Mathematical Proof: What It Is and What It Ought to Be

Peter Renz

Prove all things, hold fast that which is good
. . . I Thessalonians 5:21

Certainty?

Absolute certainty can be an unproductive and unwise goal, even in mathematics. The rigor aspired to should be only that needed for the purposes at hand. Be ready to doubt what has been proved, for some proofs depend on hidden assumptions, while others can be interpreted in various ways. Ask that a convincing argument be given for an assertion before accepting it, and continue to test assertions by judging the reasonableness and usefulness of the conclusions drawn. Ask also whether the hypotheses are adequate and acceptable. Here I will use "to prove" in the broader sense of "to test," which includes testing by pure reason, by intuition, and in practice. In this sense, the testing, or proving, of assertions is a process in which all things are proved and the good retained.

Formalism and rigor have their uses. Hilbert's investigations into the foundations of geometry uncovered the hidden assumptions of classical geometry. Finding these "defects" in Euclidean geometry did not invalidate it, for it had already proved its worth. Hilbert's work made explicit the assumptions that should be added to the list of postulates of a formal theory agreeing with Euclid's results.

Before Hilbert's work, Euclid's postulates were thought to be completely adequate. This was wrong, but no harm was done. The error was discovered, and matters put right. In the same way, the errors of today (if any) may be discovered and corrected. Our successors will make repairs and hold fast to the good, just as we have.

Hilbert used rigorous formal methods because they suited his purposes. He wished to give mathematics an unshakable, formal foundation. Hence, he was concerned with axioms. His methods and axioms may not help a geometer who wants to know whether there is a single tile, copies of which can cover a plane only nonperiodically—an open question in plane geometry in 1980 (see Box 1). Intuitive methods are more important for solving such problems.

Barring the inaccurate reading of symbols and the failure to follow rules, it is possible to establish with certainty that an expression is a theorem in a certain formal system. However, despite this certainty, questions remain. Do the axioms of that system correctly reflect the properties of the objects of that theory in its intended interpretation? Do the intended mathematical objects capture important

Box 1: Nonperiodic Tilings

In 1961 Hao Wang became interested in a class of problems requiring the periodic tiling of the plane by sets of unit squares whose edges were colored in special ways. The colors on the edges of neighboring tiles are required to match. Such sets of tiles are called Wang dominoes. These problems and their relation to logic and computer science are explained in Wang's November 1965 *Scientific American* article, "Games, Logic, and Computers." Wang believed that, if such a set of tiles could be used to tile the plane while obeying the matching rules, they could tile the plane periodically also following the matching rules.

Robert Berger provided that there are sets of Wang tiles that can only tile the plane nonperiodically, that is, so that no translate of the pattern of tiles will match up with itself. Berger's first set of nonperiodic Wang dominoes had more than 20,000 different types of dominoes, but he later reduced the number of different types required to just over one hundred. Raphael M. Robinson found a set of six tiles that can only tile the plane nonperiodically. Roger Penrose invented a set of two tiles that can only tile the plane nonperiodically. His tiles are astonishing, fascinating, and beautiful. The most accessible account of them is in Martin Gardner's Mathematical Games column, *Scientific American*, January 1977. The reader may want to watch for forthcoming books by Berlekamp, Conway, and Guy (*Winning Ways*, Academic Press, in press) and Grünbaum and Shephard (*Tilings and Patterns*, W. H. Freeman and Company, in press), which will give a more detailed and up-to-date treatment of nonperiodic tilings.

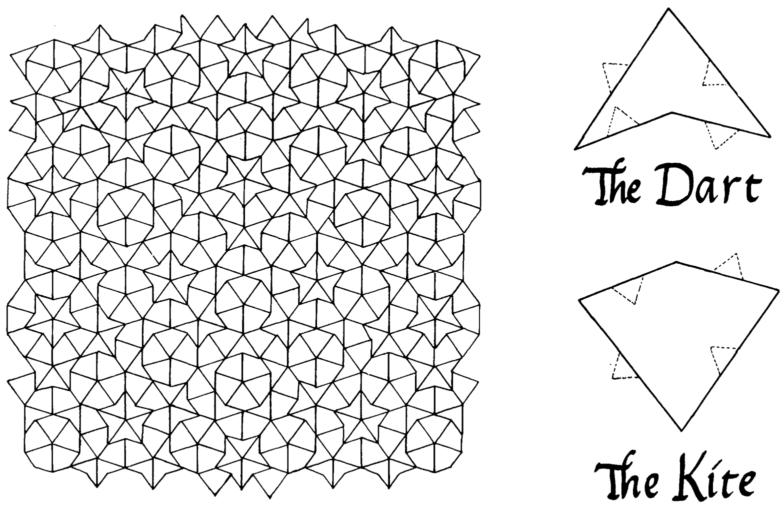
Work in logic led to the discovery of nonperiodic tilings, but bold geometric insight led to the Penrose tiles. What combination of logic, analysis, inspiration, geometry, and even art, will it take to resolve the question of whether there can be a single tile, copies of which can only tile the plane nonperiodically?

and useful aspects of reality? Are the rules of deduction correct, and do they apply to the domain in question?

To be specific, no proof can be given that Hilbert's axioms capture all aspects of intuitive geometry or that intuitive geometry, Riemannian geometry, or any other geometry is either a completely accurate or a completely satisfactory description of physical space. Such assertions cannot be proved mathematically. They lie outside of mathematics itself. However, the adequacy of Hilbert's axioms and of various models of space can be tested by seeing how well they serve us in viewing or working with reality.

Proving or testing theories is a continual process. We hold fast to those that are useful and discard or replace those that are not.

It seems most likely that a good eye and strong geometric intuition will be most important in answering this question. Roger Penrose has given an interesting account of these tiles and how he invented them in "Pentaplexity," first published in *Eureka*, Number 39, and reprinted in *The Mathematical Intelligencer*, Volume 2, Number 1 (1979).



The Penrose tiles: the dart and kite can tile the plane only nonperiodically. Tabs and cutouts as shown in dotted lines on the enlarged tiles prevent their being assembled to form a parallelogram that could tile the plane periodically. The tiles have a top and bottom surface that prevents their being turned over, so the cutouts and ears cannot be matched to form a parallelogram.

The Uses of Proof

If proofs do not establish the absolute validity of theorems, what good are they? Usefulness or adequacy must be judged relative to the purposes at hand. If we want to know whether a given expression is a theorem of Zermelo-Fraenkel set theory, then a formal derivation is probably in order. If we wish to convince students that the Pythagorean theorem is true, then one or two figures and a bit of informal argument suffice.

Mark Kac remarked that a formal proof of the Pythagorean theorem from Hilbert's axioms was written by a student of Hugo Steinhaus's, a work of nearly 80 pages. This is not the sort of proof you would want to present to a class. Indeed, as

Teacher: "... and now I want to prove this theorem."
Pupil: "Why bother to prove it, teacher? I take your word for it."

Professor: „... Und nun will ich Ihnen diesen Lehrsatz jetzt auch beweisen."
Junge: „Wozu beweisen, Herr Professor? Ich glaub' es Ihnen so."

Figure 1.

From *The Mischief Book* by Wilhelm Busch, as translated by Abby Langdon Alger and published by R. Worthington in 1880, reproduced from *Max and Moritz, from the Pen of Wilhelm Busch*, edited and annotated by H. Arthur Klein, Dover Publications, 1962.

Wilhelm Busch's cartoon (Figure 1) suggests, it may be enough to state the theorem.

