What's Up with Set Theory?

Gary Venter
gary.venter@gmail.com

A few articles have been hinting about big, exciting developments in set theory. That sounds like an oxymoron already. It has to do with finding the right axioms for set theory. The ancient Greeks used axioms and proofs as solid grounding for their mathematics, starting with definitions and obvious truths, and deriving everything from there. Set theory now is the foundation for all mathematics. Yet its axioms are not quite obvious, and if not carefully crafted, they could even be inconsistent. Now we know that they are also incomplete - there are true statements of set theory that cannot be proven - and there are areas where more axioms are needed. This has been an issue for the last six decades.

It sounds like a resolution is at hand. We finally have axioms that can clarify the structure of the set-theory universe. While there are some things that an axiom system can never prove, we can live with those. But re the axioms, the problem now is that there is not just one solution, but two - and they are mutually exclusive.

It's impossible to explain all of that in full detail in a short article. This note is an attempt to summarize the key issues for a math-friendly but non-expert reader. There is now some material helpful at that level, and slightly beyond it. Some good examples are Natalie Wolchover (2013 & 2021) in Quanta Magazine. A more detailed discussion is in Jefferey Schatz (2019), his philosophy PhD dissertation on how to decide between the two competing axiom systems.

In this note, the difficult details - especially a new way to create sets called "forcing" - are summarized conceptually, which hopefully will be enough to get an intuitive feeling for the situation. We start with the history that led us to the current, but still fairly new, ZFC axioms of set theory and their goal to formalize mathematics in a complete and decidable way. Next addressed are the problems that arose, starting with the discovery that axioms strong enough to formalize mathematics could not be proven to be consistent. And so if they are consistent, there have to be true but unprovable theorems - the fact that they're consistent being one such.

That is a general problem with any axiom system. For set theory, the ZFC axioms more specifically turn out to have problems sorting out how to measure infinity. They have actually come up with ways to do that, and there are big and then bigger then still bigger sizes of infinity, and those just keep on going. But how to do that in practice is difficult already for the real numbers used in calculus - set theorists can't quite figure out how to express the size of that infinitely large set. The set of subsets of the natural numbers has the same issue - and this generalizes to any infinite sets of subsets. We summarize these issues, then look at the solutions coming out: make new axioms. How to do that and the recent alternative answers are discussed, and a discussion section wraps it up.

History of Axiom Systems

Set theory is about axioms, and theorems derived from them. A set is just a collection of things - like the set of pillows on my sofa - but a more specific definition is needed for mathematics. Axioms came to us from the ancient Greeks. They were ideas so obvious or sensible that they do not need any justification, and which everything else can be derived from. Euclid's Elements, circa 300 BC, compiled the work in geometry and number theory from the preceding centuries of Greek mathematicians, such as the Pythagoreans from around 500 BC. Euclid reported on their axioms, theorems, and proofs. The Pythagoreans had their famous theorem about the hypotenuse, and from that they realized that some numbers, like the square root of 2, could not be expressed as a ratio of integers - they were not ratios, so were deemed "irrational." Pythagoras himself considered this result to be scandalous. The Greeks were influenced by the Babylonians and Egyptians before them, and India had its own systems that influenced all these. Ancient China had logic systems but these had been abandoned until Indian ideas like Buddhism revived them. The Greek Stoic school's logic system closely resembled modern predicate logic. See Mates (1953).

The late 1800s and early 1900s experienced a major revival, renewal, and formalization of logical reasoning. Peano, Cantor, Frege, Dedekind, Hilbert, Zermelo, Russell, Gödel, Tarski, and Cohen are some of the famous contributors. Frege in 1879 introduced symbolic logic, with formal symbols like $\exists$ and $\forall$ (for some … and for every …) with formal rules for the expression of logical statements. This was needed because logic using natural language could run into ambiguities, and these were problems for mathematical proofs. A somewhat silly example is "Every girl kissed a boy." Was that a particular boy, or possibly a different one for different girls? Both of those versions can be formalized as different statements in the formal system, e.g., "there is a boy that every girl kissed", or "for every girl, there is a boy that she kissed." Same for "All the inhabitants of the islands were men or women." That could be "all the inhabitants of all the islands were men or they were all women," or "all the inhabitants of each island were men or they were women," or "all the inhabitants of all the islands identified simply as male or female."

Russell's paradox from 1901 upended the efforts of Frege. He showed that Frege allowed the set of all sets that are not members of themselves. If that set is not a member of itself, then by definition it is a member of itself, which would then make it not a member of itself, … . Contradiction! This formulation came from allowing a set of all sets, which was of course a member of itself. That is now a taboo in all current systems.

Zermelo set out to create a new, consistent, set of axioms for set theory, first in Zermelo (1908). Fraenkel (1919) and Skolem (1922) made a few modifications. Zermelo (1930) is a finalized report. The result is the ZF axioms, now called the ZFC axioms, to emphasize that they contain the axiom of choice (AC). That axiom states that if A is a set of non-empty sets, then there is another set B that contains an element from each set in A. There is some question as to whether that statement is self-evident. Gödel and Cohen did separate proofs that AC is consistent with and independent of the other ZF axioms. But some other axioms have been tried as alternatives to it now and then.

