Non-Cantorian Set Theory

In 1963 it was proved that a celebrated mathematical hypothesis put forward by Georg Cantor could not be proved. This profound development is explained by analogy with non-Euclidean geometry

by Paul J. Cohen and Reuben Hersh

The abstract theory of sets is currently in a state of change that in several ways is analogous to the 19th-century revolution in geometry. As in any revolution, political or scientific, it is difficult for those participating in the revolution or witnessing it to foretell its ultimate consequences, except perhaps that they will be profound. One thing that can be done is to try to use the past as a guide to the future. It is an unreliable guide, to be sure, but better than none.

We propose in this article to use the oft-told tale of non-Euclidean geometry to illuminate the now unfolding story of nonstandard set theory.

A set, of course, is one of the simplest and most primitive ideas in mathematics, so simple that today it is part of the kindergarten curriculum. No doubt for this very reason its role as the most fundamental concept of mathematics was not made explicit until the 1880's. Only then did Georg Cantor make the first nontrivial discovery in the theory of sets.

To describe his discovery we must first explain what we mean by an infinite set. An infinite set is merely a set with an infinite number of distinct elements; for example, the set of all "natural" numbers (1, 2, 3 and so on) is infinite. So too is the set of all the points on a given line segment.

Cantor pointed out that even for infinite sets it makes sense to talk about the number of elements in the set, or at least to state that two different sets have the same number of elements. Just as with finite sets, we can say that two sets have the same number of elements—the same "cardinality"—if we can match up the elements in the two sets one for one. If this can be done, we call the two sets equivalent.

The set of all natural numbers can be matched up with the set of all even numbers, and also with the set of all fractions [*see illustration below*]. These two examples illustrate a paradoxical property of infinite sets: an infinite set can be equivalent to one of its subsets. In fact, it is easily proved that a set is infinite if, and only if, it is equivalent to some proper subset of itself.

All of this is engaging, but it was not new with Cantor. The notion of the cardinality of infinite sets would be interesting only if it could be shown that not all infinite sets have the same cardinality. It was this that was Cantor's first great discovery in set theory. By his famous diagonal proof he showed that the set of natural numbers is *not* equivalent to the set of points on a line segment [*see illustration on opposite page*].

Thus there are at least two different kinds of infinity. The first, the infinity of the natural numbers (and of any equivalent infinite sets), is called aleph nought ($\aleph_0$). Sets with cardinality $\aleph_0$ are called countable. The second kind of infinity is the one represented by a line segment. Its cardinality is designated by a lowercase German c ($\mathfrak{c}$), for "continuum." *Any* line segment, of arbitrary length, has cardinality $\mathfrak{c}$ [*see illustration on page 106*]. So does any rectangle in the plane, any cube in space, or for that matter all of unbounded n-dimensional space, whether n is 1, 2, 3 or 1,000!

Once a single step up the chain of infinities has been taken, the next follows naturally. We encounter the notion of the set of all subsets of a given set [*see illustration on page 111*]. If the

2,	4,	6,	8,	10,	12,	14,	16,	18,	20,	22,	...
1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11,	...
1/1,	1/2,	2/1,	1/3,	2/2,	3/1,	1/4,	2/3,	3/2,	4/1,	1/5,	...

SET IS TERMED COUNTABLE if it can be matched one for one with the natural numbers (*middle row*). Thus the set of all even numbers (*top row*) is countable. The set of all fractions (*bottom row*) is also countable. The fractions shown here are the ones used by the German mathematician Georg Cantor (1845–1918); they are not in their natural order but in order according to the sum of the numerator and the denominator. Both examples show that an infinite set, unlike a finite set, can be equivalent to one of its subsets.

original set is called A, this new set is called the power set of A and is written 2^A. And just as we obtain the power set 2^A from A, we can next obtain $2^{(2^A)}$ from 2^A, and so on as long as we please.

Cantor proved that whether A is finite or infinite, 2^A is never equivalent to A. Therefore the procedure of forming the set of all subsets generates an endless chain of increasing, nonequivalent infinite sets. In particular, if A is the set of natural numbers then it is easy to prove that 2^A (the set of all sets of natural numbers) is equivalent to the continuum (the set of all points on a line segment). In brief,

$$2^{\aleph_0} = c.$$

At this point a question may occur to the reader. Is there an infinite set with cardinality *between* $\aleph_0$ and c? That is, is there on a line segment an infinite set of points that is not equivalent to the whole segment, and also not equivalent to the set of natural numbers?

This question occurred to Cantor, but he was unable to find any such set. He concluded—or rather conjectured—that no such thing exists. This guess of Cantor's acquired the name "the continuum hypothesis." Its proof or disproof was first on the celebrated list of unsolved mathematical problems drawn up by David Hilbert in 1900. Only in 1963 was it finally settled. It was settled, however, in a sense utterly different from what Hilbert had in mind.

To tackle this problem one could no longer rely on Cantor's definition of a set as "any collection into a whole of definite and separate objects of our intuition or our thought." In fact, this definition, seemingly so transparent, turned out to conceal some treacherous pitfalls. An instance is the sad experience suffered by Gottlob Frege in 1902. Frege was about to publish a monumental work in which arithmetic was reconstructed on the foundation of set theory, that is, on the foundation of "intuitive" set theory as it was then known on the basis of Cantor's work. At this point Frege received a letter from the young Bertrand Russell, which he acknowledged by adding this postscript to his treatise: "A scientist can hardly meet with anything more undesirable than to have the foundation give way just as the work is finished. In this position I was put by a letter from Mr. Bertrand Russell as the work was nearly through the press."

Russell's blow consisted in pointing out a simple conundrum. There are two kinds of sets. First there are those, such as "the set of all objects describable in

1. .18347984639001...
2. .36948570110924...
3. .50472200173996...
4. .99801230109487...
5. .00102305497610...
6. .51546798371238...
7. .55119871350426...

