NONSTANDARD ANALYSIS

This mathematical theory has restored infinitesimals to good standing. They had been employed since antiquity, but often with doubts, to solve such problems as finding a circle's area

by Martin Davis and Reuben Hersh

Nonstandard analysis, a new branch of mathematics invented 10 years ago by the logician Abraham Robinson, marks a new stage of development in several famous and ancient paradoxes. Robinson, now at Yale University, has revived the notion of the "infinitesimal"—a number that is infinitely small yet greater than zero. This concept has roots stretching back into antiquity. To traditional, or "standard," analysis it seemed blatantly self-contradictory. Yet it has been an important tool in mechanics and geometry from at least the time of Archimedes.

In the 19th century infinitesimals were driven out of mathematics once and for all, or so it seemed. To meet the demands of logic the infinitesimal calculus of Isaac Newton and Gottfried Wilhelm von Leibniz was reformulated by Karl Weierstrass without infinitesimals. Yet today it is mathematical logic, in its contemporary sophistication and power, that has revived the infinitesimal and made it acceptable again. Robinson has in a sense vindicated the reckless abandon of 18th-century mathematics against the strait-laced rigor of the 19th century, adding a new chapter in the never ending war between the finite and the infinite, the continuous and the discontinuous.

In the controversies over the infinitesimal that accompanied the development of the calculus, Euclid's geometry was the standard against which the moderns were measured. In Euclid both the infinite and the infinitesimal are deliberately excluded. We read in Euclid that a point is that which has position but no magnitude. This definition has been called meaningless, but perhaps it is just a pledge not to use infinitesimal arguments. This was a rejection of earlier concepts in Greek thought. The atomism of Democritus had been meant to refer not only to matter but also to time and space. But then the arguments of Zeno had made untenable the notion of time as a row of successive instants, or the line as a row of successive "indivisibles." Aristotle, the founder of systematic logic, banished the infinitely large or small from geometry.

Here is a typical example of the use of infinitesimal arguments in geometry:

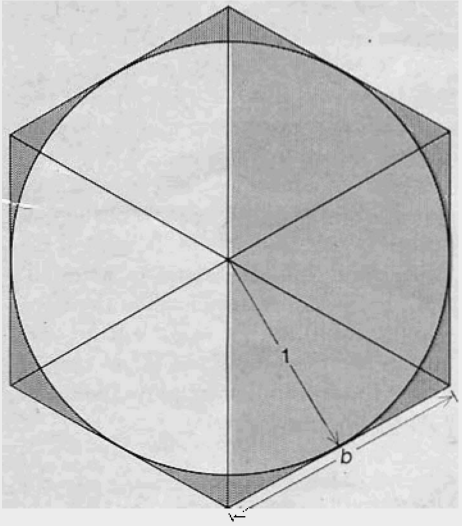
"We wish to find the relation between the area of a circle and its circumference. For simplicity we suppose that the radius of the circle is 1. Now, the circle can be thought of as composed of infinitely many straight-line segments, all equal to each other and infinitely short. The circle is then the sum of infinitesimal triangles, all of which have altitude 1. For a triangle the area is half the base times the altitude. Therefore the sum of the areas of the triangles is half the sum of the bases. But the sum of the areas of the triangles is the area of the circle, and the sum of the bases of the triangles is its circumference. Therefore the area of the circle of radius 1 is equal to one half its circumference."

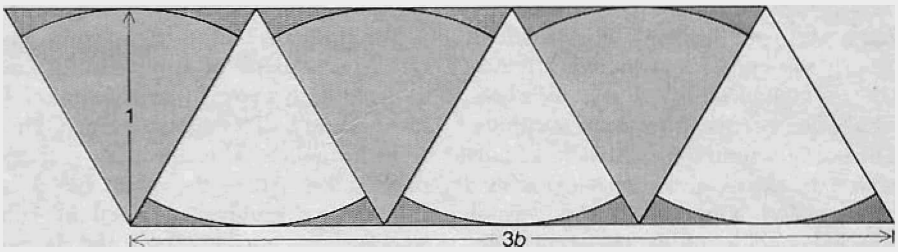
This argument, which Euclid would have rejected, was published in the 15th century by Nicholas of Cusa. The conclusion is of course true, but objections to the argument are not hard to find. The notion of a triangle with an infinitely small base is elusive, to say the least. Surely the base of a triangle must have length either zero or greater than zero. If it is zero, then the area is zero, and no matter how many terms we add we can get nothing but zero. On the other hand, if it is greater than zero, no matter how small, we will get an infinitely great sum if we add infinitely many terms. In neither case can we get a circle of finite circumference as a sum of infinitely many identical pieces.

The essence of this rebuttal is the assertion that even a very small nonzero number becomes arbitrarily large if it is added to itself enough times. Because the assertion was first made explicit by Archimedes, it is called the Archimedean property of the real numbers. An infinitesimal, if it existed, would be precisely a non-Archimedean number: a number greater than zero, which nevertheless remained less than 1, say, no matter how (finitely) many times it was added to itself. Archimedes, working in the tradition of Aristotle and Euclid, asserted that every number is Archimedean; there are no infinitesimals. Archimedes, however, was also a natural philosopher, an engineer and a physicist. He used infinitesimals and his physical intuition to solve problems in the geometry of parabolas. Then, since infinitesimals "do

METHOD OF EXHAUSTION is employed to prove indirectly that the area of a circle with radius 1 is half its circumference. In the proof on the opposite page a polygon is circumscribed on the circle (*top*), creating a number of triangles for which the areas can be calculated readily. By increasing the number of sides of the polygon, as in polygon *B* and polygon *N*, the triangles increase in number and become thinner, and the difference in area of circle and polygon becomes smaller. The difference will never be zero, however, for a polygon having any finite number of sides. Standard analysis avoids this difficulty by stating that, as the number of sides increases to infinity, the polygon's area approaches the circle's area as a limit. Nonstandard analysis avoids the concept of limit for a more suggestive explanation using a polygon with infinitely many sides, each side having infinitesimal length.

