

THE INFINITE

THOMAS JECH

The Pennsylvania State University

Until the 19th century the study of the infinite was an exclusive domain of philosophers and theologians. For mathematicians, while the concept of infinity was crucial in applications of calculus and infinite series, the infinite itself was (to paraphrase Gauss) just “the manner of speaking.” Indeed, 18th century mathematics, even with its frequent use of infinite sequences and infinitesimals, had no real need to study the concept of infinity itself.

The increasing use of abstract functions of real and complex variable and the emerging theory of real numbers in the work of Cauchy, Weierstrass and Dedekind brought about the necessity to use infinity not only as “a manner of speaking” but rather introduce into mathematical usage the concept of infinite set.

The earliest mathematician to study sets as objects of a mathematical theory was Bernhard Bolzano in the 1840’s. He gave a convincing philosophical argument for the existence of infinite sets and was also first to explicitly formulate the concept of one-to-one correspondence between sets. It seems that Bolzano was not aware of the possibility that there may exist infinite sets of different size, and that (as some historians believe) he envisioned a mathematical universe where all infinite sets are countable. While some of his ideas were clearly ahead of his time, he did

Research for this article was supported in part by a grant from the International Research & Exchange Board (IREX), with funds provided by the Andrew W. Mellon Foundation, the National Endowment for the Humanities, and the U.S. Department of State. None of these organizations is responsible for the views expressed. The author is grateful to the Kurt Gödel Society in Vienna and to the Center for Theoretical Studies in Prague for the opportunity to discuss the topics of this article. I am particularly thankful to Professor Petr Vopěnka for his interpretation of Bolzano’s work .

consider paradoxical the fact that the set of all positive integers can be in a one-to-one correspondence with a “smaller” subset, a fact that had been already pointed out by Galileo centuries earlier.

It was Georg Cantor in the 1870’s who realized the significance of one-to-one correspondence and who embarked on a systematic study of abstract sets and of cardinal and ordinal numbers. His was also the first application of set theoretical methods, namely his “nonconstructive” proof of the existence of transcendental numbers. The significance of Cantor’s proof lies in the fact that it employs, in its proof of uncountability of \mathbf{R} , the idea that any, *arbitrary*, bounded increasing sequence of real numbers has a least upper bound. This assumption is related to the concept of real numbers as Dedekind cuts in \mathbf{Q} , or as limits of Cauchy sequences of rationals, or as infinite decimal expansions. These three possible approaches to the definition of real numbers are all equivalent (in a sense) and all require the existence of *arbitrary* infinite sets of integers.

Very few mathematicians object to the idea of an arbitrary set of integers. We will argue later that the idea of existence of “large” cardinal numbers that some mathematicians find objectionable (the “never-never land” in the words of one detractor) is conceptionally no different from Cantor’s idea that is inherent in his proof.

A natural question arising from Cantor’s theory of cardinal numbers is the following: Do there exist sets of real numbers that are uncountable, yet not equivalent in cardinality to the set of all real numbers? The conjecture that no such sets exist has become known as the *Continuum Hypothesis*. Hilbert attached such importance to this question that it figured on the top of his famous list of problems in 1900.

The continuum problem has eventually been solved in a nontraditional way. It was proved that the continuum hypothesis is neither provable nor refutable, thus being *independent*.

To show that a mathematical statement is independent, one has to specify a

system of postulates, *axioms* of the theory, which are then shown to be insufficient to prove or to refute the statement. For instance, the parallels postulate is independent of the rest of Euclid's postulates. The proof of independence is accomplished by exhibiting *models* (such as the Bolyai-Lobačevskij model of non-Euclidean geometry). So clearly the assertion that the continuum hypothesis is independent requires that we specify the axiom system.

There is a more fundamental reason for axiomatizing set theory. With very few insignificant exceptions, all mathematical arguments and concepts can be, in principle, reduced to concepts and statements formalized in the language of set theory. Thus an axiom system for the theory of sets is in fact an axiom system for all of mathematics.