⋮ ⋮

SET OF REAL NUMBERS IS UNCOUNTABLE, as Cantor showed in his famous diagonal proof. Here a sample of the set is listed in decimal form and at random. If one takes the first digit of the first number, the second digit of the second and so on *(color)*, one obtains a real number whose infinite decimal expansion is .1640277.... If one randomly changed every digit in the expansion, one might get .2751388.... A moment's thought will show that the new number is different in at least one place from every number on the list. Hence the number was not present on the list, and it has been proved that the list was incomplete.

exactly 11 English words," having the peculiar property that they themselves satisfy their defining property; in other words, sets that contain themselves as elements. We call them R sets, the R standing for Russell. Then there are all other sets—sets that do not belong to themselves. Call them the non-R sets. Now, said Russell, consider the collection of all non-R sets. (The word "collection" is introduced here simply as a convenient synonym for "set.") Call this set M. Then M is either an R set or a non-R set. But if M is a non-R set, then it belongs to M, by definition of M, so that it is an R set, by definition of R sets. This is a contradiction. On the other hand, if M is an R set, then by definition of M it does not belong to M. It does not belong to itself, that is, it is not an R set, which is again a contradiction.

The moral is this: The free use of Cantor's intuitive notion of a set can lead to contradictions. Set theory can serve as a secure foundation for mathematics only if a more sophisticated approach is employed to steer clear of antinomies, as contradictions of the type proposed by Russell later came to be known.

It has happened before that unwelcome paradoxes have intruded into a seemingly clear mathematical theory. There are the paradoxes of Zeno, which revealed to the Greeks unsuspected complexities in intuitive concepts of lines and points. We can draw an analogy: As Russell found a contradiction in the unrestricted use of the intuitive concept of set, so Zeno had found contradictions in the unrestricted use of the intuitive concepts of "line" and "point."

In its beginning with Thales in the sixth century B.C. Greek geometry had relied on an unspecified intuitive concept of "line" and "point." Some 300 years later, however, Euclid had given these concepts an axiomatic treatment. For Euclid geometric objects were still intuitively known real entities, but insofar as they were the subject of geometrical reasoning they were specified by certain unproved assertions ("axioms" and "postulates"), on the basis of which all their other properties were supposed to be proved as "theorems." We do not know if, and to what extent, this development was a response to paradoxes such as Zeno's. There is no doubt, however, that to the Greeks geometry was made much more secure by virtue of depending (at least so they believed and intended) only on logical inference from a small number of clearly stated assumptions.

The analogous development for set theory took not 300 years but only 35. If

Cantor played the role of Thales—the founder of the subject, who was able to rely on intuitive reasoning alone—then the role of Euclid was played by Ernst Zermelo, who in 1908 founded axiomatic set theory. Of course, Euclid was really only one of a long succession of Greek geometers who created "Euclidean geometry"; so also Zermelo was only the first of half a dozen great names in the creation of axiomatic set theory.

Just as Euclid had listed certain properties of points and lines and had regarded as proved only those theorems in geometry that could be obtained from these axioms (and not from any possibly intuitive arguments), so in axiomatic set theory a set is regarded simply as an undefined object satisfying a given list of axioms. Of course, we still want to study sets (or lines, as the case may be), and so the axioms are chosen not arbitrarily but in accord with our intuitive notion of a set or a line. Intuition is nonetheless barred from any further formal role; only those propositions are accepted that follow from the axioms. The fact that objects described by these axioms actually may exist in the real world is irrelevant to the process of formal deduction (although it is essential to discovery).

We agree to act as if the symbols for "line," "point" and "angle" in geometry, or the symbols for "set," "is a subset of" and so on in set theory, are mere marks on paper, which may be rearranged only according to a given list of rules (axioms and rules of inference). Accepted as theorems are only those statements that are obtained according to such manipulations of symbols. (In actual practice only those statements are accepted that clearly *could* be obtained in this manner if one took enough time and trouble.)

Now, in the history of geometry one postulate played a special role. This was the parallel postulate, which says that through a given point there can be drawn precisely one line parallel to a given line. The difficulty with this statement as an axiom is that it does not have the self-evident character one prefers in the foundation stones of a mathematical theory. In fact, parallel lines are defined as lines that never meet, even if they are extended indefinitely ("to infinity"). Since any lines we draw on paper or on a blackboard have finite length, this is an axiom that by its nature cannot be verified by direct observation of the senses. Nonetheless, it plays an indispensable role in Euclidean geometry. For many centuries a leading problem in geometry was to *prove* the parallel postulate, to show that it could be obtained as a theorem from the more self-evident Euclidean axioms.

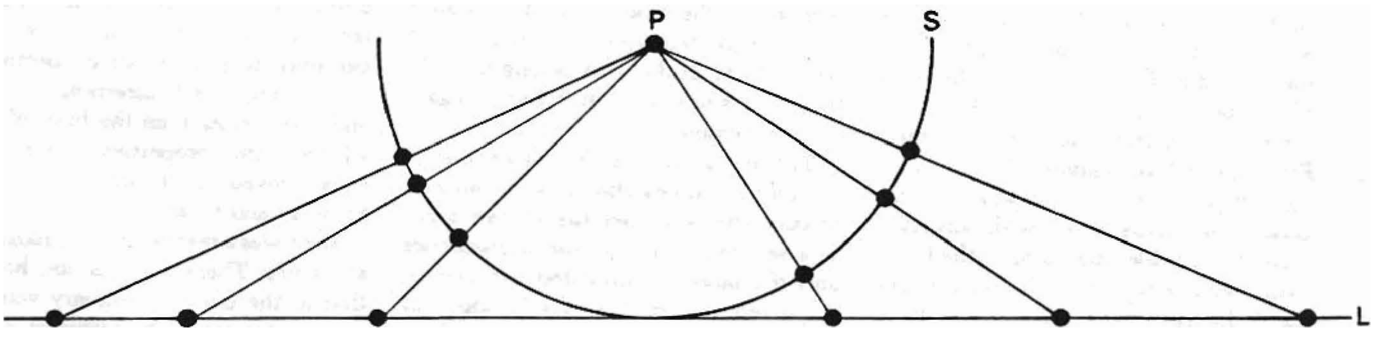
In abstract set theory, it so happens, there also was a particular axiom that some mathematicians found hard to swallow. This was the axiom of choice, which says the following: If α is any collection of sets $\{A, B, \ldots\}$, and none of the sets in α is empty, then there exists a set Z consisting of precisely one element each from A, from B and so on through all the sets in α. For instance, if α consists of two sets, the set of all triangles and the set of all squares, then α clearly satisfies the axiom of choice. We merely choose some particular triangle and some particular square and then let these two elements constitute Z.