POLYGON A





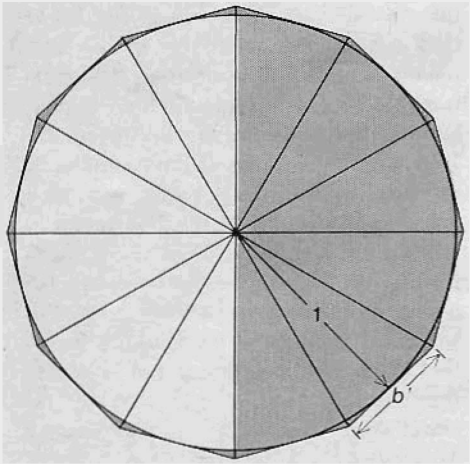
PERIMETER $P=6b$

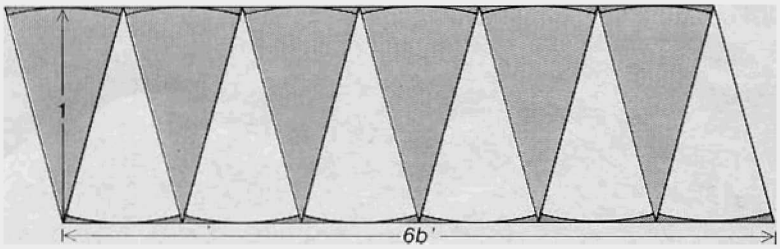
AREA OF POLYGON $A=3b=\frac{1}{2}P$

CIRCUMFERENCE OF INSCRIBED CIRCLE $\ll P$

AREA OF INSCRIBED CIRCLE $\ll \frac{1}{2}P$

POLYGON B





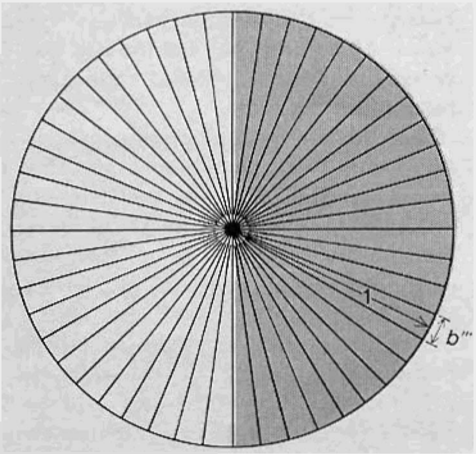
PERIMETER $P'=12b'$

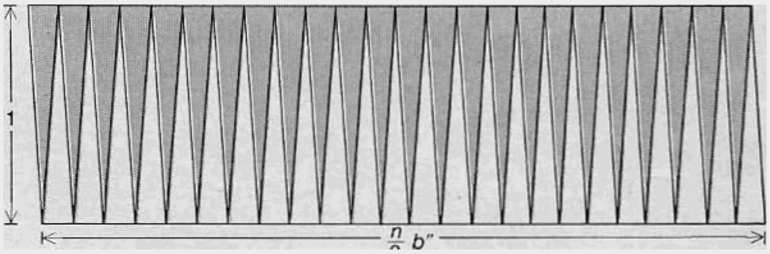
AREA OF POLYGON $B=6b'=\frac{1}{2}P'$

CIRCUMFERENCE OF INSCRIBED CIRCLE $< P'$

AREA OF INSCRIBED CIRCLE $< \frac{1}{2}P$

POLYGON N





PERIMETER $P''=nb''$

AREA OF POLYGON $N=\frac{nb''}{2}=\frac{1}{2}P''$

CIRCUMFERENCE OF INSCRIBED CIRCLE $\approx P''$

AREA OF INSCRIBED CIRCLE $\approx \frac{1}{2}P''$

not exist," he gave a "rigorous" proof of his results, using the "method of exhaustion," which relies on an indirect argument and purely finite constructions. The rigorous proof is given in his treatise *On the Quadrature of the Parabola,* which has been known since antiquity. The use of infinitesimals, which actually served to discover the answer, is in a paper called "On the Method," which was unknown until its sensational discovery in 1906.

Archimedes' method of exhaustion, which avoids infinitesimals, is in spirit close to the "epsilon-delta" method with which Weierstrass and his followers in the 19th century drove infinitesimal methods out of analysis. It is easy to explain if we refer to our example of the circle as an infinite-sided polygon. We wish to get a logically acceptable proof of the formula "The area of a circle with a radius of one unit equals half the circumference," which we discovered by a logically unacceptable argument.

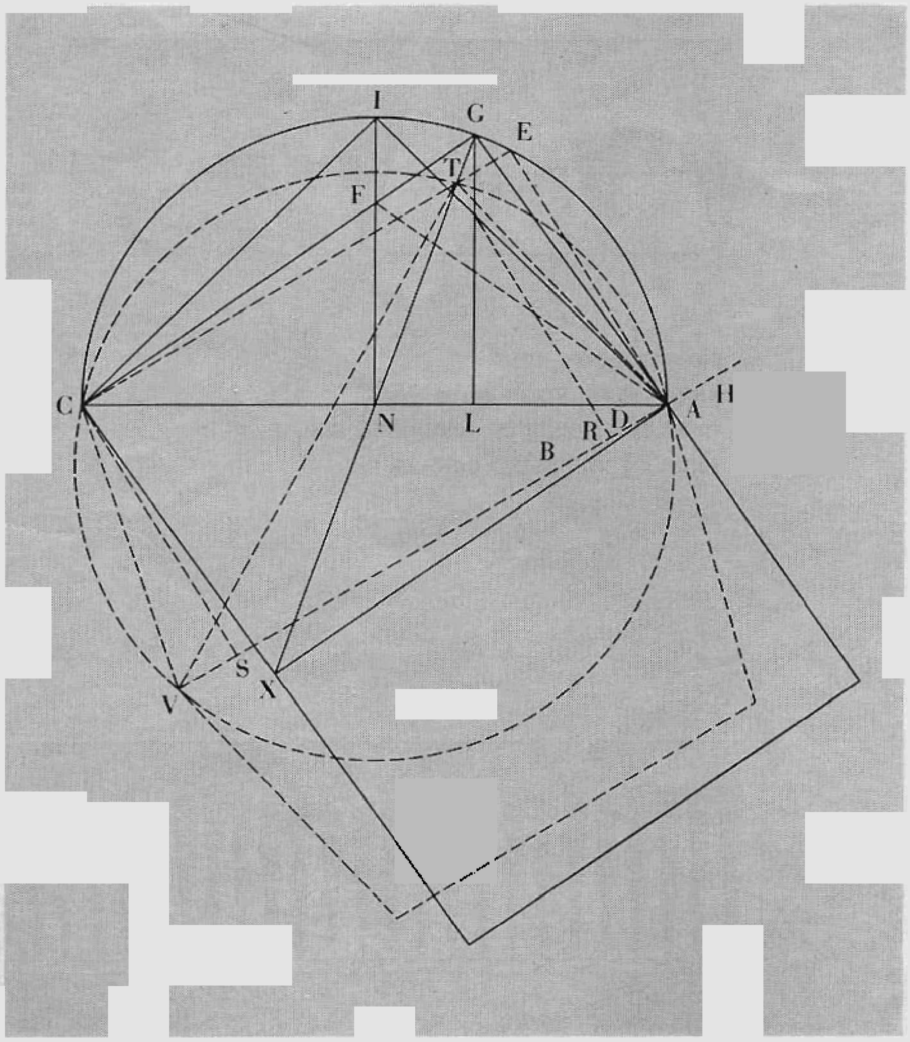
We reason as follows. The formula asserts the equality of two quantities associated with a circle with a radius of 1: its area and half its circumference. Thus if the formula is false, one of these quantities is larger than the other. Let A be the positive number obtained by subtracting the smaller from the larger. Now, we can circumscribe about the circle a regular polygon with as many sides as we wish. Since the polygon is composed of a finite number of finite triangles with altitude 1, we know that its area is half its perimeter. By making the number of sides sufficiently large we can arrange for the polygon's area to differ from the area of the circle by less than half of A (whatever its value is taken to be); at the same time the perimeter of the polygon will differ from the perimeter of the circle by less than half of A. But then the area and the semiperimeter of the circle must differ by less than A, which contradicts the supposition from which we started. Hence the supposition is impossible and A must be zero, as we wished to prove.