Zermelo argued that self-evident or not, we need AC for its results, and that is what is important when setting up a system of axioms. The argument based on results has become known as the extrinsic approach to axiom justification. That is now a criteria used for deciding among the new contenders. Another is that proposed axioms should be independent of and consistent with the other axioms.

Hilbert in the 1920s set out a program to express and develop all of mathematics from set theory. That was straightforward, because you can define numbers and other mathematical objects as types of sets. But his program also called for the system to be able to establish the truth or falsity of any mathematical statement through proofs and for the axioms themselves to be provably consistent. That didn't turn out so well.

Undecidable Statements

Kurt Gödel shocked the mathematical world with the proof in his 1931 papers that if the axioms of arithmetic or any stronger system like set theory are consistent, then there are true statements in the theory that cannot be proven from the axioms, one such being the fact that the axioms are in fact consistent. That is the incompleteness theorem - a set of axioms can never decide every question. And the axioms cannot be proved to be consistent anyway. Hilbert's program had crashed.

Gödel's proved his theorems by coding all mathematical statements as numbers that could be decoded through their prime factorizations. That was not so hard because by then the axioms and proofs in several key mathematical areas, like number theory, calculus, and geometry had been written out in formal symbolic logic notation. So he just had to develop a coding system to put all those symbols and combinations of them, like proofs, into a language written in numbers. Then he showed further that properties of statements, like being an equation, etc. could be coded similarly. One example: a statement has the property of being provable if it the last line of a derivation (which was itself defined as a specific combination of statements, etc.). The hard part was next: the self-reference lemma. This showed that any statement that asserts a property can be applied to the number encoding that statement. For instance, there is a number that encodes the statement "This statement's number is greater than 100." So that number is a sentence that refers to itself. The proof of that lemma requires digging into the coding, and I'm willing to assume that it is right. Such self-reference sounds dangerously close to Russell's set of all sets, but Gödel delicately tiptoed around the potential inconsistencies.

From there, the rest is straightforward. The sentence "This statement is not provable" could now be expressed as a statement in arithmetic. Note that if it is provable, then it is false. But a false statement cannot be provable under the assumption that the axioms are

consistent. So if the axioms are consistent, this statement must be unprovable. Then what it says is true. Not quite a contradiction there: it is a true, unprovable statement.
Once you've proved that if the axioms are consistent then this statement is unprovable, you can't prove that the axioms are consistent, as then you would have proved the statement. Thus if the axioms are consistent, you cannot prove that they are consistent.

A further consequence is that you cannot formalize the property of "truth" within arithmetic like you can for "provable." If you could, you can get the statement "This statement is not true." Then if it is true, it is not true. But if it is not true, it is true - an actual contradiction!

Even though incompleteness upended Hilbert's program, after a while practicing mathematicians started to dismiss the incompleteness of arithmetic as largely irrelevant. Gödel's unprovable statement was nothing like the potential theorems they were working on, and they were not coming across actual problems of undecidability. Further, there were some proofs of consistency of axiom systems coming out. In 1936, Gerhard Gentzen showed that you could prove Peano's axioms were consistent if you used a slightly stronger set of axioms that included transfinite induction. In 1926, Tarski reformulated the axioms for geometry, and showed that plane geometry is consistent, complete and decidable. Every statement can be proved or disproved from the axioms, and there is a rule for how to decide them. Geometry is too small a system to code math into, so it is not included in the incompleteness theorem. When you could prove consistency of a set of axioms, it could only be within a stronger system. Nonetheless, Hilbert's program was doomed. Not much later, the axioms of set theory were having even bigger problems.

Sizes of Infinities

Two ways have been developed for counting the sizes of sets, namely the ordinal and cardinal numbers. For finite sets these are the same, but it gets interesting for infinite sets. An ordinal is a well-ordered set, i.e., it has an ordering relationship, written $a < b$, between its elements, and there is a least element. The natural numbers are all ordinals, as is the set of all of them. The ordinals as defined in set theory start with $0 = \emptyset$, the empty set. Then $1 = \{0\}$ (the set containing 0), $2 = \{0,1\}$, $3 = \{0,1,2\}$, … . The definition is that an ordinal is the set of all smaller ordinals. This is well-ordered by defining "$<$" as the membership relationship: if a is a member of b, then $a<b$. This process first defines the finite ordinals 0, 1, 2, … . Then the ordinal ω is defined as the set of all finite ordinals, which by definition of "$<$" as set membership is greater than all of them. That's an example of a limit ordinal - one that is not a successor of another ordinal. Still, every ordinal is the set of all smaller ordinals. Continue $\omega+1 =$ union of ω and $\{ \omega \}$, which contains ω as an element, and alsoo as a subset. Then continue with $\omega+2$, etc. to get 2ω, 3ω, ..,. The set of all of those multiples of ω is ω^2. Keep adding 1 then taking limits to get ω^3, …, ω^ω, then towers of powers of more and more ω's. The ordinal after all of those towers, i.e., the set of all of them, is called ω_1. Then again add 1, etc. and keep going.