Most people find the axiom of choice, like the parallel postulate, intuitively very plausible. The difficulty with it is in the latitude we allow α: "any" collection of sets. As we have seen, there are endless chains of ever bigger infinite sets. For such an inconceivably huge collection of sets there is no way of actually choosing one by one from all its member sets. If we accept the axiom of choice, our acceptance is simply an act of faith that such a choice is possible, just as our acceptance of the parallel postulate is an act of faith about how lines would act if they were extended to infinity. It turns out that from the innocent-seeming axiom of choice some unexpected and extremely powerful conclusions follow. For example, we are able to use inductive reasoning to prove statements about the elements in *any* set, in much the same way that mathematical induction can be used to prove theorems about the natural numbers 1, 2, 3 and so on.

The axiom of choice played a special role in set theory. Many mathematicians thought its use should be avoided whenever possible. Such a form of axiomatic set theory, in which the axiom of choice is *not* assumed to be either true or false, would be one on which almost all mathematicians would be prepared to rely. In what follows we use the term "restricted set theory" for such an axiom system. We use the term "standard set theory" for the theory based on the full set of axioms put forward by Zermelo and Abraham Fraenkel: restricted set theory *plus* the axiom of choice.

In 1938 this subject was profoundly illuminated by Kurt Gödel. Gödel is best known for his great "incompleteness" theorems of 1930–1931 [see "Gödel's Proof," by Ernest Nagel and James R. Newman; SCIENTIFIC AMERICAN, June, 1956]. Here we refer to later work by Gödel that is not well known to nonmathematicians. In 1938 Gödel proved the following fundamental result: If restricted set theory is consistent, then so is standard set theory. In other words, the axiom of choice is no more dangerous than the other axioms; if a contradiction can be found in standard set theory, then there must already be a contradiction hidden within restricted set theory.

But that was not all Gödel proved. We remind the reader of Cantor's "continuum hypothesis," namely that no in-



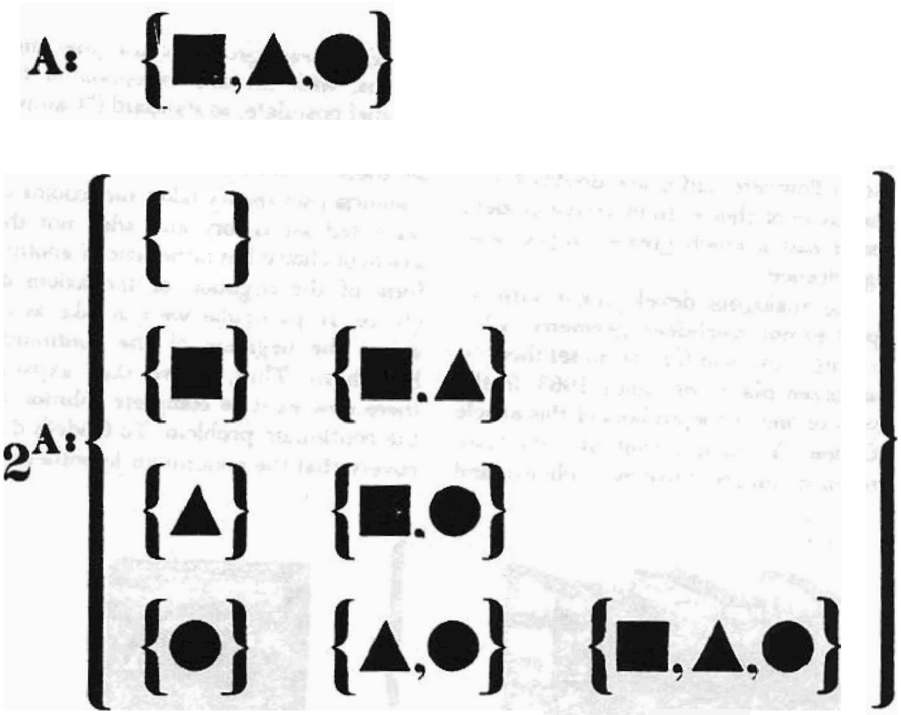
INFINITE LINE AND FINITE LINE SEGMENT can also be shown to have a one-to-one correspondence. Here P is the center of a semicircle S that is tangent to an infinite line L. A ray from P cuts S at only one point. In this way rays from P give a one-to-one match between points on S and points on L. As the ray changes direction from left to right no point is omitted from either S or L. Thus a one-to-one correspondence exists between the points on an infinite line and the points on a finite segment of arbitrary length.

finite cardinal exists that is greater than $\aleph_0$ and smaller than c. Gödel also showed that we can safely take the continuum hypothesis as an additional axiom in set theory; that is, if the continuum hypothesis plus restricted set theory implies a contradiction, then again there must already be a contradiction hidden within restricted set theory. This was a half-solution of Cantor's problem; it was not a *proof* of the continuum hypothesis but only a proof that it cannot be disproved.