If the adequacy or worth of a proof is to be judged by the purposes at hand, what purposes should be considered? Claiming neither completeness nor ranking by importance, I propose the following.

To clarify relations between properties.

(Example. The intermediate value theorem for a continuous function f on a closed interval $[a,b]$ says that if $f(a) < f(b)$ and c is any number strictly between $f(a)$ and $f(b)$, then $f(x) = c$ for some x between a and b. This is called the intermediate value property for continuous functions on intervals. The key observation in the formal proof is that, if there is no such x, then the open set S defined by $\{x \in [a,b] \mid f(x) < c\}$ is identical to the closed set defined by $\{x \in [a,b] \mid f(x) \leqslant c\}$. Hence, if there is no such c, then S is both an open and closed subset of $[a,b]$. The proof continues by using the fact that the only subsets that are open and closed in $[a,b]$ are $\emptyset$ and $[a,b]$. Because $f(a) < c, a \in S$, so $S \neq \emptyset$. Similarly, because $f(b) > c, b \notin S$ and $S \neq [a,b]$. Thus we are presented with a contradiction if $f(x) \neq c$ for all $x \in [a,b]$. It follows that $f(x) = c$ for some $x \in [a,b]$.

This proof also shows that any continuous map from a space X to the real numbers must have the intermediate value property if the only open and closed subsets of X are $\emptyset$ and X. This gives rise to the notion of *connectedness*. A *topological space* X (for example a subset X of the plane or of three-space) *is connected* if the only subsets of X which are both open and closed in X are $\emptyset$ and X. Simple topological arguments show that the continuous image of a connected set is connected. The intermediate value theorem for continuous functions on intervals is a special case of this general theorem and the fact that intervals in the real line are connected.

I remark that, although the proof outlined above can be formalized and generalized, it is not likely to be as convincing to a student as a picture combined with the simple argument that the graph of a continuous function f on $[a, b]$ where $f(a) < c < f(b)$ must cross the line $y = c$. With this result, as with the Pythagorean theorem, our search for rigor and precision leads us away from directly perceived intuition. We gain more precision and generality but lose the directly convincing force of pictures and diagrams.

Here a proof spawns a definition, and the definition allows a conceptual simplification and broad generalization of the original theorem. The property of connectedness emerges from the proof of the intermediate value theorem.)

To give us the pleasure of working out the argument and figuring out the proof.

(This is a pleasure we all know. The reader should supply his or her favorite example.)

To help us remember important or useful results.

(Example. All of us can think of examples of derivations that help recall a result. I find this true of various identities that I use infrequently such as the formula for $\tan(x + y)$

$$\left(= \frac{\sin(x+y)}{\cos(x+y)} = \frac{\sin x \cos y + \cos x \sin y}{\cos x \cos y - \sin x \sin y} = \frac{\tan x + \tan y}{1 - \tan x \tan y} \right),$$

and if you should forget for a moment the formulas for $\sin(x + y)$ and $\cos(x + y)$ they can be derived immediately from de Moivre's formula ($e^{ix} = \cos x + i \sin x$) in conjunction with the unforgettable identity $e^{ix}e^{iy} = e^{i(x+y)}$). But here, too, I expect the reader will have his or her own favorite example.)

To guide us along formally correct paths where our intuition may be weak or misleading.

(Example. Beginning with the formal identity

$$(1 - x)(1 + x + x^2 + \cdots + x^n) = 1 - x^{n+1}$$

we can conclude that $(1 - x)^{-1} = \sum x^k (k \geqslant 0)$ if $|x| < 1$. This holds with the appropriate modification even if x is allowed to be an element of a Banach algebra. The initial formula was applied to real numbers where inverses are familiar. However, the formula and its derivation hold even in unfamiliar contexts.)

To guide computations.

(Example. The proof that the greatest common divisor (g.c.d.) of n and m is the least positive integer of the form $ln + km$, in which l and k are integers, is based upon the Euclidean algorithm for computing the g.c.d. of any two integers. Both the proof and the computation can be generalized from the ring of integers to the ring of polynomials.)

To explore the properties of formal systems.

(Example. Here the logical structure of the theory is the object of study, and the notion of proof is a formal one within the theory. This kind of analysis gives us such theorems as: "The theorems of propositional calculus are precisely the tautologies, and there is an effective procedure [constructing a truth table] for determining whether or not a well-formed formula is a tautology and hence a theorem.")

To offer a different perspective.

(Example. The integers may first be regarded very concretely in terms of counting and arithmetic. Next we may look at them in terms of set theory and cardinal and ordinal arithmetic. A change to the axiomatic point of view (Peano) may shed new light. Model theory (the work of Skolem and others on the nonstandard integers) may show limitations of the axiomatic approach. A hypermodern category theory approach (Peano-Lawvere) may shed still further light and offer certain simplifications. If we look at all of these, no single point of view dominates all others, but the proofs of the properties of the integers developed from these separate points of view all contribute to our appreciation of the integers.)

Answering Unasked Questions: Proofs as Detours

In the classroom an overelegant or overrefined approach may be counterproductive. In introducing students to surds such as $\sqrt{2}$, proving that $\sqrt{2}$ exists as a real number would be a long and arduous detour—almost certainly counterproductive. Why not encourage students to work with the notion of a square root and to discover what they can about $\sqrt{2}$ from its defining properties: $(\sqrt{2})^2 = 2$ and $\sqrt{2} \geqslant 0$? They might discover that $1.41 < \sqrt{2} < 1.42$, and they may be led to discover that $\sqrt{2}$ is not the ratio of two integers and hence not a repeating decimal.

A square root is far more elementary than the complete real number system. To prove the existence of $\sqrt{2}$ as an element of a postulated complete ordered field is to explain the familiar in terms of the unknown. After all, that complete ordered field, being uncountable, contains in addition to $\sqrt{2}$ numbers that can't even be named by any finite concatenation of symbols (see Box 2). In showing that $\sqrt{2}$ exists in this setting, the instructor first asks the student to cast aside any naive belief in the existence of $\sqrt{2}$ and then to embrace an existence proof based on the existence of a more esoteric object, namely, the real number system.

Similarly, it may be best simply to tell the beginning calculus student that the formula $d(x^n)/dx = nx^{n-1}$ is valid for all constant exponents n, not only for

Box 2: How Many Numbers Can We Name?

If for simplicity we assume we have augmented English with mathematical symbols and punctuation so that we have in all a hundred symbols to work with, we can easily figure out the maximum number of numbers that can be named by strings of at most six symbols (or k symbols), namely 100^6 (or 100^k), where we assume some symbol represents a blank space.

Further, if we assign each symbol (including the symbol for a blank space) a unique value from zero to ninety-nine, then each finite string of such symbols corresponds to a unique number in base one hundred. In fact, the system of all such finite strings in "alphabetical" order is clearly isomorphic to the system of natural numbers written in base one hundred.

Not all such strings will actually name real numbers, so the namable real numbers are in one-to-one correspondence with a subset of the pseudo-names that make up this countable list. Standard arguments state that all subsets of countable sets are at most countable. Hence we conclude that the number of real numbers that can be named or described is at most countable. Because the usual theory shows there are uncountably many real numbers, this means that almost all real numbers are unnamable or undescribable.