This argument is logically impeccable. Compared with the directness of the first analysis, however, there is something fussy, even pedantic, about it. After all, if the use of infinitesimals gives the right answer, must not the argument be correct in some sense? Even if we cannot justify the concepts it employs, how can it really be wrong if it works?

Such a defense of infinitesimals was not made by Archimedes. Indeed, in "On the Method" he is careful to explain that "the fact here stated is not actually demonstrated by the argument used" and that a rigorous proof had been published separately. On the other hand, Nicholas of Cusa, who was a cardinal of the church, preferred the reasoning by infinite quantities because of his belief that the infinite was "the source and means, and at the same time the unattainable goal, of all knowledge." Nicholas was followed in his mysticism by Johannes Kepler, one of the founders of modern science. In a work less well known nowadays than his discoveries in astronomy, Kepler in 1612 used infinitesimals to find the best proportions for a wine cask. He was not troubled by the self-contradictions in his method: he relied on divine inspiration, and he wrote that "nature teaches geometry by instinct alone, even without ratiocination." Moreover, his formulas for the volumes of wine casks are correct.

The most famous mathematical mystic was no doubt Blaise Pascal. In answering those of his contemporaries who objected to reasoning with infinitely small quantities, Pascal was fond of saying that the heart intervenes to make the work clear. Pascal looked on the infinitely large and the infinitely small as mysteries, something that nature has proposed to man not for him to understand but for him to admire.

The full flower of infinitesimal reasoning came with the generations after Pascal: Newton, Leibniz, the Bernoulli brothers (Jakob and Johann) and Leonhard Euler. The fundamental theorems of the calculus were found by Newton and Leibniz in the 1660's and 1670's. The first textbook on the calculus was written in 1696 by the Marquis de L'Hôpital, a pupil of Leibniz' and Johann Bernoulli's. Here it is stated at the



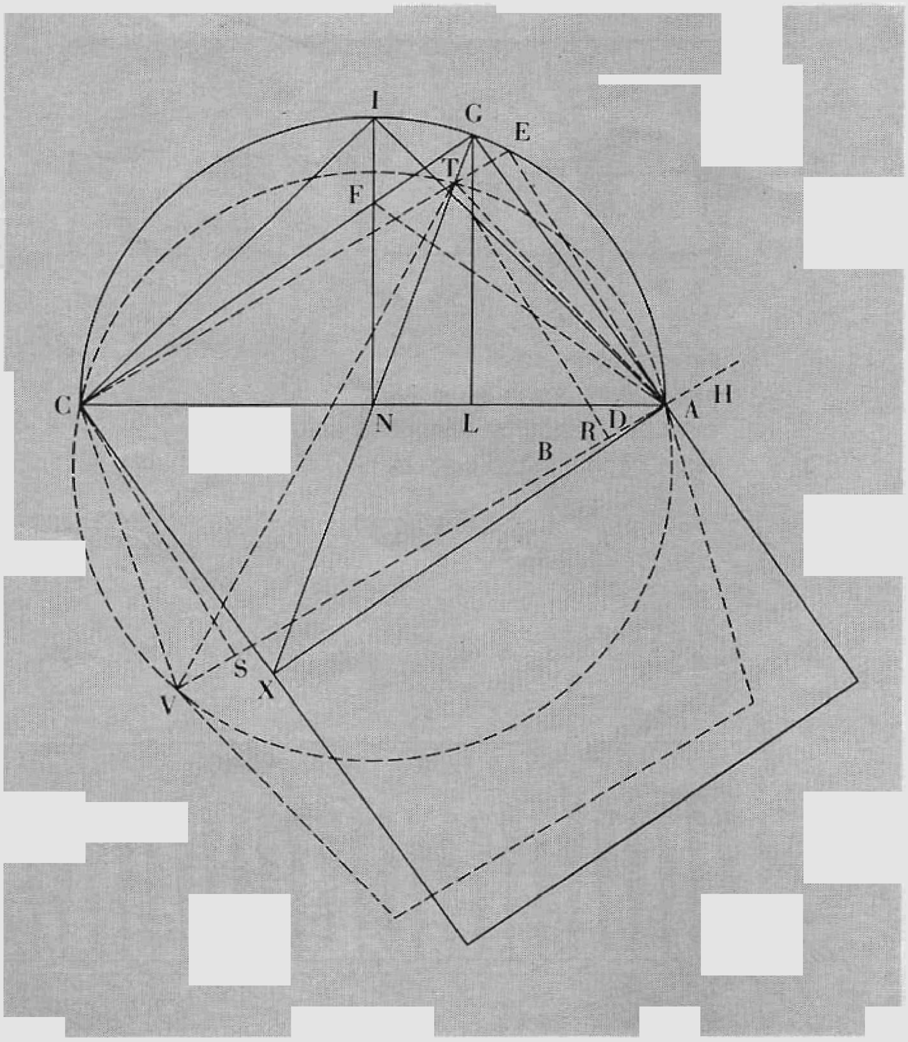
PROBLEM OF WINE CASK was attacked by Johannes Kepler by means of infinitesimals in his *Nova stereometria doliorum vinariorum,* published in Austria in 1615 and 1616. The problem Kepler set for himself was finding the best proportions for a wine cask. A page from a 19th-century reproduction of the paper that was published in Europe is shown.