Cardinal numbers are measures of how many elements are in a set, regardless of order. The cardinal sizes of sets are compared by matching up elements of the two sets. For instance, 12 eggs match up 1:1 with the dimples in an egg carton. If there is any such 1:1 mapping, the sets have the same cardinality.

The infinite sets start off with N(0) being the cardinality of the natural numbers. Any set that has a 1:1 mapping with the natural numbers also has cardinality N(0), i.e., is a countable set. For instance, the even numbers are countable. The 1:1 mapping of the evens with the natural numbers is that an even number b maps to b/2. That gets all the natural numbers.

To show that a and b have the same cardinality, you don't have to find an actual 1:1 map. It is often easier to map a to a subset of b, then b to a subset of a. For instance, the positive rational numbers can match with the positive integers by this method. Any positive integer n is already a ratio n/1. To show that all the rationals can be mapped into the natural numbers, picture an infinite grid of pairs of positive integers, representing numerators and denominators of fractions, put into a rectangle going upward 1, 2, 3, ..., for the numerators, and going across 1, 2, 3 ... for the denominators. The bottom corner is the point (1,1). You can number the whole grid with a single string of integers by snaking up along the diagonals. Start with 1 for (1,1) then continuing 2, 3, 4, 5, etc. for (2,1), (2,2), (3,1), (3,2), (3,3), etc. That gives every pair a positive integer match. Each rational number a/b corresponds to the pair (a,b). But some of those pairs will have common factors and so will not be in lowest terms. Toss those out. Then you have a mapping from the positive rational numbers into a subset of the positive integers. Thus there are N(0) rational numbers. This is a form of the diagonal argument.

Call the set of all countable ordinals Ω_1. By the definition of ordinals, it is an ordinal - it consists of all the ordinals before it. It cannot be countable, or it would be a member of itself, so it is the smallest uncountable ordinal. (Any smaller ordinal would be an element of it, as is true of all ordinals.) It is then defined as N(1), the first cardinal number greater than N(0). (The usual notation for these cardinals is $aleph_0$ and $aleph_1$, ... using the Hebrew letter aleph, but here N(x) will only be used for cardinal numbers, and using parens instead of subscripts will avoid having subscripts within superscripts. This simplifies the notation a bit.) The cardinal number of a set is defined as the smallest ordinal number it can be put into a 1:1 correspondence with. This gives that N(2) is the ordinal that is the set of all ordinals of cardinality of N(1) or less, etc. for N(3), N(4) ... . This leads up to N(ω), the set of ordinals N(j) with integer j, and the cardinals just keep going up from there N(ω+1), etc., as do the ordinals.

The real numbers in [0,1) are each strings of N(0) digits. In binary notation, which uses 0.1 = 1/2, 0.01 = 1/4, etc., you can also write reals as strings of bits, 0 or 1. Then any real number x in [0,1) defines a subset A_x of the natural numbers by the rule: put the number k in A_x if the kth bit of x is a 1, and leave it out if the kth bit is a 0. Moreover, every set of natural numbers specifies the bits of the real number x with the same mapping. This is an actual 1:1 map between the reals and the subsets of the natural numbers. The real numbers thus have the same cardinality as the set of all subsets of the natural numbers, and this is called $2^{N(0)}$. (A finite set with j elements has 2^j subsets.) The next question is, where is $2^{N(0)}$ on the list of cardinals, N(1), N(2), ...? It turns out that the numbering systems are not specified sufficiently to answer that question without more axioms.

To start, first note that there are more real numbers than there are natural numbers. That means there is no list 1, 2, 3, ... that includes all of the reals. You can show that by starting with an arbitrary list of N(0) real numbers, and then showing that there has to be another real number that is not in the list. Just define the kth bit of the new real number by 1 - the kth bit of the kth real number in the list. That new number won't be anywhere on the list because it differs in at least one bit with every number on the list. This is another diagonal argument. Therefore no countable list of natural numbers can include all the reals, and $2^{N(0)}$ > N(0). In fact, for any cardinal N(j), $2^{N(j)}$ is strictly greater than N(j).

Cantor hypothesized that $2^{N(0)}$ is N(1), it's smallest possible value. This is called the continuum hypothesis, or CH. Cantor was never able to prove that from the ZFC axioms, and now we know that CH is undecidable in ZFC. Details on that are below. Unlike in number theory, this shows that there is a real question in set theory that mathematicians care about but which is undecidable from the established axioms. One or more new axioms is needed to specify the cardinality of the real numbers.

In a way this is not so surprising, as the ZFC axioms were built up recently to try to create a formal mathematical structure that would resemble our intuitive ideas about sets and which could be used to formalize the rest of mathematics. Arithmetic and geometry had gone through similar developments over centuries, and we now have well established axioms for them. But set theory is new and still developing. And counting sizes of infinities is a new process within it. Mathematicians feel that we should be able to figure out how many subsets of natural numbers there should be - that has to be a specific cardinal number. So it's: good start, Zermelo, but keep going. We need more axioms.