To understand how Gödel achieved his results we need to understand what is meant by a model for an axiom system. Let us return for a moment to the axioms of plane geometry. If we take these axioms, including the parallel postulate, we have the axioms of Euclidean geometry; if instead we keep all the other axioms as before but replace the parallel postulate by its negation, we have the axioms of a non-Euclidean geometry. For both axiom systems—Euclidean and non-Euclidean—we ask: Can these axioms lead to a contradiction?

To ask the question of the Euclidean system may seem unreasonable. How could there be anything wrong with our familiar, 2,000-year-old high school geometry? On the other hand, to the nonmathematician there certainly is something suspicious about the second axiom system, with its denial of the intuitively plausible parallel postulate. Nonetheless, from the viewpoint of 20th-century mathematics the two kinds of geometry stand more or less on an equal footing. Both are sometimes applicable to the physical world and both are consistent, in a relative sense we shall now explain.

First we show that non-Euclidean geometry is consistent. In order to do this we merely replace the word "line" everywhere by the phrase "great circle," a line formed on the surface of a sphere by a plane passing through the center of the sphere. We now regard the axioms as statements about points and great circles on a given sphere. Moreover, we agree to identify each pair of diametrically opposite points on the sphere as a single point. If the reader prefers, he can imagine the axioms of non-Euclidean geometry rewritten, with the word "line" everywhere replaced by "great circle," the word "point" everywhere replaced by "point pair." Then it is evident that all the axioms are true, at least insofar as our ordinary notions about the surface of a sphere are true. In fact, from the axioms of Euclidean solid geometry one can easily prove as theorems that the surface of a sphere is a non-Euclidean



SET OF ALL SUBSETS OF A GIVEN SET is illustrated. The square, triangle and circle at top form the three-element set A. This set has 2^3, or 8, subsets (provided that the whole set and the empty set are somewhat improperly included). This new set consisting of eight elements is called the power set of A, and it is denoted 2^A. If A has n elements, the power set of A has 2^n elements. If A is infinite, 2^A is also infinite, and it is not equivalent to A.

surface in the sense we have just described. In other words, we now see that if the axioms of non-Euclidean geometry led to a contradiction, then so would the ordinary Euclidean geometry of spheres lead to a contradiction. Thus we have a *relative* proof of consistency; if Euclidean three-dimensional geometry is consistent, then so is non-Euclidean two-dimensional geometry. We say that the surface of the Euclidean sphere is a model for the axioms of non-Euclidean geometry. (In the particular model we have used the parallel postulate fails because there are no parallel lines. It is also possible to construct a surface, the "pseudosphere," for which the parallel postulate is false because there is more than one line through a point parallel to a given line.)

The invention of non-Euclidean geometry, and the recognition that its consistency is implied by the consistency of Euclidean geometry, was the work of many great 19th-century mathematicians; we mention the name of Bernhard Riemann in particular. Only in the 20th century was the question raised of whether or not Euclidean geometry itself is consistent.

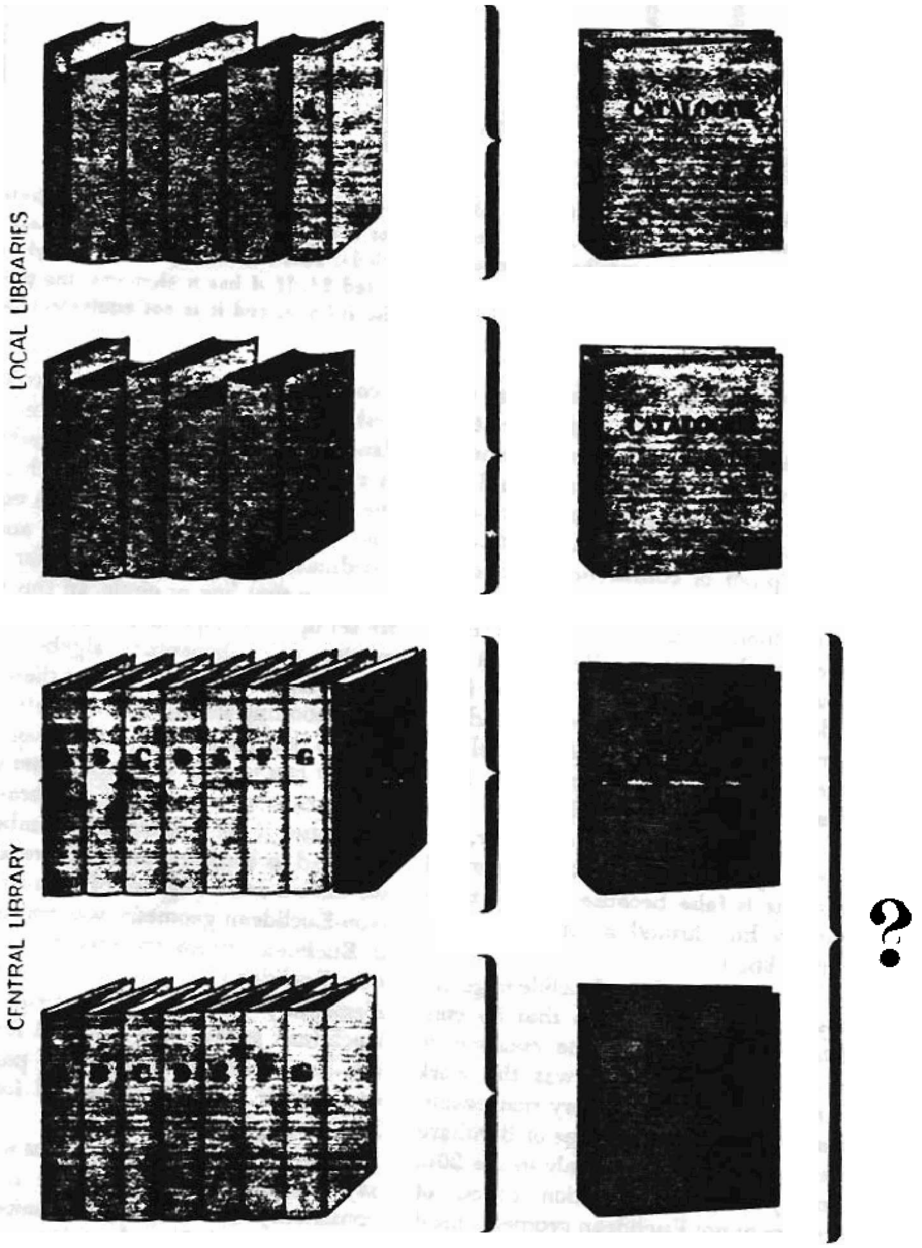
This question was asked and answered by David Hilbert. Hilbert's solution was a simple application of the idea of a coordinate system. As many college freshmen learn, to each point in the plane we can associate a pair of numbers: its x and y coordinates. Then with each line or circle we can associate an equation: a relation between the x and y coordinates that is true only for the points on that line or circle. In this way we set up a correspondence between geometry and elementary algebra. For every statement in one subject there is a corresponding statement in the other. It follows that the axioms of Euclidean geometry can lead to a contradiction only if the rules of elementary algebra—the properties of the ordinary real numbers—can lead to a contradiction. Here again we have a relative proof of consistency. Non-Euclidean geometry was consistent if Euclidean geometry was consistent; now Euclidean geometry is consistent if elementary algebra is consistent. The Euclidean sphere was a model for the non-Euclidean plane; the set of pairs of coordinates is in turn a model for the Euclidean plane.

