What V≠L stands for in Mathematics and beyond

Numbers are collective animals and their essence lies in their relationships. In other words, what is essential about numbers is how they add, multiply, and divide each other. Therefore, the best way to understand numbers is to calculate them. However, gaining an understanding of the basic features of numbers requires a more theoretical outlook.

Here is one such feature: the series 1, 2, 3, ... has a beginning but no end. It is open-ended. In other words, it is infinite. To refer to this open-endedness we use dots (...), or the English words "and so on," or the Latin abbreviation "etc." As Wittgenstein famously remarked, "and so on" can be highly ambiguous. How can one be sure that after having been encouraged to go “and so on,” everybody will go exactly in the same direction? To some extent, this direction is controlled by educational institutions. Schoolteachers insist that 1, 2, 3 is followed by 4, 5, 6. It is what the expression "1, 2, 3, ..." is supposed to stand for. However, this kind of control is limited, because at some point it will be necessary to stop counting and say “and so on” or write out the dots.

There are two ways to interpret the dots. They can be seen as an indicator of the possibility of continuing 1, 2, 3, ... without limit by adjoining a new number over and over again. The other option is to interpret the dots as a symbol that stands for the rest of the ready-made infinite series en bloc. These two interpretations of the dots rely on two different notions of infinity, which in scholastic language are called *potential* and *actual* infinity. Potential infinity appears to be “safer,” for it doesn't require anything more than an operation, which associates the next number with any given number. This is easy: N is always followed by N+1. So saying that 1, 2, 3, ... is potentially infinite is the same as saying that this +1 operation can be applied over and over again, without limit. It is less clear how to conceive of 1, 2, 3, ... as *actually* infinite but obviously much stronger assumptions are necessary. For 1, 2, 3, … to be actually infinite would seem to require that the +1 operation be carried out an infinite number of times. Or, perhaps it is better to forget about the +1 and conceptualize the numbers in an entirely different way. However, the notion of actual infinity by itself doesn't offer an alternative way of thinking about numbers, and thus remains confusing and perhaps contradictory. (What comprises the entire difference between a finite string of numbers and an infinite series of numbers, so the argument goes, is precisely the fact that the latter, unlike the former, is not and cannot possibly be completed. So the claim that 1, 2, 3, ... is actually infinite is tantamount to saying that it is and is not completed at the same time.)

For the reasons described above, for a long time the potential infinite was generally considered safe and the actual infinite was considered suspicious or even plainly unsound. In the 20^{th} century this traditional attitude towards the actual infinite changed dramatically, while the traditional view that the potential infinite is safe and almost trivial didn't change much. However, my bet is that in the next century this view of the potential infinite will also change. The notion of the potential infinity is far more problematic and more interesting than it appears. It is easy to conceive of the possibility that adding one will yield a new number; but the very notion of possibility doesn't have any obvious meaning in mathematics. One can imagine a cake without actually eating it. On this ground, one can distinguish between possible and actual cakes. But how can one distinguish between possible and actual numbers in the same way? It is sheer absurdity to think of the numbers 3 and 4 differently only based on the fact that 3 shows up in the expression "1, 2, 3, ..." and 4 does not, because the series 1, 2, 3, ... and 1, 2, 3, 4, ... are exactly the same, or at least so says the common convention.

One way to take the notion of potential infinity in mathematics seriously is to consider an idea hinted by Wittgenstein and, as legend would have it, defended by Kolmogorov: the usual concept of numbers and arithmetic generally fails to apply to big numbers in the same way that Euclidean geometry generally fails to apply to big distances. The common belief that "and so on" leaves no room for genuine surprises may turn out to be wrong. We need to keep an open mind and not assume that everything is clear in advance.
****
****

The man who first clearly stated that mathematics needs and can tolerate the concept of the actual infinite was Georg Cantor (1845-1918). This was more than just a philosophical claim; Cantor established an entire new mathematical discipline for the study of infinite collections, which today is called set theory. Cantor's major discovery was that the actual infinite allows for degrees of infinity. This is less obvious than one might initially think. For example, one may believe that the infinite series of even integers 2, 4, 6, ... is smaller than the infinite series of all integers 1, 2, 3, ... because the former series is wholly contained in the latter. However, the two series are in fact the same size. This claim is explained as follows: any number N from the series 1, 2, 3, ... can be associated with an even number 2N, and conversely, any even number M from the series 2, 4, 6, ... can be associated with a number M/2. This establishes a one-to-one correspondence between members of the two series. Two given collections (sets) are said to be of the same size if and only if there exists a one-to-one correspondence between their elements. This definition works for both finite and infinite collections, but the implications are different. If there exists a one-to-one correspondence between the elements of set A and some portion of the elements of set B then B is at least as big as A. In the finite case this definition of "at least as big as" always implies "strictly bigger than." But as the above example shows, in the infinite case this is not necessarily true. However, it may be shown that if A is at least as big as B, and B is also at least as big as A, then A and B are of the same size, just as in the finite case. If A and B have different sizes and B is as at least as big as A, then B is strictly bigger than A.