Adding Axioms to ZFC

The ZFC axioms start with a few definitions to say what a set is (e.g., it is defined by what its elements are), and then describe ways to specify some sets, like the empty set, the set containing that, etc. From there, you can use formulas of set theory to define subsets of a set (but you can't define a set with a formula except for subsets of existing sets). If you have two sets, then some other set has those two as its elements. Unions of sets (made by combining them) are sets, and for every set, there is a set of its subsets. There is an infinite set. If c is a set and f is a function on the elements c_i of c so that $f(c_i)$ is a set, then there is another set that contains all of the $f(c_i)$. The axiom of choice says that if c is a set whose elements are all non-empty sets, then there is a set that has one element from each element of c. Those are enough rules to build up the structures needed for most of mathematics. But more are needed for the unanswered questions about sets themselves. What we are looking for new axioms that give other ways to define more sets. But we want to make as sure as we can that the new axioms do not lead to contradictions. Plus we want them to give us an expanded vision of set theory and to solve some problems, like CH.

To spell out the proposed new axioms, two hierarchies of sets are needed, namely V, the Von Neumann universe of sets, and L, the constructible universe.

V is indexed by the ordinal numbers. It starts with $V_0 = \varnothing$. Then V_{b+1} is the set of all subsets of V_b. You can continue this to limit ordinals c like $c = \omega$, the set of finite ordinals. V_c is the union of all of the V_b stages for $b < c$. V is the whole Von Neumann universe, but it is not a set - it is not defined by the methods of defining sets that the ZFC axioms describe. It is too big - like the set of all sets. It is often considered to be a class, which is a grouping too big to be a set. There is a class of all ordinal numbers as well - it's not a set either. Von Neumann studied V, but it was initially defined in Zermelo (1930). The letter V was used, from the Latin word "verum," to represent the universe of all sets as early as Peano in 1889.

The constructible hierarchy L is also indexed by ordinals. L is built up in stages like V is. L_a is the constructible sets in V_a. They are the sets that can be defined by a formula of set theory in terms of the elements in the lower stages of L. As Wikipedia contributors (2023) put it:

"In von Neumann's universe, at a successor stage, one takes V_{a+1} to be the set of all subsets of the previous stage, V_a. By contrast, in Gödel's constructible universe L, one uses only those subsets of the previous stage that are definable by a formula in the formal language of set theory, with parameters from the previous stage and, with the quantifiers interpreted to range over the previous stage."

The constructible sets thus can only include subsets which can be defined explicitly, so you know what they are. But they are not necessarily all of the sets. Focusing on constructible sets is called "the inner model program."

Forcing and Forcing Axioms

One key concept in making new sets is forcing, introduced by Paul Cohen (1963) to prove that there are models of the ZFC axioms in which there are more than N(1) real numbers and sets of natural numbers. Gödel (1940) had proved that there are models with exactly N(1) reals as CH says. But after Cohen's proof, we know that not all models of ZFC agree with CH, which is what makes CH independent of those axioms.

Even before Cohen's proof, Gödel felt intuitively that there were N(2) reals (see Gödel 1940), and that the CH would eventually prove to be an example of his incompleteness theorem. In 1947 he wrote "the role of the continuum problem in set theory will be this, that it will finally lead to the discovery of new axioms which will make it possible to disprove Cantor's conjecture." So Cohen's proof must not have been a shock to him.

Forcing is a way to add sets to a model to generate a new model that has some desired property. In this case Cohen started with a model of ZFC set theory, and added enough sets to give an expanded model of ZFC with the property that there are N(2) real numbers, and N(2) subsets of natural numbers.

Cohen looked closely at the definition of the cardinality of a set X. It is N(z) if there is a 1:1 mapping between X and N(z). A mapping is actually itself a set: a set of ordered pairs <a, b> with a in X and b in N(z), with no repeats of a's or b's. The map is 1:1 if all the elements of X and of N(z) are included. A model of ZFC is a set in which all the axioms hold. But not all the sets in V have to be in the model for it to satisfy the axioms. For instance, not all of the actual 1:1 maps have to be in the model. Since CH was already known to be consistent with ZFC, there is a model that has a 1:1 map of its set of reals with its N(1). Forcing starts with that, then builds another model that has a map of its reals with its N(2).

Exactly how forcing can add sets to a model with a specific desired property is where the work comes in, and it requires several new concepts, such as filters, splitting conditions, forcing conditions, and posets. The detailed definitions of those are beyond the scope of this note. But the 10,000-foot view is that a filter is a set of the large subsets of another set and a poset, or partially ordered set, is an ordering of the sets in a filter. Forcing builds up a new desired set one point at a time. It starts with a filter of sets having a desired property, like being uncountable. Then the posets are sequential intersections of more and more sets in the filter. If done right, the intersection of all those sets ends up with a specific chosen element. Then repeat all that with different starting points enough times to fill up the new desired set. That manages to create a set that will have the desired property - like being in a 1:1 map with N(2) - in the new model.