With these examples before us we can say that Gödel's proof of the relative consistency of the axiom of choice and of the continuum hypothesis is analogous to Hilbert's proof of the relative consistency of Euclidean geometry. In both instances the standard theory was justified in terms of a more elementary one. Of course, no one ever seriously doubted the reliability of Euclidean geometry, whereas such outstanding mathematicians as L. E. J. Brouwer, Hermann Weyl and Henri Poincaré had grave doubts about the axiom of choice. In this sense Gödel's result had a much greater impact and significance.

The analogous development with respect to non-Euclidean geometry—what we might call non-Cantorian set theory—has taken place only since 1963, in the work of one of the authors of this article (Cohen). What is meant by "non-Cantorian set theory"? Just as Euclidean and non-Euclidean geometry use the same axioms, with the one exception of the parallel postulate, so standard ("Cantorian") and nonstandard ("non-Cantorian") set theory differ only in one axiom. Non-Cantorian set theory takes the axioms of restricted set theory and adds not the axiom of choice but rather one or another form of the negation of the axiom of choice. In particular we can take as an axiom the negation of the continuum hypothesis. Thus, as we shall explain, there now exists a complete solution of the continuum problem. To Gödel's discovery that the continuum hypothesis is not disprovable is added the fact that it is also not provable.

Both Gödel's result and the new discoveries require the construction of a model, just as the consistency proofs for geometry that we have described required a model. In both cases we want to prove that if restricted set theory is consistent, then so is standard set theory (or nonstandard theory).

Gödel's idea was to construct a model for restricted set theory, and to prove that in this model the axiom of choice and the continuum hypothesis were theorems. He proceeded in the following way. Using only the axioms of restricted set theory [*see illustration on page 114*], we are guaranteed first the existence of at least one set (the empty set) by Axiom 2; then by Axiom 3 and Axiom 4 we are guaranteed the existence of an infinite sequence of ever larger finite sets; then by Axiom 5, the existence of an infinite set; then by Axiom 7, of an endless sequence of ever larger (nonequivalent) infinite sets, and so on. In essentially this way Gödel specified a class of sets by the manner in which they could actually be constructed in successive steps from simpler sets. These sets he called the "constructible sets"; their existence was guaranteed by the axioms of restricted set theory. Then he showed that within the realm of the constructible sets the axiom of choice and the continuum hypothesis can both be proved. That is to say, first, from any constructible collection a of constructible sets (A, B, ...) one can choose a constructible set Z consisting of at least one element each from A, B and so on. This is the axiom of choice, which here might more properly be called the theorem of choice. Second, if A is any infinite constructible set, then there is no constructible set "between" A and 2^A (bigger than A, smaller than the power set of A and equivalent to neither). If A is taken as the first infinite cardinal, this last statement is the continuum hypothesis.



RUSSELL'S PARADOX is illustrated by supposing that in a certain country it is the custom of librarians to list their books not in a card catalogue but in a looseleaf catalogue; that is, the catalogue itself is a book. Some librarians list the catalogue itself in the catalogue (*top*); some do not (*second from top*). The first kind of catalogue is called an *R*-set, after Bertrand Russell; *R*-sets are sets that include themselves. What happens, however, if the head librarian of the country decides to make a master catalogue of all the catalogues that do not list themselves? Does his own catalogue belong in the master catalogue or not?

Hence a "generalized continuum hypothesis" was proved in the case of *constructible* set theory. Gödel's work would therefore dispose of these two questions completely if we were prepared to adopt the axiom that only constructible sets exist. Why not do so? Because one feels it is unreasonable to insist that a set must be constructed according to *any* prescribed formula in order to be recognized as a genuine set. Thus in ordinary (not necessarily constructible) set theory neither the axiom of choice nor the continuum hypothesis had been proved. At least this much was

"COMMON NOTIONS"

1. Things that are equal to the same thing are also equal to one another.

2. If equals are added to equals, the wholes are equal.

3. If equals are subtracted from equals, the remainders are equal.

4. Things that coincide with one another are equal to one another.

5. The whole is greater than the part.

"POSTULATES"

1. (It is possible) to draw (exactly one) straight line from any point to any point.