Of course, the problem of determining whether or not a finite string of symbols describes a real number (or even an integer) is in general hopeless. For example, consider: "The least integer n such that for all integers k greater or equal to n, whenever k is prime, $k + 2$ is not prime." If there are finitely many twin primes, this phrase describes an integer. If there are infinitely many twin primes, this phrase does not describe an integer. Worse yet, as we shall see, even the words "finite" and "infinite" used here are subject to certain uncertainties.

integral values of n. There is every reason to prove $d(x^n)/dx = nx^{n-1}$ for positive integers n. The usual proofs are convincing, and they reveal why the theorem is true. For example, the proof that uses the binomial theorem shows that the derivative is simply the coefficient of the linear term, Δx, when $(x + \Delta x)^n$ is expanded in terms of Δx. What could be closer to the meaning of the derivative for a polynomial?

If you feel $d(x^n)/dx = nx^{n-1}$ should be proved for nonintegral n in an introductory course, consider that honesty might first require that a^b be defined rigorously for reals a and b with $a > 0$. Many approaches are possible, but those that are mathematically easiest (for example, the approach using natural logarithm, $\ln(x)$, defined as $\ln(x) = \int_1^x t^{-1}\,dt$) are so far divorced from the ordinary notion of powers that they seem to be black magic. Such proofs may undermine rather than reinforce the student's confidence and trust in mathematics. Even worse, they give a false picture of the way mathematics is done.

Practice may advantageously precede proof in teaching implicit differentiation. The method is powerful and useful. Students will quickly learn to apply it and appreciate that it works. However, it is not easy to state and prove any satisfactory implicit differentiation theorem in beginning calculus. As evidence of this, I note that none of the following careful authors gives the theorem or its proof when introducing the method: Richard Courant [1937], Tom Apostol [1967], Michael Spivak [1967].

The mathematician's desire for a deeper understanding of certain intuitively evident facts often leads to proofs that are less evident and less convincing than a direct appeal to intuition. Few students would doubt that a continuous function f on $[0, 1]$ such that $f(0) = -1$ and $f(1) = +1$ must take on the value $f(x) = 0$ for some x in $[0, 1]$. As noted earlier, the proof of this theorem is mathematically revealing and leads to important general ideas. Yet a detailed proof of this result depends upon the deepest properties (completeness, connectedness) of the real number system. At the foundational level this proof is thrown into dispute by criticisms of constructivists and intuitionists. It would seem better to base the initial treatment of this property on geometric intuition rather than giving a proof that depends upon foundational matters that neither can nor should be properly aired in an introductory calculus course. The same comment applies to an extent to proofs of Rolle's theorem based upon the maximum value property of continuous functions. Why not rely on intuition in the cases where intuition is strong and correct and where a deeper proof is obscure and elusive (and generally not given)?

Existence: ($\exists$(Ham $\wedge$ Eggs) $\Rightarrow$ $\exists$Ham)

Existence and uniqueness proofs are often stressed. However, what it means for a mathematical object to exist is certainly debatable. We agree that Euclid gave a proof that the number of primes is not finite. However, must each of us affirm the existence of an infinite set of primes, and what does it mean to assert such existence? If we assert, "There are infinitely many primes," does this commit us to believing in the existence of infinite sets, in particular an infinite set of primes? Do such sets have an existence and reality on a par with that of the existence and reality of the hammer that catches one's thumb or one's thumb itself—to mention objects whose existence seems to be supported by direct personal experience? Or can we sidestep the question of existence and merely commit ourselves to a mathematical theory with existential quantifiers that includes a theorem whose normal interpretation would be, "There exists an infinite set of prime numbers"? The second view is purely formal and without commitments to the reality of such sets. The branch of metaphysics dealing with the theory of existence and being is ontology, and the nature of one's ontological commitments is a mostly philosophical matter, not a mathematical one.

The importance and meaning of the existence and uniqueness of $\sqrt{2}, {}^5\sqrt{7}, {}^3\sqrt{5}$, $\sqrt{2} + {}^5\sqrt{7}$, and so on, within the real numbers is that we need not worry about manipulating expressions like $(\sqrt{2} + \sqrt[5]{7})^5$ if we believe that they refer to unique numbers existing within the real number system. Within the context of the reals, such expressions are meaningful and follow the standard algebraic rules. In contrast, if we apply these rules to expressions that are ambiguous or have no interpretation in the reals, for example,

$$1 = 1^{1/2} = ((-1)(-1))^{1/2} = (-1)^{1/2}(-1)^{1/2} = ((-1)^{1/2})^2 = (-1)^1 = -1,$$

we may end up with an inconsistency. This incorrect string of equations can be ruled out or explained away in ways of no particular interest here. What is of interest is the conclusion that we can't assume that a^b has a unique and unambiguous interpretation satisfying $(a^b)^c = a^{bc}$ and $(ab)^c = a^c b^c$ for all reals a, b, c. The first string of equations shows that this assumption leads to a contradiction, and hence these assumptions cannot hold in any model. No such model can exist no matter what existence is taken to mean.

At a minimum, the statement that there exists an x within system S having property P means that $\exists x P(x)$ may be adjoined to the body of formulas assumed to hold in S without entailing a contradiction. This statement about an existentially quantified expression is formal and, as it stands, without ontological implications. This is mathematical existence in its weakest sense.

In the strongest sense, an assertion of existence is an assertion that something is. One can think of the reals existing as a concrete reality or as a platonic ideal. To say that there is a system $\mathbf{R}$ (unique up to isomorphism) satisfying the axioms of a complete ordered field in this sense is to assert the existence of a model (unique up to isomorphism) for those axioms. If one really believes (or knows) such a system exists in this sense, then *a fortiori*, one believes (or knows) that the axioms of the system must be consistent and categorical. I have neither that knowledge nor that belief.

Turning aside from absolute assertions, it is possible to give *relative existence* proofs. For example, if there is a model for Euclidean geometry, then a model for hyperbolic geometry can be constructed within that Euclidean model. It is not easy to construct and then verify the properties of models for the hyperbolic non-Euclidean plane within the Euclidean plane, but a sketch of one such construction, the Poincaré disk model, is presented in Box 3, page 92.

It is easier to construct a model for the natural numbers, $\mathbf{N}$, within a model $\mathbb{S}$ of set theory. For this construction it is convenient to redefine $\mathbf{N}$ to be the nonnegative integers (that is, to include 0) rather than having $\mathbf{N}$ consist of only the positive integers. In order to model $\mathbf{N}$ (redefined to contain 0) in $\mathbb{S}$ we must know the essential properties of $\mathbf{N}$. Peano listed these when he axiomatized the natural numbers. He specified that $\mathbf{N}$ was a system consisting of a distinguished initial element, here 0, and equipped with a successor mapping, σ, that maps $\mathbf{N}$ into $\mathbf{N}$. (Intuitively $\sigma(n)$ may be thought of as $n + 1$, but in Peano's axiomatization it is σ that will be used to define addition, not addition that is used to define σ.) Peano further required that the initial element, 0, not be the successor of any element and *he required that* $\mathbf{N}$ *consist precisely of the initial element*, 0, *and all of its successors and nothing more*. This is the intuitive content of Peano's axioms. The formal statement of them is given in Box 4, page 94.

To construct a model for $\mathbf{N}$ within set theory, we need only find the appropriate candidates to represent 0 and σ. What better candidate for $0 \in \mathbb{S}$ than $\emptyset$, the empty set? The choice for σ is not obvious, but one has been found. We define the successor of a set $S \in \mathbb{S}$ to be $\sigma^*(S) = S \cup \{S\}$, where $\{S\}$ is the set whose only element is S. Starting with $\emptyset \leftrightarrow 0$ and using σ^* and σ we may work out the correspondence between the natural numbers $\mathbf{N}$ and their model $\mathbf{N}^*$ in $\mathbb{S}$ as indicated in Box 5. To ensure that $\mathbf{N}^*$ satisfies Peano's Axiom 5, we require that $\mathbf{N}^*$ be the minimal set such that $\emptyset \in \mathbf{N}^*$ and $((\sigma^*(\mathrm{S}) \in \mathbf{N}^*)$ whenever $S \in \mathbf{N}^*)$.