The axiomatization of set theory based on Cantor's "definition" of sets turns out to be inconsistent, for rather trivial reasons. (The idea is to consider as sets all collections of objects that satisfy an arbitrary "property". This approach is clearly untenable, as exemplified by the "set"

$$\{a : a \notin a\}$$

whose existence leads to a logical contradiction.)

Since the 1920's, the widely accepted system of axioms for set theory is the Zermelo-Fraenkel axiomatic set theory (ZF). It consists of several axioms (and axiom schemata) that encompass all uses of sets in mathematics. This axiomatic theory is almost universally accepted as the logical foundation of current mathematics.

Most axioms of Zermelo-Fraenkel are construction principles that merely formalize certain manipulations with sets. The two notable exceptions are the *axiom of infinity* and the *axiom of choice*.

The axiom of infinity postulates the existence of an infinite set or, equivalently, the existence of the set \mathbf{N} of all natural numbers. This axiom is of course the essence of set theory. If we replace the axiom of infinity by its negation, we get a

theory equivalent to *Peano arithmetic*, the axiomatic theory of elementary number theory.

Historically, the most interesting axiom of ZF is the axiom of choice. Unlike the other axioms, it is highly nonconstructive, as it postulates the existence of choice functions without giving a specific description of such functions. It was introduced by Zermelo in his proof of well-ordering, and is in fact equivalent to the principle that states that *every* set can be well ordered. As it seems impossible to *define* a well ordering of the set \mathbf{R} , among others, this principle was violently opposed by some mathematicians of the first half of this century.

Some of the objections are quite understandable in light of such consequences of the axiom of choice as the Hausdorff-Banach-Tarski paradoxical decomposition of the unit ball. It should be noted that the axiom of choice is independent of the rest of ZF (by the work of Gödel and Cohen), which puts set theory in situation roughly similar to that of Euclidean v. non-Euclidean geometry. From the formalist's point of view it does not make any difference whether the axiom of choice is accepted as an axiom or not. In view of its usefulness in algebra and analysis, the prevailing opinion favors the axiom. (I would like to point out that even such an innocuous statement as "the union of countably many countable sets of real numbers is countable" requires the axiom of choice).

It is of these Zermelo-Fraenkel axioms that the continuum hypothesis was shown to be independent. First, in 1938, Kurt Gödel constructed a model of set theory ("the constructible universe") in which the continuum hypothesis holds. Gödel's model is a canonical model; it is the least possible model, which any other model (that contains all ordinal numbers) has to include. In 1963, Paul Cohen discovered a method of constructing "generic extensions" of models of set theory, including models in which the continuum hypothesis fails. Cohen's method turned out to be quite general and resulted in a number of independence proofs. It so happens that a number of conjectures in different areas of mathematics cannot be decided on the

basis of Zermelo-Fraenkel axioms, and have been proved independent by Cohen's method. (In addition to the continuum hypothesis we mention Suslin's Problem, Borel's Conjecture, Whitehead's Problem and Kaplansky's Conjecture.)

In view of the apparent incompleteness of Zermelo-Fraenkel axioms a reasonable suggestion offers itself, namely to find a complete system of axioms, in which every possible statement of mathematics would be formally decidable. This was the so called *Hilbert's Program*. A mortal blow to this program were the two *Incompleteness Theorems* of Gödel.

The First Incompleteness Theorem shows in effect that there is no hope for a complete axiomatization of mathematics. It states that *every* axiomatic theory that includes elementary arithmetic is incomplete (if it is consistent). In particular, both Peano arithmetic and Zermelo-Fraenkel set theory are incomplete, and will remain incomplete even if we add further axioms.

Gödel's proof of the Incompleteness Theorem yields an undecidable sentence, but unlike the Continuum Hypothesis in set theory, Gödel's sentence is not particularly meaningful. It does provide a starting point in the search for easily understandable statements expressible in the language of arithmetic that are independent of Peano arithmetic. Such statements have indeed been found, by Harrington and Paris. Their examples are combinatorial principles, expressible in the language of arithmetic, and true (i.e. provable in set theory) but unprovable in Peano arithmetic. It remains to be seen whether some of the unproved conjectures in elementary number theory may perhaps be examples of undecidable statements.