2. (It is possible) to extend a finite straight line continuously in a straight line.

3. (It is possible) to describe a circle with any center and distance.

4. All right angles are equal to one another.

5. If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

EUCLID'S AXIOMS were of two kinds: "common notions" and "postulates." The Scottish physicist and mathematician John Playfair (1748–1819) is identified with an axiom that may be shown to be equivalent to Euclid's Postulate 5: Through a given point *A* not on a given line *m* there passes one line that does not intersect *m*. A non-Euclidean geometry is obtained by replacing "one" with either "none" or "more than one." It should be said that Euclid's axioms are not clear or complete by modern standards.

certain: either of them could be assumed without causing any contradiction unless the "safe" axioms of restricted set theory already are self-contradictory. Any contradiction they cause must already be present in constructible set theory, which is a model for ordinary set theory. In other words, it was known that neither could be disproved from the other axioms but not whether they could be proved.

Here the analogy with the parallel postulate in Euclidean geometry becomes particularly apt. That Euclid's axioms are consistent was taken for granted until quite recently. The ques-

$\forall$	FOR ALL	$\leftrightarrow$	IF AND ONLY IF	$\in$	IS A MEMBER (ELEMENT) OF
$\exists$	THERE EXISTS	$\vee$	OR	$=$	EQUALS
$\exists!$	THERE EXISTS UNIQUELY	&	AND	$\neq$	DOES NOT EQUAL
$\cup$	UNION	$\sim$	NOT	ϕ	THE EMPTY SET
$\rightarrow$	IMPLIES	$\subseteq$	IS A SUBSET OF		

1. AXIOM OF EXTENSIONALITY

$\forall x, y\, (\forall z\, (z \in x \leftrightarrow z \in y) \rightarrow x = y).$

Two sets are equal if and only if they have the same members

2. AXIOM OF THE NULL SET

$\exists x\, \forall y\, (\sim y \in x).$

There exists a set with no members (the empty set)

3. AXIOM OF UNORDERED PAIRS

$\forall x, y\, \exists z\, \forall w\, (w \in z \leftrightarrow w = x \vee w = y).$

If x and y are sets, then the (unordered) pair $\{x,y\}$ is a set

4. AXIOM OF THE SUM SET OR UNION

$\forall x\, \exists y\, \forall z\, (z \in y \leftrightarrow \exists t\, (z \in t \,\&\, t \in x)).$

If x is a set of sets, the union of all its members is a set. (For example, if $x = \{\{a,b,c\},\{a,c,d,e\}\}$, then the union of the two elements of x is the set $\{a,b,c,d,e\}$

5. AXIOM OF INFINITY

$\exists x\, (\phi \in x \,\&\, \forall y\, (y \in x \rightarrow y \cup \{y\} \in x)).$

There exists a set x that contains the empty set and that is such that if y belongs to x, then the union of y and $\{y\}$ is also in x. The distinction between the element y and the singleton set $\{y\}$ is basic. This axiom guarantees the existence of infinite sets

6_n. AXIOM OF REPLACEMENT

$\forall t_1 \ldots t_k\, (\forall x\, \exists! y\, A_n(x,y;t_1 \ldots t_k) \rightarrow \forall u\, \exists v\, B(u,v))$ where $B(u,v) \equiv \forall r\, (r \in v \leftrightarrow \exists s\, (s \in u \,\&\, A_n(s,r;t_1 \ldots t_k))).$

This axiom is difficult to restate in English. It is called 6_n rather than 6 because it is really a whole family of axioms. We suppose that all the formulas expressible in our system have been enumerated; the nth is called A_n. Then the axiom of replacement says that if for fixed $t_1 \ldots t_k$, $A_n(x,y;t)$ defines y uniquely as a function of x, say $y = \varphi(x)$, then for each u the range of φ on u is a set. This means, roughly, that any ("reasonable") property that can be stated in the formal language of the theory can be used to define a set (the set of things having the stated property)

7. AXIOM OF THE POWER SET

$\forall x\, \exists y\, \forall z\, (z \in y \leftrightarrow z \subseteq x).$

This axiom says that there exists for each x the set y of all subsets of x. Although y is thus defined by a property, it is not covered by the replacement axiom because it is not given as the range of any function. Indeed, the cardinality of y will be greater than that of x, so that this axiom allows us to construct higher cardinals

8. AXIOM OF CHOICE

If $a \rightarrow A_a \neq \phi$ is a function defined for all $a \in x$, then there exists another function $f(a)$ for $a \in x$, and $f(a) \in A_a$.

This is the well-known axiom of choice, which allows us to do an infinite amount of "choosing" even though we have no property that would define the choice function and thus enable us to use 6_n instead.

9. AXIOM OF REGULARITY

$\forall x\, \exists y\, (x = \phi \vee (y \in x \,\&\, \forall z\, (z \in x \rightarrow \sim z \in y))).$

This axiom explicitly prohibits $x \in x$, for example.

ZERMELO-FRAENKEL AXIOMS FOR SET THEORY are listed. In order to state these theorems it is necessary to use the symbols of set theory, a glossary of which is given at top. This axiom system was put forward by Ernst Zermelo and Abraham Fraenkel.

tion that interested geometers was whether or not they are independent, that is, whether the parallel postulate could be proved on the basis of the others. A whole series of geometers tried to prove the parallel postulate by showing that its negation led to absurdities. It seems that Carl Friedrich Gauss was the first to see that these "absurdities" were simply the theorems of a new, non-Euclidean geometry. But what Gauss had the courage to think he did not have the courage to publish. It was left for János Bolyai, Nikolai Ivanovich Lobachevsky and Riemann to carry out the logical consequences of denying the parallel postulate. These consequences were the discovery of "fantastic" geometries that had as much logical consistency as the Euclidean geometry of "the real world." Only after this had happened was it recognized that two-dimensional non-Euclidean geometry was just the ordinary Euclidean geometry of certain curved surfaces (spheres and pseudospheres).