There are certain technical problems with this construction. The mapping σ^* must be shown to satisfy Peano's conditions. Even the definition of σ^* raises questions, because σ^* appears to be a function defined on the class of all sets.

Box 3: A Model for a Non-Euclidean Geometry Constructed in the Euclidean Plane

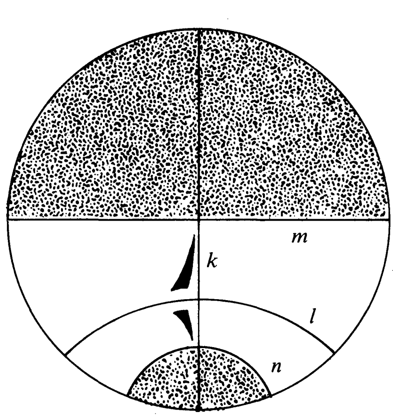
Poincaré devised a beautiful model for the hyperbolic plane within the Euclidean plane. This model shows that hyperbolic geometry is as consistent as Euclidean geometry. The key to all such relative models is to reinterpret the primitive notions of the system so that they have new properties. In this case we want the axioms of Euclidean geometry to be satisfied except for the postulate that asserts the uniqueness of parallels.

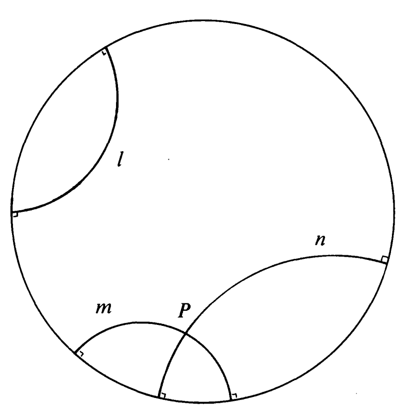
The primitive notions of plane geometry are points, lines, incidence, order and betweenness, angles, and congruence. To distinguish between the standard Euclidean use and the meaning in the Poincaré disk, I will put quotes around these words when they refer to the model.

The "plane" itself will consist of all the points interior to a fixed unit disk in the regular plane. The "points" of the model will be ordinary points. The "lines" will be arcs or segments that meet the boundary of our unit disk at right angles. The "incidence" of "lines" with each other and with "points" will be incidence in the usual sense for curves and points lying within the open disk. Likewise, "order" and "betweenness" will be the usual notions restricted to the "lines" and "points" lying within the open disk. No ambiguities of circular order enter, because the circular arcs that constitute "lines" in the interior of the disk never make up more than part of the circumference of a full circle. "Angles" between "lines" are measured in the usual way as angles between the arcs or segments representing the "lines."

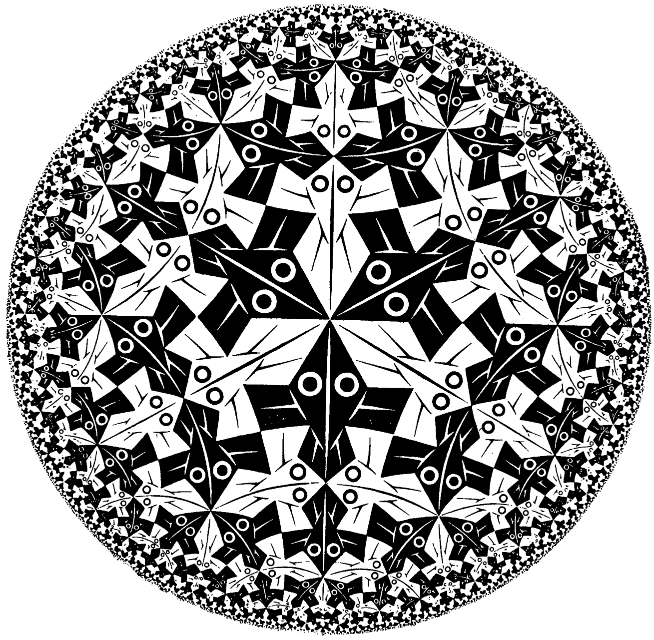
The idea of "congruence" is developed by defining the "reflection" in a "line" l to be the inversion of the open disk in the circle giving rise to l if l

"Lines" l, m, n in the Poincaré disk are segments or arcs orthogonal to the circle bounding the disk. The "line" n is the "reflection" (that is, the circular inversion) of "line" m in "line" l. "Line" k is "reflected" onto itself in "line" l. The two shaded regions are "congruent" under this reflection as are the hatched triangles. See figure below left.





"Lines" m and n parallel to "line" l through "point" P in the Poincaré disk. See above right.



CIRCLE LIMIT IV
by M. C. Escher
BEELDRECHT,
Amsterdam/VAGA,
New York 1981
Collection Haags
Gemeentemuseum

comes from a circle or by the ordinary reflection of the open disk in the line representing l if l is a diameter of the fixed unit disk. The "isometries" of the Poincaré disk are then all mappings generated by compositions of "reflections." Two figures are "congruent" if one can be carried exactly onto the other by an "isometry."

The usual axioms about incidence, points and lines, and order and betweenness are easily verified. The fact that inversions preserve angles and preserve the class consisting of circles and lines tells us that the "isometries" of the disk model preserve "lines" and "angle measure." These isometries obviously distort Euclidean distance, but there is a hyperbolic "distance" that they preserve. The definition of this "distance" is technical and may be found in texts such as Greenberg [1980]. It is the length you would get by fixing an appropriate unit "segment" in the model and then moving it about to measure things using "isometries" just as you would a ruler in the ordinary plane.

Despite the distortions of these "isometries," the lengths so defined will always be consistent, and the congruence postulates of plane geometry will be satisfied. These two facts are not at all evident and are the heart of the verification of the Poincaré model. For details, see Greenberg [1980] or Hilbert and Cohn-Vossen [1932].

Granted the correctness of the above assertions, it is clear that the "lines" m and n are distinct parallels to "line" l through "point" P. This shows the relative consistency of the denial of the uniqueness of parallels with the axioms of Euclidean geometry. Beyond that, this model opens up the possibility of many fascinating and beautiful representations of constructions and tilings in the hyperbolic plane. M. C. Escher's Circle Limit pictures are tilings of the Poincaré disk and suggest the richness of the possibilities.

Box 4: The Peano Postulates for the Natural Number System

The natural number system $\mathbf{N}$ consists of a set $\mathbf{N}$ together with a distinguished initial element 1 and a successor function $\sigma: \mathbf{N} \to \mathbf{N}$ satisfying the following postulates:

(1) $1 \in \mathbf{N}$;
(2) If $x \in \mathbf{N}$, then $\sigma(x) \in \mathbf{N}$;
(3) If $x \in \mathbf{N}$, then $\sigma(x) \neq 1$;
(4) If $\sigma(x) = \sigma(y)$ for $x, y \in \mathbf{N}$, then $x = y$;
(5) For all sets $M \subseteq \mathbf{N}$, if M satisfies (i) $1 \in M$ and (ii) $x \in M$ implies $\sigma(x) \in M$, then $M \supseteq \mathbf{N}$.