Wikipedia contributors (2024) outline a way to do the forcing process using the Löwenheim–Skolem theorem. It says that there are countable models of the whole ZFC set theory. That sounds impossible, since there are uncountable sets in ZFC. It can be done by restricting which 1:1 maps are in the model. A countable model M of ZFC has a countable ordinal $N(1)^M$ in it that inside the model looks uncountable, because the model has in it no 1:1 maps of $N(1)^M$ with the natural numbers. The same can be done for $N(2)^M$, etc.

Wolchover (2021) illustrates this for a real number example, but it is a terse explanation. I believe this is the idea: start with a real-number line that has $N(1)^M$ points on it inside the model M, but look at it in V where those points constitute only a countable subset of the reals. Each point on the line splits the reals, excluding that point, into two segments: those greater than the point, and those less. Do this then for another of those points, then another, etc. until you have a splitting for each original point. Since $N(1)^M$ is a countable ordinal in V, its elements are ordered 1, 2, 3, … . Go through the $N(1)^M$ points on the line in this order and pick either the upper or lower section for each. Cohen shows it is possible to do this in a way so each step has a segment you can pick that overlaps with the selected segments from all the previous steps, so take that one. The selected segments are a filter, and the sequence of intersections of the segments is a poset. Cohen shows that all of those segments in the end will overlap (intersect), but only at a single real number in V that is not in M. This is a countable sequence of open intervals that intersects at a single real number. I guess that's plausible in that a real number is a countable list of digits, and the digits up to any point put the number in the interval of reals that agree up to that digit.

Anyway, then use that single number in building a new model W that adds points to M. Repeat with another starting point to get another new element of W, etc. You can map the additional points 1:1 to the countable ordinal $N(2)^M$ as you add them, and get an uncountable number of new reals, as measured in the new model. This forcing process is designed to make the new set of reals uncountable within W. Then you can use the other axioms of ZFC to expand W to be a model of ZFC by including subsets, power sets, etc. That's broadly how forcing gets new sets with specified properties into a model of ZFC.

Spelling out the details would be a big step, but now at least we can see that the central concept of forcing is adding sets to a model to get a new model having a desired property. This can be extended to "forcing axioms," which say that the type of set added to M to get W actually exists within V. Schatz (2019) notes that a key point here is that if ZFC is consistent, then ZFC expanded with a forcing axiom is also consistent. That's because there is a model of ZFC + the new axiom, namely the W just created. Forcing axions provide a way to define sets by describing properties. In ZFC you can use a formula to define a subset, but forcing axioms define new sets that do not have to be subsets of another set.

Forcing axioms can make $2^{N(0)}$ be N(2), or N(780), or anything. But once you add such an axiom, using forcing to add more sets has to be consistent with that axiom. To add a forcing axiom, you need more justification than just because I can. The search for new axioms looks at what the possible axioms can tell you about the universe of sets. Part of this gets into the study of what are called large cardinals.

Large Cardinals

A different way to expand set theory is with large-cardinal axioms. These specify new higher cardinal numbers that are not definable with the set-building processes of the ZFC axioms. They are very large. While there is no precise definition of large, sets known to be inconsistent with ZF are usually excluded. Also, if k is a large cardinal, it a model of ZFC - that is all of ZFC set theory could be contained within it without violating any axioms. If such a cardinal exists, ZFC is consistent because it has a model. But if ZFC is consistent it cannot prove that it is due to the incompleteness theorem, so it cannot prove the existence of large cardinals. Each one needs its own axiom.

One example is inaccessible cardinals. If a cardinal k is strongly inaccessible, then for any cardinal c less than k, the set of its subsets 2^c is also smaller than k. Also k cannot be the union of fewer than k smaller sets. A powerset 2^c of a set c is a set by ZFC, as are unions of known sets, but these won't get you to an inaccessible according to this definition. Hence the term. But you could make an axiom that they exist.

There is an increasing hierarchy of larger and larger cardinals. But none can be created using the ones below it. A lot of them have definitions that are not intuitive or not obviously so large. Still, each has all the traits of the ones below it, and an axiom for it would imply that all the ones below it exist. All the large cardinals are inaccessible, for instance.

An important type of large cardinal is measurable cardinals. These have a measure, which is a definition of large vs. small subsets according to a few rules. The large ones are the cardinal itself and things like that, and small subsets include those containing a single element, plus a few more rules. If k is a measurable cardinal, the intersection of fewer than k large sets is also measurable. That makes measurable sound pretty distant from smaller sets. Measurable cardinals have some relationship to forcing axioms.

Schatz reports a particularly critical aspect due to Silver (1971): if a measurable cardinal exists, the constructible sets as a whole are a very weak part of V, and cannot even distinguish among uncountable cardinals. That result created a detour in the inner model program. Eventually a sort of ok weakening of constructability emerged that still provides information about the makeup of the sets, and could be extended up to measurable cardinals. Schatz reports that they found a "canonical, weakly core inner model of a measurable." Doing this gets more difficult for the larger cardinals beyond measurable and such research is still underway.