The Second Incompleteness Theorems dashes all hopes that the axiom of infinity can ever be justified by elementary means. The theorem states that in every axiomatic theory that includes Peano arithmetic, the statement that the theory itself is consistent (which is expressible in Peano arithmetic)¹ is unprovable (if the theory is consistent). Thus Peano arithmetic cannot prove its own consistency, and

¹The statement can be formulated roughly as follows: There exists no finite sequence of sentences that constitutes a formal proof of a contradiction.

neither can Zermelo-Fraenkel. On the other hand, the axiom of infinity provides, in set theory, a way to prove the consistency of arithmetic (or of set theory without the axiom of infinity).

Using the infinite set \mathbf{N} , one can define the usual operations $+$ and \cdot and thus obtain a model of Peano arithmetic, thereby establishing its consistency. As Peano arithmetic (or the theory of finite sets which is equivalent to it) cannot prove the consistency of Peano arithmetic, and set theory does, it is a consequence of Gödel's Second Theorem that the axiom of infinity *cannot* be proved consistent relative to the other axioms of Zermelo-Fraenkel. Therefore, our acceptance of the existence of infinite sets, even of infinite sets of integers, is an act of faith, faith that set theory is not contradictory.

The theory of cardinal numbers initiated by Cantor some 100 years ago proved to be unusually rich and profound. While the basic arithmetic of cardinals, addition and multiplication, follows very simple rules, exponentiation of cardinals makes an interesting mathematical theory. Already the Continuum Hypothesis is a problem in cardinal exponentiation. Since the cardinality of \mathbf{R} is 2^{\aleph_0} where \aleph_0 is the cardinality of \mathbf{N} , the continuum hypothesis can be stated as $2^{\aleph_0} = \aleph_1$, where \aleph_1 is the least uncountable cardinal.

A central problem of modern set theory is to describe the behavior of the cardinal function 2^{\aleph_α} . One outcome of the work following Cohen's discovery was a solution of this problem for *regular* cardinals \aleph_α (Easton's Theorem): the result is that subject to some trivial constraints, the value of 2^{\aleph_α} when \aleph_α is regular is quite arbitrary. In contrast to Easton's theorem, the problem of calculating 2^{\aleph_α} when \aleph_α is *singular* (the *Singular Cardinals Problem*) is far from being solved, despite some spectacular results (Silver, Magidor, Shelah).

The cardinal number of an infinite set A is called *singular* if A is the union of a smaller number of smaller sets, i.e. $A = \bigcup_{i \in I} A_i$ where both I and each A_i have a smaller cardinality than A . Cardinals that are not singular are *regular*. For

instance, \aleph_0 is regular: the set \mathbf{N} is not the union of finitely many finite sets. On the other hand, the cardinal $\aleph_\omega = \sup\{\aleph_n : n = 0, 1, 2, \dots\}$ is singular, as it is the limit of a countable sequence of smaller cardinals.

Which brings us to the most important issue in present day set theory: *large cardinals*. When we consider a typical limit cardinal, i.e. a cardinal \aleph_α whose index α is a limit ordinal, such as $\aleph_\omega, \aleph_{\omega_1}, \aleph_{\omega_2+\omega}$ and so on, we notice that such cardinal is singular. However, as Hausdorff already observed some 70 years ago, it is not inconceivable that a limit cardinal could be regular. Hausdorff called such cardinals *inaccessible*. Inaccessible cardinals, if they exist, have to be extremely large, much greater than cardinals that come up in ordinary mathematical practice. Moreover, their existence is unprovable in Zermelo-Fraenkel.

The fact that inaccessible cardinals cannot be proved to exist is a manifestation of Gödel's Second Incompleteness Theorem. If we assume that an inaccessible cardinal exists then we can *prove* that set theory is consistent. Thus not only one cannot prove in ZF that an inaccessible exists, but one cannot even prove that its existence is consistent. The situation is analogous to the proof of consistency of finite set theory from the axiom of infinity. Thus adopting inaccessible cardinals constitutes a similar act of faith as adopting infinite sets.