Postulate (5) differs from the other postulates in that it is quantified over *all subsets* M of $\mathbf{N}$, whereas the other axioms are quantified over *elements* of $\mathbf{N}$. In the technical language of logic, axioms that are allowed to range over sets, not simply individuals, are *second order axioms*. The natural numbers can be axiomatized in a first-order language by replacing (5) by a countable schema of axioms represented by (5′) below.

(5′) If P is any property such that (i) $P(1)$ is true and (ii) for all $k \in \mathbf{N}$, $P(k)$ implies $P(\sigma(k))$, then $P(k)$ is true for all $k \in \mathbf{N}$.

As stated, (5′) appears to be a single axiom, but we can make its content more explicit by giving an explication of what is meant by a property P. By "a property" we mean a property statable in terms of the language of the theory of the integers, $\mathbf{N}$. This language is assumed to have a finite or countable* number of primitive signs, terms, and relations, and the allowable expressions in this language are assumed to consist of certain finite strings of the symbols for these elements. If we order the symbols in the language, all finite strings of such symbols may be put into order first by string length; and strings of the same length are to be ordered lexicographically once we decide an alphabetical ordering for the basic symbols. This scheme throws all such strings into a one-to-one correspondence with the natural numbers (as we understand these informally). Hence the number of such strings (and hence the number of properties, P, they describe) is at most countable. This argument parallels that used in Box 2.

By making the p's in (5′) explicit in this way we arrive with a countable family of first-order axioms that are implied by (5′) and that can replace (5) if we wish to have a first-order set of axioms for $\mathbf{N}$.

*As we will see, the terms "finite" and "countable" do not have absolute meanings. However, I use them here in their usual informal senses. As long as we are consistent in interpreting these notions, all the usual results hold no matter how these notions are interpreted in a model.

Box 5: The Natural Numbers in the Context of Set Theory

A model for the nonnegative integers, $\mathbf{N} = \{0, 1, 2, 3, 4, \ldots\}$, can be constructed in the universe of sets $\mathbb{S}$ by setting up the following correspondence.

$$
\begin{array}{ccl}
\underline{\text{Integers}} & & \qquad\qquad\qquad\qquad \underline{\text{Sets}} \\
0 & \longleftrightarrow & S_0 = \emptyset \\
1 & \longleftrightarrow & S_1 = S_0 \cup \{S_0\} = \emptyset \cup \{\emptyset\} = \{\emptyset\} \\
2 & \longleftrightarrow & S_2 = S_1 \cup \{S_1\} = \{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\} \\
3 & \longleftrightarrow & S_3 = S_2 \cup \{S_2\} = \{\emptyset, \{\emptyset\}\} \cup \{\{\emptyset\}, \{\emptyset\}\}\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \\
\vdots & & \vdots \\
k & \longleftrightarrow & S_k = S_{k-1} \cup \{S_{k-1}\} \\
 & & S_k = \{\emptyset, \\
 & & \qquad \{\emptyset\}, \\
 & & \qquad \{\emptyset, \{\emptyset\}\}, \\
 & & \qquad\quad \vdots \\
 & & \text{Continued for } k \text{ lines}\} \\
\vdots & \vdots & \vdots
\end{array}
$$

This tedious construction allows an interpretation of the idea of a successor ($k \to k + 1$ for integers k) in terms of sets ($S \to S \cup \{S\}$ for sets S). Such considerations allow us to imbed number theory in set theory. What we gain is the knowledge that all number theoretic notions can be stated in set-theoretic terms and that the consistency of set theory would imply that of number theory. What we lose is the simplicity and directness of our common intuition of number. The number three expressed in full in this system is $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$ compared to which the Roman numeral III is simplicity itself.

Because of this we cannot think of σ^* as a *set* of ordered pairs in the usual way. If σ^* were regarded as a *set* of ordered pairs, then the first elements of those pairs (which constitute the class $\mathbb{S}$ of all sets) would also be a set, and it is known that systems in which the collection of all sets is regarded itself as a set are plagued by paradoxes (for example, Russell's paradox, the Burali-Forti paradox, and so on). A good account of these paradoxes may be found in Quine [1962].

A second question is whether any family of sets, W, in $\mathbb{S}$ has both properties (i) $(\emptyset \in W)$ and (ii) $(\sigma^*(S) \in W$ whenever $S \in W)$. This is essentially the question of whether $\mathbb{S}$ contains infinite sets or infinite families of sets. There are axioms (such as the axiom of infinity) that require $\mathbb{S}$ to contain such sets. Finally, there is the question of whether $\mathbb{S}$ contains a minimal set with properties (i) and (ii). This question can be disposed of. It is clear that if each member of a family $\mathfrak{F}$ of sets possesses properties (i) and (ii), then the intersection, $\bigcap F$ $(F \in \mathfrak{F})$, also possesses properties (i) and (ii). The intersection of all sets with properties (i) and (ii) can be expected to yield a minimal set with properties (i) and (ii). Within $\mathbb{S}$ we call this minimal set $\mathbf{N}^*$.

Note that the construction of $\mathbf{N}^*$ as an intersection of sets is dependent upon the sets and operations available in $\mathbb{S}$, and it is not absolute, but is relative to $\mathbb{S}$. However, given $\mathbb{S}$ and sufficient apparatus in $\mathbb{S}$, $\mathbf{N}^*$ can be constructed and is unique. In the context of the universe of sets, $\mathbb{S}$, $\mathbf{N}^*$ is the standard model of the natural numbers.

In contrast to this abstract construction of $\mathbf{N}^*$ consider the construction of a group of order three and the proof that all groups of order three are isomorphic. These are as concrete as anyone might wish; the group symbols and the operation table are easily written down in their entirety. Can an equally persuasive demonstration be given of the existence and uniqueness of an object $\mathbf{N}$ representing the intended model for the natural numbers? A possible affirmative answer has been outlined above based on set theoretic constructions. We might say that the *intended* model for $\mathbf{N}$ is the model naturally constructed in the *intended* (or "real") model for set theory. This begs the question. How can the intended model of set theory be distinguished from the unintended models? Worse yet, this is an attempt to explain and describe the integers (which are relatively familiar) in terms of sets which are mysterious and unfamiliar. The diversity of possible models for the integers will be indicated in the next section.

Nonmathematical existence proofs may be likened to saying, "If we had ham, we could have ham and eggs, if we had eggs." Mathematical existence proofs are often far weaker. They are like saying, "If we had ham and eggs, we could have ham." That is, in mathematics we assume the existence of a more complex and comprehensive object to infer the existence of a simpler and more specialized one. "If the 3real numbers exist, then $\sqrt{2}$ exists within the real numbers." "If we are given a universe of sets, we can find within it well-defined structures representing the integers and the real numbers." A fixed context is assumed for these statements, and existence and uniqueness are proved only relative to that context. In fact, the desired entities ($\sqrt{2}$ and so forth) are assumed to exist within the universe considered. Our constructions merely single them out.

These constructions are in the spirit of the construction of the model $\mathbf{N}^*$ for the natural numbers constructed within set theory as outlined above. These can be found in standard texts on the foundations of analysis and real numbers. We emphasize that these constructions are made relative to some structure like the universe of sets, a structure that is itself not completely specified.

What Has Work on the Foundations of Mathematics Revealed?

The real foundations of mathematics are deep. To discover them we would have to understand what it is that makes mathematical reasoning convincing and why mathematics is useful. Such questions are empirical, psychological, and philosophical and are not usually considered part of the mathematical study of the foundations of the science.