not exist," he gave a "rigorous" proof of his results, using the "method of exhaustion," which relies on an indirect argument and purely finite constructions. The rigorous proof is given in his treatise *On the Quadrature of the Parabola,* which has been known since antiquity. The use of infinitesimals, which actually served to discover the answer, is in a paper called "On the Method," which was unknown until its sensational discovery in 1906.

Archimedes' method of exhaustion, which avoids infinitesimals, is in spirit close to the "epsilon-delta" method with which Weierstrass and his followers in the 19th century drove infinitesimal methods out of analysis. It is easy to explain if we refer to our example of the circle as an infinite-sided polygon. We wish to get a logically acceptable proof of the formula "The area of a circle with a radius of one unit equals half the circumference," which we discovered by a logically unacceptable argument.

We reason as follows. The formula asserts the equality of two quantities associated with a circle with a radius of 1: its area and half its circumference. Thus if the formula is false, one of these quantities is larger than the other. Let A be the positive number obtained by subtracting the smaller from the larger. Now, we can circumscribe about the circle a regular polygon with as many sides as we wish. Since the polygon is composed of a finite number of finite triangles with altitude 1, we know that its area is half its perimeter. By making the number of sides sufficiently large we can arrange for the polygon's area to differ from the area of the circle by less than half of A (whatever its value is taken to be); at the same time the perimeter of the polygon will differ from the perimeter of the circle by less than half of A. But then the area and the semiperimeter of the circle must differ by less than A, which contradicts the supposition from which we started. Hence the supposition is impossible and A must be zero, as we wished to prove.

This argument is logically impeccable. Compared with the directness of the first analysis, however, there is something fussy, even pedantic, about it. After all, if the use of infinitesimals gives the right answer, must not the argument be correct in some sense? Even if we cannot justify the concepts it employs, how can it really be wrong if it works?



PROBLEM OF WINE CASK was attacked by Johannes Kepler by means of infinitesimals in his *Nova stereometria doliorum vinariorum*, published in Austria in 1615 and 1616. The problem Kepler set for himself was finding the best proportions for a wine cask. A page from a 19th-century reproduction of the paper that was published in Europe is shown.

Such a defense of infinitesimals was not made by Archimedes. Indeed, in "On the Method" he is careful to explain that "the fact here stated is not actually demonstrated by the argument used" and that a rigorous proof had been published separately. On the other hand, Nicholas of Cusa, who was a cardinal of the church, preferred the reasoning by infinite quantities because of his belief that the infinite was "the source and means, and at the same time the unattainable goal, of all knowledge." Nicholas was followed in his mysticism by Johannes Kepler, one of the founders of modern science. In a work less well known nowadays than his discoveries in astronomy, Kepler in 1612 used infinitesimals to find the best proportions for a wine cask. He was not troubled by the self-contradictions in his method; he relied on divine inspiration, and he wrote that "nature teaches geometry by instinct alone, even without ratiocination." Moreover, his formulas for the volumes of wine casks are correct.

The most famous mathematical mystic was no doubt Blaise Pascal. In answering those of his contemporaries who objected to reasoning with infinitely small quantities, Pascal was fond of saying that the heart intervenes to make the work clear. Pascal looked on the infinitely large and the infinitely small as mysteries, something that nature has proposed to man not for him to understand but for him to admire.

The full flower of infinitesimal reasoning came with the generations after Pascal: Newton, Leibniz, the Bernoulli brothers (Jakob and Johann) and Leonhard Euler. The fundamental theorems of the calculus were found by Newton and Leibniz in the 1660's and 1670's. The first textbook on the calculus was written in 1696 by the Marquis de L'Hôpital, a pupil of Leibniz' and Johann Bernoulli's. Here it is stated at the outset as an axiom that two quantities differing by an infinitesimal can be considered to be equal. In other words, the quantities are at the same time considered to be equal to each other and not equal to each other! A second axiom states that a curve is "the totality of an infinity of straight segments, each infinitely small." This is an open embracing of methods that Aristotle had outlawed 2,000 years earlier.

Indeed, wrote L'Hôpital, "ordinary analysis deals only with finite quantities; this one penetrates as far as infinity itself. It compares the infinitely small differences of finite quantities; it discovers the relations between these differences, and in this way makes known the relations between finite quantities that are, as it were, infinite compared with the infinitely small quantities. One may even say that this analysis extends beyond infinity, for it does not confine itself to the infinitely small differences but discovers the relations between the differences of these differences."

Newton and Leibniz did not share L'Hôpital's enthusiasm. Leibniz did not claim that infinitesimals really existed, only that one could reason without error as if they did exist. Although Leibniz could not substantiate this claim, Robinson's work shows that in some sense he was right after all. Newton tried to avoid the infinitesimal. In his *Principia Mathematica*, as in Archimedes' *On the Quadrature of the Parabola*, results that were originally found by infinitesimal methods are presented in a purely finite Euclidean fashion.