Cantor's basic example of an infinite set, which is strictly bigger than 1, 2, 3, ..., has two representations. It can be described as the set of all points on any given line or on any other continuous geometrical object. Or, it can be described as the set of all series (finite and infinite) contained in the series 1, 2, 3, ..., i.e., as the set of all subsets of 1, 2, 3, ... Why these two sets are thought to be the same size is not very important in the present context so I will skip this part of the story.

Here is a shortened version of Cantor's diagonal argument, which shows that the set of all subsets of 1, 2, 3, ..., referred to as its *powerset*, is strictly bigger than 1, 2, 3, ... itself. Let's first associate with every subset of 1, 2, 3, ... an infinite series of 0's and 1's, which is built as follows: if N is an element of the given subset one puts a '1' at the N^{th} place of the corresponding 0-1 series, and otherwise one puts a '0'. If we want to include all the members of 1, 2, 3, ... in a subset, then the above rule gives us 1, 1, 1, ...; if we want to include none of them the rule gives 0, 0, 0, ... and so on. This establishes a one-to-one correspondence between the 0-1 series and subsets of 1, 2, 3, ... Now suppose that every single one of these 0-1 series are listed and enumerated from top down:

1) **1**0110001000010000100101...
2) 0**0**101001011011010101001...
3) 10**1**10011100010100010011...
.
.
.

Take the diagonal of this matrix and invert it by replacing every 0 with 1 and every 1 with 0. By doing this we will get a particular 0-1 series that we will call D. By our hypothesis, D must be found somewhere in the list. However this is clearly impossible because of the inversion! Suppose that D coincides with a series S that resides at the Nth position in the list. If D has a 1 at the Nth place, then S should have a 0 at the N-th place—yet the converse has to be true as well. This contradiction shows that all 0-1 series—and hence all subsets of 1, 2, 3, ...—cannot be arranged into an infinite series. We have thus found a set that is strictly bigger than 1, 2, 3, ... This example easily generalises to a theorem, which says that the powerset of any given set is strictly bigger than this given set. So using the powerset operation repeatedly yields an infinite hierarchy of infinite sizes, just as completing the operation +1 repeatedly yields the infinite series 1, 2, 3, ...

Cantor's mathematical advances didn't resolve philosophical controversies about the actual infinite but showed that infinite collections are mathematically treatable in a non-trivial way. He expressed his personal stance towards traditional philosophical concerns about the actual infinite in his famous slogan “The essence of mathematics lies in its freedom” (*Das Wesen der Mathematik liegt in ihrer Freiheit*), describing the freedom from what Cantor called a "metaphysical control". However, the new freedom of mathematical creation didn't come without a price. Cantor's Mengenlehre, i.e. set theory, turned out to be full of contradictions, which became known under the terms of antinomies and paradoxes. The simplest paradox, sometimes referred to as Cantor's paradox, resembles the traditional argument that the notion of an actual infinity is contradictory to begin with. Consider the set U of all sets and its powerset PU. According to the theorem described above, PU is bigger than U, which is impossible since every element of PU is also an element of U by definition of U. Another paradox of a very different nature is known as Russell's paradox. Think of a set A containing all sets, which are their own elements (like U, for example) and another set B, containing all sets which are not their own elements (like 1, 2, 3, ...). Then the classical logical law of *tertium non datur* leaves us with only two mutually exclusive possibilities: either B is its own element and hence is an element A, or B is not its own element and hence is an element of B. So by making assumptions about sets, which seem innocent enough, and relying on the usual laws of logic, we arrive at a stunning contradiction: if B is its own element it is not its own element, and the converse is also true. Note the analogy with Cantor's diagonal proof. One man's proof turns out to be another man's paradox!