Within mathematics, foundations is the study of axiomatics and the exploration of the various bases (logic, set theory, and so forth) for the subject itself. The discovery of the independence of the parallel postulate and the existence of models for non-Euclidean geometry in the Euclidean plane are striking early results in foundations (Box 3, page 92).

The discovery of non-Euclidean geometries focused attention on axiomatics in other areas. Progress was made by axiomatizing the natural numbers (Peano) and by building set-based theories of the real numbers and the natural numbers (Dedekind and others). Russell and his school sought to establish a secure foundation for mathematics in logic alone (logicism), while Hilbert and his school attempted to use formal methods to establish the completeness and consistency of the axiomatic basis for mathematics. Positive results had been obtained and, despite some objections from constructivists and intuitionists and some problems with various paradoxes, the prospects for success seemed excellent. For an overview see p. 164 (I.1) "Foundations of Mathematics" in Mathematical Society of Japan [1977]; Chapter III, "A Critique of Mathematical Reasoning," in Kleene [1952], or Chapter I, "The Antinomies," in Fraenkel and Bar-Hillel [1958].*

The hopes these approaches raised for new and completely secure foundations for mathematics were dampened by, among other things, the discovery of nonstandard models for the integers (such systems admit *uncountable* models) and for the real numbers and set theory (such systems admit *countable* models). These results are due to Lowenheim and Skolem. Though the existence of such models may at first seem astonishing if not contradictory, on further consideration it proves to be natural and consistent.

These results reflect the limitations of language. We will see some of these limitations in the subsection on nonstandard models. Because only a countable collection of real numbers can be named (see Box 1), it is not surprising that a countable model can be found for the real numbers. I once asked Paul Bernays whether this was a reasonable way to think of the paradox of a countable model for the reals. He answered "yes" and the tone of his answer carried the unspoken message "as everybody knows."

Further setbacks to the program of providing adequate and provably consistent foundations were found in the limitative results of Kurt Gödel and Alfred Tarski. Gödel showed that the usual theory of the natural numbers, if consistent, contains statements that are true in the intended model but unprovable. This is not simply a deficiency of the axioms used. There will always be such unprovable statements for

---

*Fraenkel and Bar-Hillel give an extraordinarily deep and wise treatment throughout the cited book, although it necessarily leaves out recent and important developments such as Cohen's invention of the method of forcing and his proof of the independence of the axiom of choice and of the continuum hypothesis (see Cohen [1966], Cohen and Hersh [1967], Jech [1970], Jech [1978]) and the related development of Boolean valued models for attacking independence problems (see Steen [1971], Scott [1967], Rosser [1969]).

any recursively enumerable axiom system (see Nagel and Newman [1956, 1958], Hofstadter [1979], Dawson [1979]). Similar limitative results were obtained by Tarski, who showed that the notion of truth for systems rich enough to contain arithmetic can only be formalized within a metatheory that is of a higher type than the theory under investigation. The escalation of complexity required to formulate the notion of truth in a system is a general problem and is not restricted to the theory of the natural numbers (see Tarski [1931] and Tarski [1969]). Finally, Gödel established that the consistency of arithmetic cannot be proved by finitistic methods, that is, by methods formalizable in arithmetic itself. This last result was a disappointment to those who hoped that at least the consistency of Peano arithmetic could be established by such means.

It would be wrong to give the impression that all the results obtained have been negative or limitative. Many important completeness results and consistency results have been proved as well. For example, Gödel proved the completeness of the predicate calculus and the relative consistency of the continuum hypothesis ($2^{\aleph_0} = \aleph_1$) and the axiom of choice with the other axioms of Zermelo-Fraenkel set theory.

Nonstandard Models: Compared to What Standard?

If we have a system $\mathbf{N}$ satisfying the Peano axioms with the induction principle given in first order form by a countable axiom schema stated in terms of properties (see Box 4), then we may adjoin to our language a special constant symbol c, and to our axioms the sequence of axioms $A(k)$, one for each integer k, stating that there exists $c \in \mathbf{N}$ such that $c > k$. Any finite collection (say, $\{A(k) | k \in F\}$ with F a finite subset of $\mathbf{N}$) of these new axioms can be satisfied in the model simply by giving c a value larger than any of the k's in the finite collection of statements $A(k)$ to be satisfied (that is, c should be larger than any k in K).

For safety we may also augment the Peano axioms and the axioms $A(k)$ by requiring that every formula P that holds in $\mathbf{N}$ also be taken as an axiom for the new system $\mathbf{N}'$ that we are going to construct. If we do this, we will ensure that, within the power of the language used to describe the integers, $\mathbf{N}$ and $\mathbf{N}'$ will be indistinguishable (technically such indistinguishable models are called *elementarily equivalent*). We note that there is no general way of deciding whether a particular statement P holds in $\mathbf{N}$ short of replacing all variables in P with all allowable values in $\mathbf{N}$; nor is it possible to systematically enumerate all of the statements P that are valid in $\mathbf{N}$ (Gödel undecidability and related results). If we assume we are given a model $\mathbf{N}$, we may as well assume also that we are given the collection of statements that are valid in $\mathbf{N}$.

The arguments above show that any finite collection of these axioms is consistent and hence the complete collection of axioms is consistent. (This is the "finiteness" theorem, and it is a consequence of the fact that the proofs allowed are finite in length and begin with finitely many assumptions. Hence, if a contradiction can be derived from the full set of assumptions, it can be derived from a finite subset of those assumptions.)

Thus, we know that the Peano axioms together with the above-mentioned complete collection of axioms, including the axioms $A(k)$ for k in $\mathbf{N}$ are consistent and that each finite subset of these axioms has a model. Thus, there exists a model for the Peano axioms together with all the axioms $A(k)$ for k in $\mathbf{N}$. This result may be obtained by invoking either the Gödel-Henkin completeness theorem or the compactness theorem applied to predicate calculus. See, for example, Bell and Slomson [1969] or other books on logic and model theory. Taken in full generality, the compactness theorem implies the axiom of choice, but Peter Hinman pointed out to me that the existence of a model like $\mathbf{N}'$ for the Peano axioms and the $A(k)$'s can be established by more constructive methods (see, for example, Hunter [1971] p. 45 et seq.). The existence of such an $\mathbf{N}'$ and its properties rests upon the existence of $\mathbf{N}$ and its properties, from which $\mathbf{N}'$ can be obtained by an inductive construction.

The $\mathbf{N}'$ obtained satisfies the Peano axioms; but it is a proper extension of the model $\mathbf{N}$ with which we began, because the constant c is given a value in $\mathbf{N}'$ greater than any of the k's in $\mathbf{N}$ (which is imbedded in $\mathbf{N}'$). We call $\mathbf{N}'$ a nonstandard model of the natural numbers because it contains the element denoted by c that is, in a manner of speaking, infinite.