The analogous step in set theory would be to deny the axiom of choice or the continuum hypothesis. By this we mean, of course, that the step would be to prove that such a negation is consistent with restricted set theory, in the same sense in which Gödel had proved that the affirmation was consistent. It is this proof that has been accomplished in the past few years, giving rise to a surge of activity in mathematical logic whose final outcome cannot be guessed.

Since it is a question of proving the relative consistency of an axiom system, we naturally think of constructing a model. As we have seen, the relative consistency of non-Euclidean geometry was established when surfaces in Euclidean three-space were shown to be models of two-dimensional non-Euclidean geometry. In a comparable way, in order to prove the legitimacy of a non-Cantorian set theory in which the axiom of choice or the continuum hypothesis is false we must use the axioms of restricted set theory to construct a model in which the negation of the axiom of choice or the negation of the continuum hypothesis can be proved as theorems.

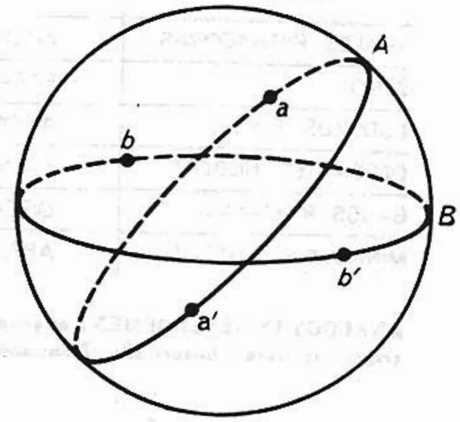
It must be confessed that construction of this model is a complex and delicate affair. This is perhaps to be expected. In Gödel's constructible sets, his model of Cantorian set theory, the task was to create something essentially the same as our intuitive notion of sets but more tractable. In our present task we have to create a model of something unintuitive and strange, using the familiar building stones of restricted set theory.

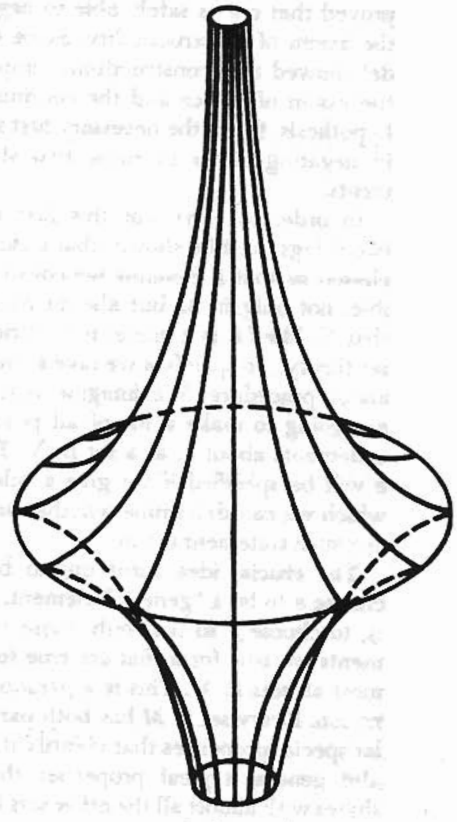
Rather than throw up our hands and say it is impossible to describe this model in a nontechnical article, we shall attempt at least to give a descriptive account of one or two of the leading ideas that are involved. Our starting point is ordinary set theory (without the axiom of choice). We hope only to prove the consistency of non-Cantorian set theory in a relative sense. Just as the models of non-Euclidean geometry prove that non-Euclidean geometry is consistent if Euclidean geometry is consistent, so we shall prove that if restricted set theory is consistent, it remains so if we add the statement "The axiom of choice is false" or the statement "The continuum hypothesis is false." We may now assume that we have available as a starting point a model for restricted set theory. Call this model *M*; it can be regarded as Gödel's class of constructible sets.

We know from Gödel's work that in order for the axiom of choice or the continuum hypothesis to fail we must add to *M* at least one nonconstructible set. How to do this? We introduce the letter *a* to stand for an object to be added to *M*; it remains to determine what kind of thing *a* should be. Once we add *a* we must also add everything that can be formed from *a* by the permitted operations of restricted set theory: uniting two or more sets to form a new set, forming the power set and so on. The new collection of sets generated in this way by *M* + *a* will be called *N*. The problem is how to choose *a* in such a way that (1) *N* is a model for restricted set theory, as *M* was by assumption, and (2) *a* is not constructible in *N*. Only if this is possible is there any hope of denying the axiom of choice or the continuum hypothesis.

We can get a vague feeling of what has to be done by asking how a geometer of 1850 who was trying to discover the pseudosphere might have proceeded. In a very rough sense, it is as if he had started with a curve *M* in the Euclidean plane, thought of a point *a* not in that plane, and then connected that point *a* to all the points in *M*. Since *a* is chosen not to lie in the plane of *M*, the resulting surface *N* will surely not be the same as the Euclidean plane. Thus it is reasonable to think that with enough ingenuity and technical skill one could show that it is really a model for a non-Euclidean geometry.

The analogous thing in non-Cantorian set theory is to choose the new set *a* as a nonconstructible set, then to generate a new model *N* consisting of all sets obtained by the operations of restricted set theory applied to *a* and to the sets in *M*. If this can be done, it will have been





ON SURFACE OF A SPHERE "straight line" is interpreted to mean "great circle" (*A and B at top*). Through any pair of diametrically opposite points (*aa′ and bb′*) there pass many great circles. If we interpret "point" to mean "point pair," then Euclid's first postulate is true. The second postulate is true if one allows the extended "straight line" to have a finite total length, or to retrace itself many times as it goes around the sphere. The third postulate is also true if one understands distance to be measured along great circles that can be retraced several times; here a "circle" means merely the set of points on the sphere at a given great-circle distance from a given point. The fourth postulate is likewise true. Playfair's postulate is false, because any two great circles intersect. Thus the sphere is a model of non-Euclidean geometry. So is the pseudosphere (*bottom*), if straight lines are interpreted as being the shortest curves connecting any two points on the surface. On the surface of the pseudosphere there are many "straight lines" that pass through a given point and do not cross a given straight line.