WEIERSTRASS	FALLING STONE: POSITION $s = 16t^2$	ROBINSON
Set $t' = 1 + \Delta t \cdot \Delta t$ is a positive real number. $s' = 16 + 32\Delta t + 16(\Delta t)^2$. $\Delta s = s' - s$ $= 32\Delta t + 16(\Delta t)^2$. $\frac{\Delta s}{\Delta t} = 32 + 16\Delta t$. Given any positive real number ϵ, however small, we choose $\delta = \frac{\epsilon}{16}$. Then for all $\Delta t < \delta$, $\frac{\Delta s}{\Delta t} - 32 = 16\Delta t < 16\delta$ $= 16 \cdot \frac{\epsilon}{16} = \epsilon$. So Instantaneous velocity = $\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = 32$.		Set $t' = 1 + dt$. dt is a positive infinitesimal number. $s' = 16 + 32dt + 16(dt)^2$. $ds = s' - s$ $= 32dt + 16(dt)^2$. $\frac{ds}{dt} = 32 + 16dt$. Since dt is infinitesimal, so is $16dt$. 32 is a standard real number. So Instantaneous velocity = standard part of $\frac{ds}{dt} = 32$.

FALLING-STONE PROBLEM is depicted as it would be solved by standard analysis (*left*) and nonstandard analysis (*right*). Standard analysis, exemplified by the 19th-century German mathematician Karl Weierstrass, computes the velocity of the falling stone at any instant without employing infinitesimals, defining the speed instead as a limit that is approximated by ratios of finite increments. Abraham Robinson of Yale University, who invented nonstandard analysis, makes the computation with a modified infinitesimal method.

Dynamics had become as important as geometry in providing questions for mathematical analysis. The leading problem was the connection between "fluents" and "fluxions," what would today be called the instantaneous position and the instantaneous velocity of a moving body.

Consider a falling stone. Its motion is described by giving its position as a function of time. As it falls its velocity increases, so that the velocity at each instant is also a variable function of time. Newton called the position function the "fluent" and the velocity function the "fluxion." If either of the two is given, the other can be determined; this connection is the heart of the infinitesimal calculus fashioned by Newton and Leibniz.

In the case of a falling stone the fluent is given by the formula $s = 16t^2$, where s is the number of feet traveled and t is the number of seconds elapsed since the stone was released. As the stone falls its velocity increases steadily. How can we compute the velocity of the falling stone at some instant of time, say at $t = 1$?

We could find the *average* velocity for a finite time by the elementary formula: velocity equals distance divided by time. Can we use this formula to find the instantaneous velocity? In an infinitesimal increment of time the increment of distance would also be infinitesimal; their ratio, the average speed during the instant, should be the finite instantaneous velocity we seek.

We let dt stand for the infinitesimal increment of time and ds for the corresponding increment of distance. (Of course ds and dt must be thought of as single symbols and not as d times t or d times s.) We want to find the ratio ds/dt, which is to be finite. To find the increment of distance from $t = 1$ to $t = 1 + dt$ we compute the position of the stone when $t = 1$, which is $16 \times 1^2 = 16$, and its position when $t = 1 + dt$, which is $16 \times (1 + dt)^2$. Using a little elementary algebra, we find that ds, the increment of distance, which is the difference of these two distances, is $32dt + 16dt^2$. Thus the ratio ds/dt, which is the quantity we are trying to find, is equal to $32 + 16dt$.

Have we solved our problem? Since the answer should be a finite quantity, we should like to drop the infinitesimal term, $16dt$, and get the answer, 32 feet per second, for the instantaneous velocity. That is precisely what Bishop Berkeley will not let us do.

The Analyst, Berkeley's brilliant and devastating critique of the infinitesimal method, appeared in 1734. The book was addressed to "an infidel mathematician," who is generally supposed to have been Newton's friend the astronomer Edmund Halley. Halley financed the publication of the *Principia* and helped to prepare it for the press. It is said that he also persuaded a friend of Berkeley's of the "inconceivability of the doctrines of Christianity"; the Bishop responded that Newton's fluxions were as "obscure, repugnant and precarious" as any point in divinity.

"I shall claim the privilege of a Freethinker," wrote the Bishop, "and take the liberty to inquire into the object, principles, and method of demonstration admitted by the mathematicians of the present date, with the same freedom that you presume to treat the principles and mysteries of Religion." Berkeley declared that the Leibniz procedure, simply "considering" $32 + 16dt$ to be "the same" as 32, was unintelligible. "Nor will it avail," he wrote, "to say that [the term neglected] is a quantity exceedingly small; since we are told that *in rebus mathematicis errores quam minimi non sunt contemnendi.*" If something is neglected, no matter how small, we can no longer claim to have the exact velocity but only an approximation.

Newton, unlike Leibniz, tried in his later writings to soften the "harshness" of the doctrine of infinitesimals by using physically suggestive language. "By the ultimate velocity is meant that with which the body is moved, neither before it arrives at its last place, when the motion ceases, nor after; but at the very instant when it arrives.... And, in like manner, by the ultimate ratio of evanescent quantities is to be understood the ratio of the quantities, not before they vanish, nor after, but that with which they vanish." When he proceeded to compute, however, he still had to justify dropping unwanted "negligible" terms from his computed answer. Newton's argument was to find first, as we have done, $ds/dt = 32 + 16dt$, and then to set the increment dt equal to zero, leaving 32 as the exact answer.

But, wrote Berkeley, "it should seem that this reasoning is not fair or conclusive." After all, dt is either equal to zero or not equal to zero. If dt is not zero, then $32 + 16dt$ is not the same as 32. If dt is zero, then the increment in distance ds is also zero, and the fraction ds/dt is not $32 + 16dt$ but a meaningless expression, 0/0. "For when it is said, let the increments vanish, i.e., let the increments be nothing, or let there be no increments, the former supposition that the increments were something, or that there were increments, is destroyed, and yet a consequence of that supposition, i.e., an expression got by virtue thereof, is retained. Which is a false way of reasoning." Berkeley charitably concluded: "What are these fluxions? The velocities of evanescent increments. And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?"

Berkeley's logic could not be answered; nevertheless, mathematicians went on using infinitesimals for another century, and with great success. Indeed, physicists and engineers have never stopped using them. In pure mathematics, on the other hand, a return to Euclidean rigor was achieved in the 19th century, culminating under the leadership of Weierstrass in 1872. It is interesting to note that the 18th century, the great age of the infinitesimal, was the time when no barrier between mathematics and physics was recognized. The leading physicists and the leading mathematicians were the same people. When pure mathematics reappeared as a separate discipline, mathematicians again made sure that the foundations of their work contained no obvious contradictions. Modern analysis secured its foundations by doing what the Greeks had done: outlawing infinitesimals.