A very large cardinal type is the supercompact cardinals. They require advanced notions to define but are generalizations of measurable cardinals. Assuming they exist is a very strong assumption and leads to a wide array of results. Hamkins etal (2012) say " Assuming the existence of a supercompact is a very strong assumption and the large cardinal strength of supercompact cardinals is seen in a wide (and bewildering) array of set-theoretic contexts, especially the development of strong forcing axioms and establishing regularity properties of sets of reals." Schatz details some of this, like a complete characterization of the properties of the real numbers. Supercompacts are very useful for establishing consistency results for forcing axioms, and much more. The constructible approximation has not been built up yet for supercompact k, and Schatz discusses several potential barriers that stand in the way. But if the weak version of constructibles gets to the supercompact stage, Woodin (2017) shows that extending that further to all of V becomes almost automatic. This goal is called ultimate L.

The large-cardinal program has had a lot of progress and has produced results that are useful and illuminating. At least one large cardinal type is likely to survive as an axiom - maybe supercompact. That would give all the smaller ones as well, like measurables. One thing they do not do, however, is solve CH. It and its negation are not affected by large-cardinal axioms, it turns out. Basically, one more axiom is needed. Two are battling it out.

Two Surviving Axioms with Two Opposing Answers to CH

Research in the 1980's discovered more ways to implement forcing, such as by doing many forcings simultaneously. A powerful but risky axiom (requiring supercompact cardinals for consistency) called the proper forcing axiom, or PFA, led to some breakthroughs. In particular, Stevo Todorčević (1989) showed that PFA proves $2^{N(0)}$ is N(2), which is one cardinal higher than the N(1) that CH postulates. This was the first natural axiom that found CH to be false, and in a specific and surprising way. Schatz reports that it implied that "the continuum is as small as it could possibly be without CH being true. Such a size for the continuum was not widely seen as very plausible at this time; if CH was false, it was often said, it would be wildly false, allowing for a wide scope of interesting behavior of the many uncountable cardinals below the size of the continuum. … (W)ith this discovery, a naturally arising axiom candidate for not-CH had been discovered. Since this axiom provided for a 'small' continuum, with Todorčević's proof the focus of research on not-CH began to shift away from the supposed large values for the continuum. The importance of this event is quite hard to overstate: while the inner model program had provided axiom candidates implying the truth of CH, it is only with the discovery of PFA's consequences for the size of the continuum that proponents of not-CH had a parallel axiom to defend. It is only at this point that forcing axioms start to become considered as serious axiom candidates."

In a parallel development, Foreman, Magidor, and Shelah (1988) found a maximal extension of PFA, which they called Martin's Maximum (MM). The name comes from the fact that any further addition of forcing posets beyond it would create contradictions. It seems that using multiple posets simultaneously lets you postulate the existence of multiple properties of objects. In a sense MM implies that forcing is complete: nothing further can be added by using forcing. It postulates the richest and most complex universe of sets that you can create with forcing axioms. MM also leads to $2^{N(0)}$ = N(2), and it proves more theorems about large cardinals. It is not really much stronger than PFA, though, and its consistency also depends on supercardinals. An extension of it, MM++, allows for more detailed specification of the forcing conditions.

A more recent very strong general axiom called (*) asserts that sets defined in a specific but very broad manner exist. This is another way to define new sets beyond ZFC and does it in a way that looks more specific than forcing. It allows definition of sets y according to the formula: for every x for some y the property F(x,y) holds, where the property F is defined over the sets already specified. Wolchover (2021) gives an example: "for every set x of N(1) reals, some real y is not in x." This says the reals are bigger than N(1).

(*) is developed using an axiom inconsistent with AC and so was believed to be inconsistent with MM, even though it also gives $2^{N(0)}$ = N(2) and many of the other results of MM. Wolchover reports that Woodin found a way to use forcing to expand the (*) universe to one that was consistent with AC. That produced MM++. So (*) and MM++ could be consistent with each other.

It was still a confusing situation for set theorists. In a recent exciting paper, Asperó and Schindler (2021) prove that (*) and MM++ are in fact equivalent. That was both surprising and comforting, and has led to a lot of enthusiasm towards accepting them as the right axiom, and finally agreeing that $2^{N(0)}$ = N(2). Oddly enough, this axiom implies that $2^{N(1)}$ = N(2) as well, which is CH-like.

There is still work going on that would support CH, however. It basically comes down to the axiom V = ultimate L, i.e., all sets are weakly constructible, and it implies CH. As noted above, this is contingent on being able to formally specify ultimate L, which would result from being able to characterize L_k for a supercompact k, which is daunting. In general, CH holds when the universe is limited to constructible sets, even in their weak form.

The CH axioms do not produce the rich, complex universe of sets that (*) = MM++ does. The ultimate L proponents argue that set theory is basically about large cardinals, and being able to completely characterize them as a specific, sort of constructible, structure would complete set theory. But V = ultimate L is not in the form of a rule for defining new sets like the other axioms are. It is more like "stop here - we're done." It is a different form of statement than the ZFC axioms, or MM++ or (*).

Deciding

Schatz explores in detail a test of alternative axiomatizations proposed by Penelope Maddy (1997) called "maximization." It's a way to compare the results produced by different proposed axioms as well as seeing how much insight each of them gives into the results of the other. The methodology is beyond the scope of this note, but Schatz' conclusion is that MM++ wins easily over V = ultimate L.

