. For example, let A be the set {R,Y,B} of three primary colors: red, yellow and blue. Then we can think of subsets of A as colors we create by mixing the primary colors. For example, the subset {R,Y} corresponds to mixing red and yellow, which is orange, and the subset {Y,B} corresponds to mixing yellow and blue, which is green. Note that among these subsets is the empty set, which corresponds to mixing no colors at all. We can think of it as color white.

Now, given a set A, we define a new set S(A) as the set of all subsets of A. That is to say, a member of S(A) is a particular subset C of A, but we view it as a single entity (not focusing on its elements). Note that we can iterate this construction further, and consider the set S(S(A)) of all subsets of S(A), and so on. This construction shows the power and versatility of set theory. By following a simple procedure, we can wish into existence a whole hierarchy of increasingly complicated objects starting from a simple one.

Let A be the set {1,2,3,...,n} of natural numbers from 1 to n, and S(A) the set of its subsets. With each element of S(A) — that is, a subset C of A — we associate a binary sequence (s(1),s(2),...,s(n)), where each s(i) is either 0 or 1.

Namely, we define s(k)=1 if k belongs to C, and s(k)=0 if k does not belong to C. Conversely, given a binary sequence (s(1),s(2),...,s(n)), we can reconstruct C: It consists of those and only those elements k of A for which s(k)=1.

For example, if A = {1,2,3,4}, then the sequence (0,1,1,0) is associated to the subset {2,3}, the sequence (1,0,1,1) is associated to the subset {1,3,4}, and the sequence (0,0,0,0) is associated to the empty set.

Viewing elements of S(A) as binary sequences can be used in solving the following

Puzzle 1: If A = {1,2,3,...,n}, how many elements are in S(A)?

Without giving away the answer, we can note that for any n>0, the set S(A) has more elements than A. Therefore, by simple counting we find that there can be no one-to-one correspondence between A and S(A). This is an important statement. And now we will give an alternative proof of this statement, which we will then use in the case of an infinite set A. This will be our path to Cantor's diagonal argument.

Namely, let f be any map from A to S(A). This means that to each number k from 1 to n we assign a binary sequence f(k) in S(A). The map f is a one-to-one correspondence if and only if f(1),...,f(n) are all different, and altogether they exhaust all possible binary sequences in S(A).

We already know that there is no such one-to-one correspondence because S(A) has more elements than A. But we don't want to rely on this fact. Instead, we will construct a binary sequence (s(1),s(2),...,s(n)) as follows:

First, we consider the binary sequence f(1), take its first digit, and replace it by its opposite. That will be our s(1). In other words, if the first digit of f(1) is 0, then we write s(1)=1, and if the first digit of f(1) is 1, then we write s(1)=0. Next, we consider f(2), take its second digit, and take its opposite. That will be our s(2), and so on. Thus, for each k from 1 to n, we define s(k) as the opposite to the k-th digit of f(k).

Puzzle 2: Show that the binary sequence (s(1),s(2),...,s(n)) constructed above differs from f(k) for any k=1,...,n.

Thus, the binary sequence we have constructed cannot possibly be assigned by f to any element of A = {1,2,...,n}. This shows that the map f is not a one-to-one correspondence. And since we can apply this argument to any map f from A to S(A), this shows that there is no one-to-one correspondence between A and S(A).

The argument we have given is Cantor's diagonal method adapted for finite sets. Why is it called "diagonal"? Imagine an n x n square table whose first row is f(1), second row is f(2), and so on, and n-th row is f(n). Then our construction can be rephrased as follows: Take the diagonal of this square. This is a sequence of n digits, right? If we take the opposite of each of its digits, we obtain our sequence (s(1),s(2), ...,s(n)). The fact that this newly constructed binary sequence differs from every row of our square shows that the map f cannot possibly be one-to-one.

Now let's apply the same method to an infinite set: the set of natural numbers N = {1,2,3,...}. So, let S(N) be the set whose elements are infinite binary sequences (s(1),s(2),s(3),...), where each s(i) is either 0 or 1.

Cantor's famous theorem can be formulated as follows: *There is no one-to-one correspondence between N and S(N)*. And now we are in position to prove it, by using an analogue of the argument we have employed in the case of finite sets.

Let f be any map from N to S(N). That is to say, to each natural number k we assign a binary sequence f(k) in S(N).

Puzzle 3: Construct a binary sequence (s(1),s(2),s(3),...) that differs from f(k) for every natural number k.

The existence of a sequence (s(1),s(2),s(3),...) with the above property immediately implies that f is not a one-to-one correspondence: indeed, we find that this sequence is not assigned by f to any k in N. And since this argument can be applied to any map f from A to S(N), we obtain that there is no one-to-one correspondence between N and S(N). Therefore Cantor's theorem is proved.

We close with two remarks:

1) Recall from the above discussion that a finite binary sequence (s(1),s(2), ...,s(n)) is the same thing as a subset of A={1,2,...,n}. Likewise, an infinite binary sequence (s(1),s(2),s(3),...) is the same thing as a subset of the set N of natural numbers (namely, the subset of those and only those natural numbers k for which s(k)=1). Thus, Cantor's theorem can be rephrased as saying that there is no one-to-one correspondence between N and the set of all subsets of N. In a similar way, one can prove the analogous statement for any set A (finite or infinite): there is no one-to-one correspondence between a given set A and the set of all subsets of A. Namely, given any map f from A to S(A), we can construct a specific subset of A which is not equal to f(a) for any a in A. As a bonus puzzle: Can you describe it?

This is the most general formulation of Cantor's result. It shows that starting with the set of natural numbers, we can obtain a never-ending hierarchy of "larger and larger" infinite sets by taking the set S(N) of all subsets of N, the set S(S(N)) of all subsets of S(N), and so on. And this is just the beginning of the theory of infinite sets.

2) Independently, one can show, using binary representation of real numbers, that there is a one-to-one correspondence between S(N) and the set of real numbers. Therefore the above theorem implies that there is no one-to-one correspondence between the set of real numbers and the set of natural numbers. Informally, one could say that the "infinity" of real numbers is "bigger" than the "infinity" of natural numbers.

It is natural to ask: is there an "infinity" in-between these two? This leads us to the famous "continuum hypothesis," one of the deepest problems in mathematical logic. But perhaps that's a subject for another post.

Edward Frenkel recently talked about infinity and the role mathematics is playing in today's world with the journalists from the Turkish website Her-An. Here's the video of his interview (to see the English translation of the written questions, click on CC at the bottom of the screen; the answers are in English):

With that we wrap up this week's challenges. As always, once you're able to read comments for this post, use Gary Hewitt's Enhancer to correctly view formulas and graphics. (Click here for an intro.) And send your favorite puzzles to gary.antonick@NYTimes.com.

Solution

Check reader comments on Friday for solutions and recap by Edward Frenkel.