To find an instantaneous velocity according to the Weierstrass method we abandon any attempt to compute the speed as a ratio. Instead we define the speed as a limit, which is approximated by ratios of finite increments. Let Δt be a variable finite time increment and Δs be the corresponding variable space increment. Then $\Delta s/\Delta t$ is the variable quantity $32 + 16\Delta t$. By choosing Δt sufficiently small we can make $\Delta s/\Delta t$ take on values as close as we like to the value 32, and so, by definition, the speed at $t = 1$ is exactly 32.

This approach succeeds in removing any reference to numbers that are not finite. It also avoids any attempt directly to set Δt equal to zero in the fraction $\Delta s/\Delta t$. Thus we avoid both of the logical pitfalls exposed by Bishop Berkeley. We do, however, pay a price. The intuitively clear and physically measurable quantity, the instantaneous velocity, becomes subject to the surprisingly subtle notion of "limit." If we spell out in detail what that means, we have the following tongue-twister:

"The velocity is v if, for any positive number ε, $\Delta s/\Delta t - v$ is less than ε in absolute value for all values of Δt less in absolute value than some other positive number δ (which will depend on ε and t)."

We have defined v by means of a subtle relation between two new quantities, ε and δ, which in some sense are irrelevant to v itself. At least ignorance of ε and δ never prevented Bernoulli or Euler from finding a velocity. The truth is that in a real sense we already knew what instantaneous velocity was before we learned this definition; for the sake of logical consistency we accept a definition that is much harder to understand than the concept being defined. Of course, to a trained mathematician the epsilon-delta definition is intuitive; this shows what can be accomplished by proper training.

The reconstruction of the calculus on the basis of the limit concept and its epsilon-delta definition amounted to a reduction of the calculus to the arithmetic of real numbers. The momentum gathered by these foundational clarifications led naturally to an assault on the logical foundations of the real-number system itself. This was a return after two and a half millenniums to the problem of irrational numbers, which the Greeks had abandoned as hopeless after Pythagoras. One of the tools in these efforts was the newly developing field of mathematical, or symbolic, logic.

More recently it has been found that mathematical logic provides a conceptual foundation for the theory of computing machines and computer programs. Hence this prototype of purity in mathematics now has to be regarded as belonging to the applicable part of mathematics.

The link between logic and computing is to a great extent the notion of a formal language, which is the kind of language machines understand. And it is the notion of a formal language that enabled Robinson to make precise Leibniz' claim that one could without error reason as if infinitesimals existed.

Leibniz had thought of infinitesimals

SYMBOL	INTENDED MEANING
$\sim$	not
&	and
$\vee$	or
$\rightarrow$	implies
$\forall$	for all
$\exists$	there exists
=	equals
x y z	variables ranging over real numbers
f g h	variables ranging over other objects
+ · <	plus, times, less than
( ) [ ]	parentheses
0 1 2	symbols for particular numbers

SYMBOLS EMPLOYED in the formal language *L*, in which calculus can be expressed, are translated into English in this partial dictionary. The formal language, which employs many more symbols than these, provides a link between the standard universe and the larger nonstandard universe that is a central concept of nonstandard mathematical analysis.

as being infinitely small positive or negative numbers that still had "the same properties" as the ordinary numbers of mathematics. On its face the idea seems self-contradictory. If infinitesimals have the same "properties" as ordinary numbers, how can they have the "property" of being positive yet smaller than any ordinary positive number? It was by using a formal language that Robinson was able to resolve the paradox. Robinson showed how to construct a system containing infinitesimals that was identical with the system of "real" numbers with respect to all those properties expressible in a certain formal language. Naturally the "property" of being positive yet smaller than any ordinary positive number will turn out *not* to be expressible in the language, thereby escaping the paradox.

The situation is familiar to anyone who has ever communicated with a computing machine. A computer accepts as inputs only symbols from a certain list that is given in advance to the user, and the symbols must be used in accordance with certain given rules. Ordinary language, as used in human communication, is subject to rules that linguists are still far from understanding. Computers are "stupid," if you have to communicate with them, precisely because unlike humans they work in a formal language with a given vocabulary and a given set of rules. Humans work in a natural language, with rules that have never been made fully explicit.

Mathematics, of course, is a human activity, like philosophy or the design of computers; like these other activities, it is carried on by humans using natural languages. At the same time mathematics has, as a special feature, the ability to be well described by a formal language, which in some sense mirrors its content precisely. It might be said that the possibility of putting a mathematical discovery into a formal language is the test of whether it is fully understood.

In nonstandard analysis one takes as the starting point the finite real numbers and the rest of the calculus as known to standard mathematicians. Call this the "standard universe," designated by the letter *M*. The formal language in which we talk about *M* can be designated *L*. Any sentence in *L* is a proposition about *M*, and of course it must be either true or false. That is, any sentence in *L* is either true or its negation is true. We call the set of all true sentences *K*, and we say *M* is a "model" for *K*. By this we mean that *M* is a mathematical structure such that every sentence in *K*, when interpreted as referring to *M*, is true. Of course, we do not "know" *K* in any effective sense; if we did, we would have the answer to every possible question in analysis. Nevertheless, we regard *K* as being a well-defined object, about which we can reason and draw conclusions.

FORMAL SENTENCE IN L	INTERPRETATION IN STANDARD UNIVERSE	INTERPRETATION IN NONSTANDARD UNIVERSE
$(\forall x)(\exists y)[x=0 \vee xy=1]$ Literally: For all x, there exists y such that either $x=0$ or $xy=1$.	Every nonzero real number has a reciprocal.	Every nonzero nonstandard real number has a nonstandard reciprocal; in particular positive infinitesimals have reciprocals that are larger than any standard, real number, i.e., they are infinite.

FORMAL SENTENCE is stated in the language *L*. The middle column gives its interpretation or meaning in the standard universe; right-hand column, in the nonstandard universe.

The essential fact, the main point, is that in addition to *M*, the standard universe, there are also nonstandard models for *K*. That is, there are mathematical structures M^*, essentially different from *M* (in a sense we shall explain) and that nevertheless are models for *K* in the natural sense of the term: there are objects in M^* and relations between objects in M^* such that if the symbols in *L* are reinterpreted to apply to these pseudo-objects and pseudo-relations in the appropriate way, then every sentence in *K* is still true, although with a different meaning.