How does $\mathbf{N}'$ look from the inside? Might we find a formula $P(x)$ that is satisfied when x is in $\mathbf{N}$ but not when x is outside of $\mathbf{N}$ in $\mathbf{N}'$? This is impossible because $P(x)$ would be valid for all x in $\mathbf{N}$ and we augmented our axiom system by including every statement P valid in $\mathbf{N}$. Thus P would be an axiom for $\mathbf{N}'$ and hence valid in the model $\mathbf{N}'$. Even if we only require that $\mathbf{N}'$ satisfy the Peano axioms together with the axioms $A(k)$ for k in $\mathbf{N}$, we can prove that no predicate $P(x)$ can be true for all x in $\mathbf{N}$ and false for every x in $\mathbf{N}'$ lying outside of $\mathbf{N}$. To see this, consider such a P. The implication "$P(x)$ implies $P(x+1)$" would be true for any x in $\mathbf{N}$ because $P(x+1)$ would be true. The implication "$P(x)$ implies $P(x+1)$" would be true for every x in $\mathbf{N}'$ lying outside of $\mathbf{N}$, because $P(x)$ would be false. Hence "$P(x)$ implies $P(x+1)$" would be true for all x in $\mathbf{N}'$ and $P(0)$ would be true. It would follow by induction that $P(x)$ is true for all x in $\mathbf{N}'$, a contradiction to P's role of distinguishing elements of $\mathbf{N}$ from those lying in $\mathbf{N}'$ outside of $\mathbf{N}$.

The fact is that $\mathbf{N}'$ can be constructed so that it is impossible to distinguish $\mathbf{N}'$ from $\mathbf{N}$ by examining properties describable in the language used to describe integers and their properties. This is a consequence of the limitations of language, which simply does not and cannot contain a predicate P that makes the desired distinction.

But what about the infinite element c that $\mathbf{N}'$ contains? The property $c > k$ holds for all k in $\mathbf{N}$, where we think of $\mathbf{N}$ as the genuine set of integers. Thus c appears to be infinite relative to $\mathbf{N}$. Within $\mathbf{N}'$ the interval from 1 to c is simply an initial segment, call it I, of what is a model for the integers, that is, a system satisfying the Peano axioms. How does this interval look in $\mathbf{N}'$? In naive set theory the first choice for the definition of finiteness is to define a set to be finite if it can be placed in one-to-one correspondence with an initial segment of the integers. In this sense I is a finite set relative to $\mathbf{N}'$. If we could describe $\mathbf{N}$ as a subset of $\mathbf{N}'$ in the language of the theory of the integers, we could show that I was infinite in an absolute sense, but we argued in the preceding paragraph that no predicate in our language can describe $\mathbf{N}$ as a subset of $\mathbf{N}'$. Thus, the approach to the distinction between finite and infinite sets by defining finite sets to be equivalent to initial segments of the integers cannot be used to prove I is infinite.

Might I be shown to be infinite by other means? Dedekind proposed that a set, S, be termed infinite precisely when S can be put into one-to-one correspondence with a proper subset of itself. If we enrich the language of the natural numbers and the model we are working in to include functions and so on in a consistent way, can we show I is Dedekind infinite? This is impossible, because the same induction proof that shows there is no one-to-one map of $1, 2, \ldots, k$ into a proper subset of itself in the usual integers is valid line for line in the theory of $\mathbf{N}'$. This too is a limitation of our language. *We have no way of describing or constructing the function which would show that I is infinite*. That function does not exist *within the system* $\mathbf{N}'$ even when this system is enriched to include functions.

If the only way that $\mathbf{N}$ can be distinguished from $\mathbf{N}'$ is by relying on the special symbols k, which represent elements of $\mathbf{N}$ in $\mathbf{N}'$, and this set of elements cannot be singled out within $\mathbf{N}'$, then what is to prevent me from having started this process with an $\mathbf{N}$ that was already nonstandard (that is, nonstandard with respect to your standard $\{1, 1+1, 1+1+1, \ldots\}$)? Nothing. Is it possible, then, that a system $\mathbf{M}$, which I think of as the standard natural numbers, the intended model for the natural numbers, $\{1, 1+1, 1+1+1, \ldots\}$, might itself be nonstandard with respect to some other system (for example, your system that you also think of as $\{1, 1+1, 1+1+1, \ldots\}$)?

It was Skolem who observed that the Peano axioms are not categorical; that is, these axioms admit various models which are not isomorphic (see Kleene [1952], p. 432 and p. 430). Skolem worked with a first-order system of axioms, but the fact that set theory is not categorical suggests that we cannot achieve categoricity for the integers by relying on axioms based on set theory, that is, by second order axioms. I go further and suggest that we might entertain different ideas about the intended model for the natural numbers. This seems inconceivable because of the strong intuitive feeling each of us has for the integers. However, the weight of evidence suggests that the system of integers is not yet, nor ever will be, specified in a completely unambiguous way. At the present it seems that there is no absolute way to distinguish certain models of the natural numbers and that, for this reason, the existence of a unique standard system commonly referred to as the natural numbers cannot be taken for granted. The same uncertainty exists in even greater measure for the notions and constructs of set theory.

Which of these are finite and which are infinite?

$0, 1, 2, 3, \ldots$	The system $\mathbf{N}$.
$0, 1, 2, 3$	A finite set, an interval in $\mathbf{N}$.
$0, 1, 2, 3, \ldots, 100^{100}$	A larger finite set, the interval from 0 to 100^{100} in $\mathbf{N}$.
$0, 1, 2, \ldots, k$	An arbitrary finite interval in $\mathbf{N}$.
$0, 1, 2, \ldots, l$	Where l is the least integer such that $l > 2$ and $n^l + m^l = q^l$ has nontrivial integer solutions for m, n, and q. Is this a finite interval?
$0, 1, 2, \ldots, c$	A pseudo-finite interval in $\mathbf{N}'$.

Standard Notions

Archimedes asserted that, given a fulcrum and a lever sufficiently long and a place to stand on, he could move the earth. Our position with regard to foundations is similar. Given a standard universe of sets and sufficient apparatus, we can construct integers, real numbers, and other objects, most yet undreamed of. It is this "given" fulcrum that we think of as standard and upon which the lever of our arguments rests. But how solid and definite is this "given" and upon what does it rest? The postulate of an absolute and unambiguously specified standard mathematical universe seems unnecessary and, by the arguments above, untenable. Efforts to construct mathematics in an absolute fashion ex nihilo are doomed to failure, except for those parts that can be constructed by operations, each step of which can be seen and examined.

There is nothing wrong with using standard objects and referring to them as such. But it is well to bear in mind that if we regard them as being specified by their axioms, the description is essentially incomplete. The other horn of the dilemma is to regard standard objects as having at least Platonic reality. This leaves us with the problem of having only an incomplete description of these objects by axioms and having no direct access to them by which their other properties can be ascertained. Belief in the existence of such objects is an act of faith on a par with belief in an unknowable deity.

Holding to the Good

Even the most elementary notions of mathematics are subject to differing interpretations. Mathematics is like an intricate tapestry, the work of human hands and minds, not like a crystal, the work of undeviating natural forces. It is insight which is father to the rules and axioms of mathematics, and it is insight, not the slavish obedience to rules, that we must seek ourselves and try to develop in students.

As Fraenkel and Bar-Hillel pointed out, the discovery of the relative nature of such seeming absolute constructs and notions as the integers and cardinality does not directly strengthen any of the recognized schools of thought or the foundations of mathematics.

> More generally, to a given axiomatic set theory Σ there always exists a more comprehensive theory in which all infinite sets of Σ prove equivalent, namely denumerable; the enumeration from outside utilizes the structure of the system as a whole which cannot be reached by operations within the system . . . . Either the concept of subset of a given set is taken as a Platonic reality, hence non-denumerability is conceived as something absolute—then axiomatization seems impracticable in view of Skolem's paradox and we retreat to naive set theory with its menace of antimonies. Or else we accept axiomatization in whatever shape, then the notions "denumerable" and "non-denumerable" turn out to be relative and we have to accept the *relativization of cardinals* first pointed out by Skolem . . . . This relativism, incidentally, concerns not only more than denumerable sets but *finite sets as well* . . . . This double relativism should not be exploited to strengthen any of the existent philosophies of mathematics. If, on the one hand, it seems to support the intuitionistic slogan that by formalistic (axiomatic) methods no reality can be reached and everything remains vague and relative, on the other hand the supposedly absolute original intuition of positive integer (Chapter IV, paragraph 5, forfeits as well its pretended evidence owing to indistinct passages from finiteness to transfinite denumerability.) Thus not only is categoricalness or monomorphism of set theory called in question but also the absolute authority and distinctness of the notion of integral number. (From pages 108 and 109 of Fraenkel and Bar-Hillel [1958].)