These and other paradoxes found in Cantor's theory of sets persuaded people that there was something wrong with this theory. However, only a small minority considered this sufficient cause to give up the new theory altogether. An attempt to save the theory, which greatly influenced all later developments in the field, was made in 1908 by Zermelo, who designed a list of axioms for Cantor’s theory. An improved version of Zermelo's axiomatic theory of sets was developed by Fraenkel (1891-1965), which was named ZF theory after its

authors, and it remains standard even today. NBG is another axiomatic theory of sets named after Neumann, Bernays and Gödel , which is essentially equivalent to ZFC but has more expressive power. Specifically, ZFC prohibits the notion of the set of all sets but offers no replacement. NBG offers a replacement that is conventionally called V: this name refers to the universe of sets, which itself is not a set but a proper *class*, i.e. an entity which like a set has some elements, but unlike a set cannot be itself an element. These and other axiomatic theories of sets that are currently on the market help to avoid all the known paradoxes of Cantor's set theory but promise no protection from new paradoxes, which may come up in the future.
****
****

The idea of putting set theory on an axiomatic basis is actually more controversial than it seems. The modern version of the axiomatic method used by Zermelo and his followers involves two principle steps: (1) one stipulates a list of axioms, which only explicitly mention abstract individuals and the abstract relations between them; (2) one finds appropriate individuals and appropriate relations, which satisfy the stipulated axioms. A system of individuals and relations satisfying the given axioms is referred to as a *model* of the theory determined by these axioms. One may reasonably ask where individuals and relations needed for building models of axiomatic theories can possibly be found. The answer is that they are found in some other theories, referred to in this context as *metatheories*. If, for example, one would like a model of [a formalised version of] plane Euclidean geometry, it would be possible to use arithmetic as a metatheory for this purpose. Instead of thinking about points in the usual intuitive way, in this case one thinks of them as particular arithmetical constructions: points become pairs of numbers.

The usual attitude towards metatheories is that they are basic theories, which can be safely taken for granted and then used for theory-building in different domains of mathematics. Even if set theory is not as safe as one might like, there is another reason to consider this particular theory as a metatheory for the rest of mathematics: it allows for the rebuilding of all major domains of mathematics in its terms. Many have considered this kind of set-theoretic rebuilding of mathematics as the best way to clarify mathematical ideas. Although results of the practical realisation of this project have been controversial, it is hardly possible to talk about set theory without taking its foundational aspect into consideration.

The fact that set theory (broadly conceived), unlike any other mathematical theory, is deeply involved in the very notion of axiomatic method makes it questionable whether one can use this method for treating set theory itself. In this case it seems like one needs to have the job done before having started it. Consider this puzzling observation known as the Skolem paradox: a countable set (i.e. one equal in size to 1, 2, 3, ...) is sufficient to build a model of ZF, i.e. to "be" or represent the universe of all sets containing an infinite hierarchy of infinite sizes! In Skolem's own eyes this was strong evidence that Zermelo's efforts to develop an axiomatic theory of sets were futile to begin with. However one may also argue that the circularity involved in building axiomatic theories of sets is not actually a vicious one, and that by exploring it one learns important things about sets and the foundations of mathematics. Thus the development of set theory continued during the 20th century in this new sophisticated model-theoretic setting. From this new point of view on set theory, Cantor's way of thinking about sets is usually qualified as

"naive".
****
****

Cantor conjectured that the infinity of points on a given line is exactly the next bigger infinity after the countable infinity, but he didn't manage to prove this claim. This conjecture became known as the Continuum Hypothesis, or CH. The first significant advance concerning CH was made in the new refined model by Kurt Gödel in 1940. Gödel constructed a model of ZF, which he called L (also known as the constructible universe), where CH provably holds. This model shows that CH is compatible with ZF, i.e. it doesn't contradict its axioms. Historically, L was the first example of an *inner model*. To obtain an inner model of ZF one first assumes some model M of ZF and then constructs out of it the desired new model through an appropriate restriction of M. If one relativises this whole construction to L, i.e. takes L instead of M to begin with and then repeats the construction, one gets L back. A related fact about L is that in this model V=L holds. It turns out that L is the only model of ZF having this property. Thus we have V≠L in any other model of ZF.

Unlike V, Gödel's constructible universe L has hardly ever been seriously considered as a refined version of the "naive" Cantor's universe, for it is too obvious that the nice properties of L have purposefully been introduced into its definition. Gödel 's own motivation behind L was not to get a proof of CH but to show that CH is an independent hypothesis, which cannot be proved or disproved in ZF. The part of the argument that Gödel missed was supplied in 1963 by Paul Cohen, who used his new method of *forcing* for building models of ZF where CH fails. Like inner models, forcing models of ZF require some assumed model (Cohen used a countable model). But instead of restricting this base model, forcing allows it to be extended in a useful way. A suggestive analogy is the extension of the field of rational numbers by irrational numbers.

Cohen's result showed that ZF indeed doesn't imply anything definite about CH: in some models (namely in L) CH holds, and in some other models it fails. Whether this independent result resolves the whole problem of CH by showing that it is in fact a non-issue, or whether it only shows that ZF fails to describe the set concept properly, remains a controversy. But in any event it clearly shows that when thinking about infinite sets one must keep an open mind and avoid taking any particular construction as the last word in the story. This is what $V \neq L$ stands for above and beyond its usual technical meaning.