A crude analogy may help the intuition. Let *M* be the set of graduating seniors at Central High School. Suppose, for argument's sake, that all these students had their picture taken for the yearbook, where the students all appear in two-inch squares. Then M^* can be the set of all two-inch squares on any page of the yearbook. Clearly, with an obvious interpretation, any true statement about a student at Central High corresponds to a true statement about a certain two-inch square in the yearbook. Still, there are many two-inch squares in the yearbook that do not correspond to any student. M^* is much bigger than *M;* in addition to members corresponding to the members of *M*, it also contains many other members.

Hence the statement "Harry Smith is thinner than George Klein," when interpreted in M^*, is a statement about certain two-inch squares. It is not true if the relation "thinner than" is interpreted in the standard way. Thus "thinner than" has to be reinterpreted, as a pseudo-relation, between pseudo-students (pictures of students). We could define the pseudo-relation "thinner than" (in quotation marks) by saying that the two-inch square labeled "Harry Smith" is "thinner than" the two-inch square labeled "George Klein" only if Harry Smith is actually thinner than George Klein. In this way true statements about students are reinterpreted as true statements about two-inch squares.

Of course, in this example the entire argument is a bit contrived. If *M* is the standard universe for the calculus, however, then M^*, the nonstandard universe, is a remarkable and interesting place.

The existence of interesting nonstandard models was first discovered by the Norwegian logician Thoralf A. Skolem, who found that the axioms of counting—the axioms that describe the "natural numbers" 1, 2, 3 and so on—have nonstandard models containing "strange" objects not contemplated in ordinary arithmetic. Robinson's great insight was to see how this exotic offshoot of modern formal logic could be the basis for resurrecting infinitesimal methods in differential and integral calculus. In this resurrection he relied on a theorem first proved by the Russian logician Anatoli Malcev and then generalized by Leon A. Henkin of the University of California at Berkeley. This is the "compactness" theorem. It is related to the famous "completeness" theorem of Kurt Gödel, which states that a set of sentences is logically consistent (no contradiction can be deduced from the sentences) if and only if the sentences have a model, that is, if and only if there is a "universe" in which they are all true.

The compactness theorem states the following: Suppose we have a collection of sentences in the language *L*. Suppose in the standard universe every finite subset of this collection is true. Then there exists a nonstandard universe where the entire collection is true at once.

The compactness theorem follows easily from the completeness theorem: if every finite subset of a collection of sentences of L is true in the standard universe, then every finite subset is logically consistent. So the entire collection of sentences is logically consistent (since any deduction can make use of only a finite number of premises). By the completeness theorem there is a (nonstandard) universe in which the entire collection is true.

A direct consequence of the compactness theorem is the "existence" of infinitesimals. To see how this amazing result follows from the compactness theorem consider the sentences:

"C is a number bigger than zero and less than $1/2$."

"C is a number bigger than zero and less than $1/3$."

"C is a number bigger than zero and less than $1/4$." And so on.

This is an infinite collection of sentences each of which can be written in the formal language L. With reference to the standard universe R of real numbers, every finite subset is true, because if you have finitely many sentences of the form "C is a number bigger than zero and less than $1/n$," then one of the sentences will contain the smallest fraction $1/n$, and $1/2n$ will indeed be bigger than zero and smaller than all the fractions in your finite list of sentences. And yet if you consider the entire infinite set of these sentences, it is false with reference to the standard real numbers, because no matter how small a positive real number c you choose, $1/n$ will be smaller than c if n is big enough.

The compactness theorem of Malcev and Henkin states that there is a nonstandard universe containing pseudoreals R^* including a positive pseudoreal number c smaller than any number of the form $1/n$. That is, c is infinitesimal. Moreover, c has all the properties of standard real numbers in a perfectly precise sense: any true statement about the standard reals that you can state in the formal language L is true also about the nonstandard reals, including the infinitesimal c—under the appropriate interpretation. (The two-inch square labeled "Harry Smith" is not really thinner than the two-inch square labeled "George Klein," but the statement "Harry Smith" is "thinner than" "George Klein" is true, under our nonstandard interpretation of "thinner than." On the other hand, properties shared by all the standard real numbers may not apply to the nonstandard pseudo-numbers, if these properties cannot be expressed in the formal language L.

The Archimedean property (nonexistence of infinitesimals) of R can be expressed by using an infinite set of sentences of L as follows (we use the symbol "$>$" as usual to mean "is greater than"). For each positive element c of R all but a finite number of the sentences below are true:

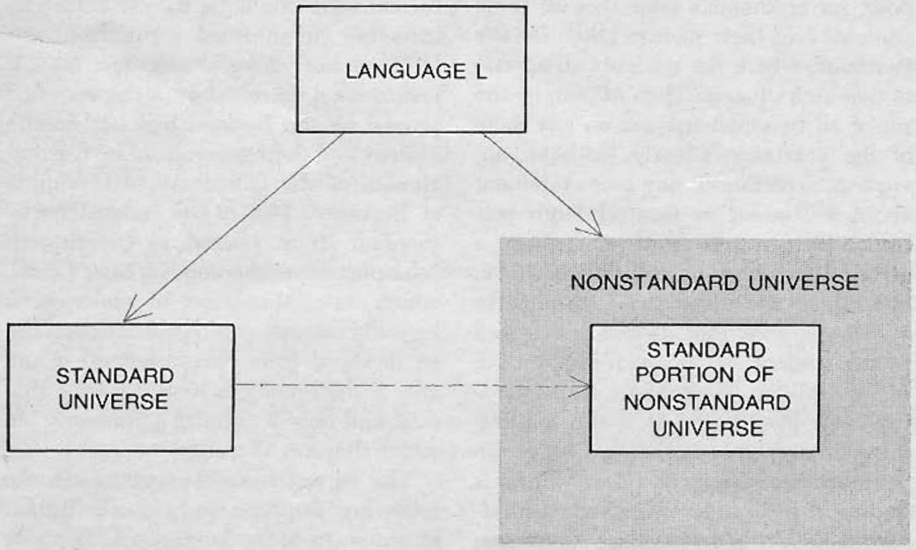
$$c > 1$$
$$c + c > 1$$
$$c + c + c > 1, \text{ and so on.}$$

This is not true, however, for the pseudo-reals R^*: if c is infinitesimal (hence pseudo-real), all these sentences are false. In other words, no finite sum of c's can exceed 1, no matter how many terms we take. The very fact that the Archimedean property is true in the standard world but false in the nonstandard one proves that the property cannot be expressed by a sentence of L; the statement we have used involves infinitely many sentences. It is precisely this distinction that makes the pseudo-objects useful. They behave "formally" like standard objects and yet they differ with respect to important properties that are not formalized by L.