Basing such a conclusion on the *results produced* by the axioms instead of their obvious truth is what Gödel (1947) calls the "extrinsic" approach. Schatz details arguments of Zermelo, Russell, Gödel, Cantor, Hilbert, etc. in the late 1800s and early 1900s that were of this nature. The result of being able to formalize all of mathematics in set theory is the goal, and so is more important than an axiom being obviously true. Informally, the test is "What's that axiom do for me, anyway?" The varied results coming from MM++ are impressive, and they provide an expansive universe of sets that theorists find meaningful and illuminating.

Ultimate L would be a wonderful and very interesting thing to have. It would give a picture of the structure of supercompact cardinals, or at least some part of them. But that would still be true if ultimate L were a proper subclass of V instead of being all of it. It is not clear why someone might prefer it being all there is. With it you could say that now we have explained all of set theory. Job completed. You might feel that under MM++ you get more in V, but all that is unexplained and unknowable stuff that is better off being ignored. Hard to say what the actual motivations are. These sound like arguments against using the extrinsic maximize test. In the end there may not be unanimity it what the axioms should be.

Discussion

In the late 1960s when everyone was still digesting Cohen's results, I took an undergraduate class in set theory at UC Berkeley. Our professor, Julia Robinson, was a stunningly brilliant mathematician and the wife and former student of the famous number theorist Rafael Robinson. She told us up front that this class would not get as far as forcing or the CH. But one day after class a few of us corralled her to get her feelings about it. She told us that when CH was proved to independent of ZFC, she at first thought that an expansive view would end up dominant, with the cardinality of the reals pretty high - the large continuum view that was still prominent 20 years later. But she said when she thought more about it, a higher cardinal for the reals just meant the existence of fewer 1:1 maps between the reals and lower cardinals. It did not mean that there were more sets, just that there were fewer mappings. Having more cardinals between N(1) and the continuum didn't make the continuum bigger, it just meant that the mappings were more restricted. And the mappings are sets themselves - ordered pairs of elements of the two sets. That left her with no clear intuition about what was right.

It is interesting that Gödel felt intuitively that the reals were N(2), although he wasn't able to convince many people of this. Maybe he was thinking about mappings as well.

The mapping issue applies to the cardinality of any set X. It is N(c) if there is an 1:1 mapping of X with N(c). The existence or nonexistence of such a mapping is not purely a property of X. It has to do with the other sets that may exist - like the mappings. We know what the reals are - all the infinite sequences of digits. Every such sequence is a real number under MM++ = (*) or V = ultimate L, and there are no other reals but these in either case. It would be hard to accept that there are different real numbers under those axioms. Yet they have different cardinalities. Thus the cardinality of the reals tells us more about the other sets in the universe given by those axioms, i.e., whether or not there exist 1:1 mappings between N(1) and the reals. Maybe N(1), the set of all countable ordinals, is different in the two universes due to the existence or not of mappings. What's countable might be up for grabs.

In either case, the cardinality of the reals is not an intrinsic property of that set but depends on other things going on in the universe where it is situated. That makes cardinality a less satisfactory measure of the size of a set than one may have hoped. Measuring the sizes of infinite sets is not totally straightforward. The choice of axiom systems largely is about the characteristics of universe of sets they create, and cardinality of power sets, etc. is a result.

The reals are becoming less essential for calculus and analysis anyway. Using set theory and model theory, Abraham Robinson (1966) details an approach to calculus and real analysis based on infinitesimals. That is what Newton and Leibnitz used informally, but it had never been properly developed axiomatically. The limits approach was developed as an alternative. Infinitesimals are easier and more intuitive than limits. A positive infinitesimal is greater than 0 but less than 1/n for every positive integer n. Their reciprocals, unbounded numbers, are greater than any integer but less than $\infty = 1/0$.

My feeling is that on extrinsic grounds, I like MM++ = (*) plus supercompact cardinals. Those axioms produce a rich universe with discoverable properties, especially about large sets. It seems fine that there N(2) reals. Since cardinality is not purely a property of the reals this is hard to worry much about.

The notion of contradiction weaves through all these discussions. That's what led Zermelo to develop the ZFC axioms in the first place. Gödel's incompleteness proof came right up to the edge of contradiction but stepped back to find that if the axioms are consistent, they must be incomplete and their consistency unprovable. MM++ itself is one step away from contradiction: it is the furthest you can push forcing without creating inconsistency. MM++ and (*) are based on assumptions that are inconsistent with each other, but somehow the universe of sets can be expanded enough to make the two consistent. Anything can be proved from inconsistent axioms. The closer you get to that, the more you can prove.

Mathematics is full of rich and mysterious systems that come from a few simple concepts. The distribution of prime numbers comes from a simple definition but has open questions and more is still being discovered. The "bewildering" scope of results about large cardinals that arise in MM++ is like that. Out of a few simple rules, a vast complex universe of sets comes out - a playground for set theorists to explore while skirting the edge of impossibility.