However, this is no cause for despair, but rather a cause to turn from an overanxious search for certainty and the ultimate foundations of the subject toward relatively free exploration of concepts and entities. We should maintain skepticism about the sense in which mathematical objects may be said to exist and about the absoluteness of our conclusions. Use inaccessible cardinal numbers or a choice function on a family of nonempty sets if you wish, but don't expect everyone to agree that such things exist. Both the methods and the results of mathematics are of value independent of any link to absolute truth or Platonic reality and without the need for them to be certifiably consistent statements about some formal system.

The power of mathematics is evident in practice and in the contributions it has made to culture, philosophy, and the life of the mind. This power and these contributions remain, whatever arguments may take place concerning the foundations of the subject. Past and present controversies on the nature and foundation of mathematics are valuable themselves for the rich philosophical material they have yielded.

Eugene P. Wigner [1967] described mathematics as "the science of skillful operations with concepts and rules invented for just this purpose." This is a pleasing description, free of ontological implications.

Willard Van Orman Quine, in his essay "On What There Is," urges conservatism about the existence of things. He suggests, in effect, that we get along with as few ontological assumptions as we can, but that we be ready to make whatever assumptions are needed. Expressing a similar flexibility, J. B. S. Haldane borrowed from Hamlet to give his thoughts on science, ontology, and philosophy: "I suspect that there are more things in heaven and earth than are dreamed of in any philosophy. That is the reason why I have no philosophy myself, and must be my excuse for dreaming" (Haldane [1927]).

Perhaps it is best to follow Haldane and neither to try to have nor to expect to find a complete and coherent philosophy for mathematics. Even if mathematics and its objects are only inventions of the human mind, it is a fascinating, useful, and surprising domain. Proofs are some of the tools that we use in creating and exploring mathematics. Use them for what they reveal, but be a slave neither to formalism nor to rigor. Free exploration should be combined with careful checking and a good measure of skepticism. In this way, progress will be assured and, in the end, errors corrected, and proofs seen in the proper perspective.

Acknowledgments

I am indebted to Peter Hinman, Victor Klee, Morris Kline, W. A. J. Luxemburg, and Sherman Stein, among others, for helpful conversations on the subject of this article. Those named above also read and made helpful comments on drafts of this article. Thanks go to Nancy Renz, who not only helped with the editing but also transformed the scrawly manuscript into beautiful typescript through many drafts, and to my editors: Donald Albers, who encouraged me to write this article and suggested many improvements and additions, and Patricia Mittelstadt, who reformed many unclear, awkward, or infelicitous passages in drafts.

There are broader debts to the ideas of others. All ideas arise in a specific context. At present there is an active reexamination of the foundations of mathematics, and this article was called forth, in part, by this interest in foundations.

BIBLIOGRAPHY

Apostol, T. M. [1967], *Calculus, Second Edition*, Ginn-Blaisdell.

Bell, J. L., and Slomson, A. B. [1969], *Models and Ultraproducts: An Introduction*, North-Holland.

Cohen, P. J. [1966], *Set Theory and the Continuum Hypothesis*, W. A. Benjamin.

Cohen, P. J., and Hersh, R. [December 1967], "Non-Cantorian Set Theory," *Scientific American*, reprinted in *Mathematics in the Modern World*, Klein, M. ed. [1968], W. H. Freeman and Company.

Courant, R. [1937], *Differential and Integral Calculus, Second Edition* translated by McShane, E. J., Interscience Publishers, New York.

Dawson, J. [November 1979], "The Gödel Incompleteness Theorem from a Length-of-Proof Perspective," *Am. Math. Monthly*, Vol. 86, No. 9.

Fraenkel, A. A., and Bar-Hillel [1958], *Foundations of Set Theory*, North-Holland Publishing Company.

Greenberg, M. J. [1980], *Euclidean and Non-Euclidean Geometries*, W. H. Freeman and Company.

Haldane, J. B. S. [1927], *Possible Worlds.*

Hamming, R. W. [February 1980], "The Unreasonable Effectiveness of Mathematics," *Am. Math. Monthly*, Vol. 87, No. 2.

Hilbert, D., and Cohn-Vossen [1932], *Geometry and the Imagination*, Julius Springer, Berlin, 1932, and Chelsea, New York, 1952.

Hofstadter, D. R. [1979], *Gödel, Escher, Bach: An Eternal Golden Braid,* Basic Books.

Hunter, G. [1971], *Metalogic: An Introduction to the Metatheory of Standard First Order Logic,* University of California Press.

Jech, T. J. [1970], Lectures in set theory with particular emphasis on the method of forcing, *Lecture Notes in Mathematics 217*, Springer.

Jech, T. J. [1978], *Set Theory*, Academic Press.

Kleene, S. C. [1952], *Introduction to Metamathematics*, Van Nostrand.

Mathematical Society of Japan [1977], *Encyclopedic Dictionary of Mathematics* edited by Shokichi Iyanaga and Yukiyosi Kawada, The MIT Press.

Nagel, E., and Newman, J. R. [June 1956], "Gödel's Proof," *Scientific American*, reprinted in *Mathematics in the Modern World* [1969], Kline, M., ed., W. H. Freeman and Company.

Nagel, E., and Newman, J. R. [1958], *Gödel's Proof,* New York University Press.

Quine, W. V. [April 1962], "Paradox," *Scientific American*, reprinted in *Mathematics in the Modern World*, Kline, M., ed., [1968], W. H. Freeman and Company.

Rosser, J. B. [1969], *Simplified Independence Proofs: Boolean Valued Models of Set Theory*, Academic Press.

Scott, D. [1967], "A Proof of the Independence of the Continuum Hypothesis, " *Mathematical Systems Theory*, Vol. 1, No. 2, Springer-Verlag.

Spivak, M. [1967], *Calculus*, W. A. Benjamin.

Steen, L. A. [August 1971], "New Models of the Real-Number Line," *Scientific American.*

Tarski, A. [1931], "The Concept of Truth in Formalized Languages," reprinted in English in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938 by Alfred Tarski* [1956], translated by J. H. Woodger, Oxford at the Clarendon Press.

Tarski, A. [June 1969], "Truth and Proof," *Scientific American.*

Weyl, H. [1949], *Philosophy of Mathematics and Natural Science*, Princeton University Press.

Wigner, E. P. [1967], from "The Unreasonable Effectiveness of Mathematics in Natural Science," reprinted in *Symmetries and Reflections.*

Peter Renz received his Ph.D. from the University of Washington and taught at Reed College and Wellesley College before becoming an editor with W. H. Freeman and Company. His published research has been in geometry, topology, analysis, and graph theory, and he has written expository pieces on more general subjects. As an editor, he has published books whose subjects range from ninth-grade algebra to advanced topics in mathematics, computer science, and physics, and he has had the occasion and the need to discuss the role of proofs with a wide range of teachers and mathematicians—particularly as this relates to classroom practice and textbooks.