Although the nonstandard universe is conceptually distinct from the standard one, it is desirable to think of it as an enlargement of the standard universe. Since R^* is a model for L, every true sentence about R has an interpretation in R^*. In particular the names of numbers in R have an interpretation as names of objects in R^*. We can simply identify the object in R^* called "2" with the familiar number 2 in R. Then R^* contains the standard real numbers in R, along with a vast collection of infinitesimal and infinite quantities, in which R is embedded.

An object in R^* (a pseudo-real number) is called infinite if it is pseudo-greater than every standard real number; otherwise it is called finite. A positive pseudo-real number is called infinitesimal if it is pseudo-smaller than every positive standard real number. If the pseudo-difference of two pseudo-reals is finite, we say they belong to the same "galaxy"; the pseudo-real axis contains an uncountable infinity of galaxies. If the pseudo-difference of two pseudo-reals is infinitesimal, we say they belong to the same "monad" (a term Robinson borrowed from Leibniz' philosophical writings). If a pseudo-real r^* is infinitely close to a standard real number r, we say r is the standard part of r^*. All the standard reals are of course in the same galaxy, which is called the principal galaxy. In the principal galaxy every monad contains one and only one standard real number. This monad is the "infinitesimal neighborhood" of r: the set of nonstandard reals infinitely close to r. The notion of a monad turns out to be applicable not only to real numbers but also to general metric and topological spaces. Nonstandard analysis therefore is relevant not just to elementary calculus but to the entire range of modern abstract analysis.

When we say infinitesimals or monads exist, it should be clear that we do not



ROLE OF FORMAL LANGUAGE in mediating between standard and nonstandard universes is portrayed. Formal language L describes the standard universe, which includes the real numbers of classical mathematics. Sentences of L that are true in standard universe are also true in nonstandard one, which contains additional mathematical objects such as infinitesimals. Nonstandard analysis thus makes the infinitesimal method precise for first time.

mean this at all in the sense it would have been understood by Euclid or by Berkeley. Until 100 years ago it was tacitly assumed by all philosophers and mathematicians that the subject matter of mathematics was objectively real in a sense close to the sense in which the subject matter of physics is real. Whether infinitesimals did or did not exist was a question of fact, not too different from the question of whether material atoms do or do not exist. Today many, perhaps most, mathematicians have no such conviction of the objective existence of the objects they study. Model theory entails no commitment one way or the other on such ontological questions. What mathematicians want from infinitesimals is not material existence but rather the right to use them in proofs. For this all one needs is the assurance that a proof using infinitesimals is no worse than one free of infinitesimals.

The employment of nonstandard analysis in research goes something like this. One wishes to prove a theorem involving only standard objects. If one embeds the standard objects in the nonstandard enlargement, one may be able to find a much shorter and more "insightful" proof by using nonstandard objects. The theorem has then been proved actually with reference to the nonstandard interpretation of its words and symbols. Those nonstandard objects that correspond to standard objects have the feature that sentences about them are true (in the nonstandard interpretation) only if the same sentence is true with reference to the standard object (in the standard interpretation). Thus we prove theorems about standard objects by reasoning about nonstandard objects.

For example, recall Nicholas of Cusa's "proof" that the area of a circle with a radius of 1 equals half its circumference. In Robinson's theory we see in what sense Nicholas' argument is correct. Once infinitesimal and infinite numbers are available (in the nonstandard universe) it can be proved that the area of the circle is the standard part of the sum (in the nonstandard universe) of infinitely many infinitesimals.

Here is how the falling-stone problem would look according to Robinson. We define the instantaneous velocity not as the ratio of infinitesimal increments, as L'Hôpital did, but rather as the standard part of that ratio; then ds, dt and their ratio ds/dt are nonstandard real numbers. We have as before $ds/dt = 32 + 16dt$, but now we immediately conclude, rigorously and without any limiting argument, that v, the standard part of ds/dt, equals 32. A slight modification in the Leibniz method of infinitesimals, distinguishing carefully between the nonstandard number ds/dt and its standard part v, avoids the contradiction, which L'Hôpital simply ignored.

Of course, a proof is required that the Robinson definition gives the same answer in general as the Weierstrass definition. The proof is not difficult, but we shall not attempt to give it here.

What is achieved is that the infinitesimal method is for the first time made precise. In the past mathematicians had to make a choice. If they used infinitesimals, they had to rely on experience and intuition to reason correctly. "Just go on," Jean Le Rond d'Alembert is supposed to have assured a hesitating mathematical friend, "and faith will soon return." For rigorous certainty one had to resort to the cumbersome Archimedean method of exhaustion or its modern version, the Weierstrass epsilon-delta method. Now the method of infinitesimals, or more generally the method of monads, is elevated from the heuristic to the rigorous level. The approach of formal logic succeeds by totally evading the question that excited Berkeley and all the other controversialists of former times, that is, whether or not infinitesimal quantities really exist in some objective sense.

From the viewpoint of the working mathematician the important thing is that he regains certain methods of proof, certain lines of reasoning, that have been fruitful since before Archimedes. The notion of an infinitesimal neighborhood is no longer a self-contradictory figure of speech but a precisely defined concept, as legitimate as any other in analysis.

The applications we have discussed are elementary, in fact trivial. Nontrivial applications have been and are being made. Work has appeared on nonstandard dynamics and nonstandard probability. Robinson and his pupil Allen Bernstein used nonstandard analysis to solve a previously unsolved problem on compact linear operators. It must nonetheless be said that many analysts remain skeptical about the ultimate importance of Robinson's method. It is quite true that whatever can be done with infinitesimals can in principle be done without them. Perhaps, as with other radical innovations, the full use of the new ideas will be made by a new generation of mathematicians who are not too deeply embedded in standard methods to enjoy the freedom and power of nonstandard analysis.