The most frequent argument against the study of large cardinals raised by non-specialists is that the sets in question are so large that they have no relevance to problems arising in "mainstream" mathematics. (Thus the "never-never land" label.) This argument is hardly convincing, considering the numerous discoveries that have provided ample evidence of close relationship between large cardinals and diverse mathematical problems. The case in point is the *measure problem*. Specifically, consider the question whether Lebesgue measure can be extended to a σ -additive measure defined on all subsets of the real line. By the work of Ulam and Solovay, this question is equiconsistent with the existence of a certain large cardinal, called *measurable*. And a measurable cardinal is inaccessible, and in fact

its existence is a considerably stronger postulate than the existence of inaccessibles.

The relationship of measurable v. inaccessible cardinals is prototypical in the bestiary of large cardinals. Firstly, every measurable cardinal is inaccessible. Secondly, the least inaccessible cardinal is not measurable, and in fact every measurable cardinal has below it a cofinal set of inaccessibles. And thirdly (appealing to Gödel's Second Incompleteness Theorem), the consistency of the existence of a measurable cardinal is unprovable in the theory that postulates the existence of inaccessible cardinals. The theory of large cardinals leads to a plethora of large cardinals that are more or less arranged in a hierarchy where the relationship between two kinds of large cardinals is like the relationship between measurables and inaccessibles described above.

The large cardinals investigated in set theory arise either by generalizing properties of other large cardinals or by analyzing problems in other areas of mathematics, mostly measure theory, model theory or combinatorics. The most convincing evidence for large cardinals, however, comes from *descriptive set theory*, the study of “nice” sets of real numbers.

When investigating various properties of sets of reals such as Lebesgue measurability, it is useful to distinguish between arbitrary sets and those sets that are definable; the definable sets can be arranged in a natural hierarchy: Borel, analytic, projective etc. (We note in passing that the assumption that all projective sets are Lebesgue measurable is equiconsistent with the existence of inaccessible cardinals.) The main tool of descriptive set theorists is *determinacy of infinite games*. One of the successes of modern set theory was the proof that the hierarchy of axioms arising from determinacy (by varying the complexity of the payoff set) corresponds very closely to the hierarchy of large cardinal axioms. These results have inextricably tied large cardinals with properties of sets of real numbers.

As a result of Gödel's Theorem, it will never be possible to prove consistency of large cardinals. And one cannot of course exclude the possibility that some of the

large cardinals may turn out to be inconsistent. In the rush to generalize various large cardinal properties in the early 70's, a certain axiom was proposed that later was refuted (a result of Kunen). It seems unlikely however that a similar fate awaits other large cardinal axioms. Firstly, these axioms have been studied for a number of years, without getting a contradiction. Moreover, the theory leads to an elegant structure theory of the set theoretic universe, admitting canonical models for large cardinals, not unlike Gödel's model of constructible sets. Secondly, large cardinal properties arise in seemingly unrelated areas. We already mentioned descriptive set theory; it should also be noted that the Singular Cardinals Problem is very closely tied to the large cardinal theory.

To conclude this discourse I would like to point out the following. The statement that the existence of some large cardinal (say measurable or supercompact or whatever) is consistent, is expressible as a sentence in the language of arithmetic. Thus we have an arithmetical sentence that is unprovable in set theory even under the assumption that the large cardinal in question exists, while provable under the assumption of some bigger large cardinal. Of course, this arithmetical sentence has no transparent meaning, but it is not inconceivable that a meaningful statement might be discovered in the future (in analogy with the Harrington-Paris result for Peano arithmetic). Thus it is conceivable that one might formulate a conjecture of elementary number theory or in finite mathematics, which would be provable under the assumption of some very large cardinal but not without it.

While this is all very speculative, I wish to mention a recent result of Richard Laver. Using a large cardinal assumption just a tad weaker than the one refuted by Kunen, Laver gives a decision procedure for the word problem for the free algebra with one generator and one left-distributive binary operation. So here is a problem of finite mathematics for which the only known solution employs one of the largest cardinals that set theorists ever invented.