References

Asperó, David. Ralf Schindler. 2021. Martin's Maximum++ implies Woodin's axiom (∗). Annals of Mathematics (2) 193 (3) 793 - 835.

Cohen, Paul Joseph 1963. The independence of the continuum hypothesis. Proceedings of the National Academy of Sciences of the United States of America. 50 (6): 1143–1148.

Devlin, Kevin J. 1974. An introduction to the fine structure of the constructible hierarchy. Studies in Logic and the Foundations of Mathematics. 79:123-163.

Foreman, M., M. Magidor, and S. Shelah. 1988. Martin’s maximum, saturated ideals, and nonregular ultrafilters. I. Annals of Math.\ematics (2) 127 (1988): 1–47.

Fraenkel, Abraham. 1919. Einleitung in die Mengenlehre. Berlin: Julius Springer

Gentzen, Gerhard. 1936. Die Widerspruchsfreiheit der reinen Zahlentheorie. Mathematische Annalen. 112: 493–565. doi:10.1007/BF01565428, S2CID 122719892 – Translated as "The consistency of arithmetic", in (Gentzen & Szabo 1969).

Gentzen, Gerhard. 1969. Szabo, M. E. (ed.). Collected Papers of Gerhard Gentzen. Studies in logic and the foundations of mathematics. Amsterdam: North-Holland.

Gödel, Kurt. 1930. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, Monatshefte für Mathematik und Physik, v. 38 n. 1, pp. 173–198.

Gödel, Kurt. 1931. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I, in Solomon Feferman, ed., 1986. Kurt Gödel Collected works, Vol. I. Oxford University Press, pp. 144–195.

Gödel, Kurt. 1940. The Consistency of the Continuum Hypothesis. Annals of Mathematics Studies. 3. Princeton: Princeton University Press.

Gödel, Kurt. 1945. Some considerations leading to the probable conclusion that the true power of the continuum is $\aleph_2$ in in Solomon Feferman, ed., 1995. Kurt Gödel Collected works, Vol. II, Publications 1938-1974, Oxford University Press, pp. 420-422.

Gödel, Kurt. 1947. What is Cantor's continuum problem? American Mathematical Monthly. 54: 515–525.

Gödel, Kurt. 1951, "Some basic theorems on the foundations of mathematics and their implications", in Solomon Feferman, ed., 1995. Kurt Gödel Collected works, Vol. III, Oxford University Press, pp. 304–323.

Hamkins, Joel David, Victoria Gitman and contributors. 2012. "Supercompact cardinal." Cantor's Attic. https://neugierde.github.io/cantors-attic/Supercompact.

Maddy, Penelope. 1997. Naturalism in Mathematics. Oxford University Press.
Mates, Benson. 1953. Stoic Logic. University of California Press.

Robinson, Abraham. 1966. Non-standard analysis. North-Holland Publishing Co., Amsterdam.

Russell, Bertrand. 1903. The Principles of Mathematics. Cambridge University Press, United Kingdom.

Schatz, Jeffrey Robert. 2019. Axiom Selection by Maximization: V = Ultimate L vs Forcing Axioms. UC Irvine. https://escholarship.org/uc/item/9875g511 .

Skolem, Thoralf. 1922. Some Remarks on Axiomatized Set Theory.

Silver, Jack H. 1971. Some applications of model theory in set theory. Annals of Mathematical Logic. 3: 45–110.

John R. Steel. 2014. Gödel's Program. In Kennedy, J., editor, Interpreting Gödel - Critical Essays. Cambridge University Press. 153 - 179.

Tarski, Alfred and Givant, Steven. 1999. Tarski's system of geometry. The Bulletin of Symbolic Logic. 5 (2): 175–214.

Todorčević, Stevo. 1989. Partition problems in topology. Volume 84 of Contemporary Mathematics. American Mathematical Society, Providence, RI.

Wikipedia contributors. 2023. Constructible universe. Wikipedia, The Free Encyclopedia. December 31.

Wikipedia contributors. 2024. Forcing (mathematics). Wikipedia, The Free Encyclopedia.

Wolchover, Natalie. 2013. "To Settle Infinity Dispute, a New Law of Logic". Quanta Magazine, Simons Foundation, USA. 11/26. https://www.quantamagazine.org/to-settle-infinity-question-a-new-law-of-mathematics-20131126/

Wolchover, Natalie. 2021. "How Many Numbers Exist? Infinity Proof Moves Math Closer to an Answer". Quanta Magazine, Simons Foundation, USA. 07/15. https://www.quantamagazine.org/how-many-numbers-exist-infinity-proof-moves-math-closer-to-an-answer-20210715/

Woodin, W. Hugh. 2017. "In Search of Ultimate-L the 19th Midrasha Mathematicae Lectures." The Bulletin of Symbolic Logic. 23 (01) (March): 1–109.

Zermelo, Ernst. 1908. "Untersuchungen über die Grundlagen der Mengenlehre I". Mathematische Annalen. 65 (2): 261–281.

Zermelo, Ernst. 1930. "Über Grenzzahlen und Mengenbereiche". Fundamenta Mathematicae. 16: 29–47.