GEOMETRY	STAGE OF DEVELOPMENT	SET THEORY
THALES, PYTHAGORAS	INTUITIVE BASIS FOR FIRST THEOREMS	CANTOR
ZENO	PARADOX REVEALED	RUSSELL
EUDOXUS, EUCLID	AXIOMATIC BASIS FOR STANDARD THEORY	ZERMELO, FRAENKEL, ETC.
DESCARTES, HILBERT	STANDARD THEORY SHOWN (RELATIVELY) CONSISTENT	GODEL
GAUSS, RIEMANN	DISCOVERY OF NONSTANDARD THEORIES	CURRENT WORK
MINKOWSKI, EINSTEIN	APPLICATION OF NONSTANDARD THEORY	? ?

ANALOGY IN DEVELOPMENT of geometry (*left*) and set theory (*right*) is traced historically. Nonstandard (non-Euclidean) geometry has been applied in such theories as Einstein's theory of relativity. Nonstandard set theory has yet to be applied in physics.

proved that one is safely able to negate the axiom of constructibility. Since Gödel showed that constructibility implies the axiom of choice and the continuum hypothesis, this is the necessary first step in negating either of these two statements.

In order to carry out this first step two things must be shown: that *a* can be chosen so that it remains nonconstructible, not only in *M* but also in *N*, and that *N*, like *M*, is a model for restricted set theory. To specify *a* we take a roundabout procedure. We imagine that we are going to make a list of all possible statements about *a*, as a set in *N*. Then *a* will be specified if we give a rule by which we can determine whether or not any such statement is true.

The crucial idea turns out to be to choose *a* to be a "generic" element, that is, to choose *a* so that only those statements are true for *a* that are true for almost all sets in *M*. This is a paradoxical notion. Every set in *M* has both particular special properties that identify it, and also general typical properties that it shares with almost all the other sets in *M*. It turns out to be possible in a precise way to make this distinction between special and generic properties perfectly explicit and formal. Then when we choose *a* to be a generic set (one with, so to speak, no special properties that distinguish it from any set in *M*), it follows that *N* is still a model for restricted set theory. The new element *a* we have introduced has no troublesome properties that can spoil the *M* we started with. At the same time *a* is nonconstructible. Any constructible set has a special character—the steps by which it can be constructed—and our *a* precisely lacks any such individuality.

To construct a model in which the continuum hypothesis is false we must add to *M* not just one new element *a* but a great many new elements. In fact, we must add an infinite number of them. We can actually do this in such a way that the elements we add have cardinality

$$\aleph_2 = 2^{(2^{\aleph_0})}$$

from the viewpoint of the model *M*. Again a rough geometric analogy may be helpful: To a two-dimensional creature living embedded in a non-Euclidean surface it would be impossible to recognize that his world is part of a three-dimensional Euclidean space. In the present instance we, standing outside *M*, can see that we have thrown in only a countable infinity of new elements. They are such, however, that the counting cannot be done by any apparatus available in *M* itself. Thus we obtain a new model *N'*, in which the continuum hypothesis is false. The new elements, which in *N'* play the role of real numbers (that is, points on a line segment), have cardinality greater than $2^{\aleph_0}$, and so there is now an infinite cardinal—namely $2^{\aleph_0}$—that is greater than $\aleph_0$ and yet smaller than *c*, since in our model *N'*, *c* is equal to

$$2^{(2^{\aleph_0})}.$$

Since we can construct a model of set theory in which the continuum hypothesis is false, it follows that we can add to our ordinary restricted set theory the assumption of the falsity of the continuum hypothesis; no contradiction can result that was not already present. In the same spirit we can construct models for set theory in which the axiom of choice fails. We can even be quite specific about which infinite sets it is possible to "choose from" and which are "too big to choose from."

Whereas Gödel produced his results with a single model (the constructible sets), we have in non-Cantorian set theory not one but many models, each constructed with a particular purpose in mind. Perhaps more important than any of the models is the technique that enables one to construct them all: the notion of "generic" and the related notion of "forcing." Very roughly speaking, generic sets have only those properties they are "forced" to have in order to be set-like. In order to decide whether *a* is "forced" to have a certain property we must look at all of *N*. Yet *N* is not really defined until we have specified *a*! The recognition of how to make this seemingly circular argument noncircular is another key element in the new theory.

What does the history of geometry suggest for the future of set theory? The most remarkable thing about non-Euclidean geometry is that it turned out to be an essential prerequisite for Einstein's general theory of relativity. Riemann created Riemannian geometry for the purely abstract purpose of unifying, clarifying and deepening the non-Euclidean geometry of Lobachevsky, Bolyai and Gauss. This geometry turned out to be the indispensable tool for Einstein's revolutionary reinterpretation of the gravitational force.

Does this example justify an expectation that non-Cantorian set theory someday will find a currently unforeseeable application in the "real" (that is, nonmathematical) world? No one today would venture an answer. Certainly we can see (with hindsight) that geometry has always furnished the essential background in which physical events take place. In that sense it should perhaps have been expected that fundamental advances in geometry would find a physical application. Set theory does not seem today to have any such organic interrelationship with physics. Still, there have been some mathematicians (Stanislaw Ulam, for example) who have proposed that abstract set theory might furnish useful models for theoretical physics. At this stage the safest thing is to refuse to predict anything about the future—except that it is unpredictable.