from *The Mathematical Experience,* by Philip J. Davis & Reuben Hersh. (1981), Houghton-Mifflin Company, Boston, MA.

Nonstandard Analysis

NONSTANDARD ANALYSIS, a new branch of mathematics invented by the logician Abraham Robinson, marks a new stage of development in several famous and ancient paradoxes. Robinson 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 nineteenth 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 eighteenth-century mathematics against the strait-laced rigor of the nineteenth 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 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 fifteenth 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.

Archimedes
c. 287 B.C.*–212* B.C.

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 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 nineteenth 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 1660s and 1670s. The first textbook on the calculus was written in 1696 by the Marquis de L'Hospital, a pupil of Leibniz and Johann Bernoulli. 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'Hospital, "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

Jakob Bernoulli
1654–1705

Sir Isaac Newton
1642–1727

Johann Bernoulli
1667–1748

Gottfried Wilhelm Leibniz
1646–1716

Blaise Pascal
1623–1662

G. F. A. de L'Hospital
1661–1704

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'Hospital'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.

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 the 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 Free-thinker," 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 in 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 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 nineteenth century, culminating under the leadership of Weierstrass in 1872. It is interesting to note that the eighteenth 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 formal language that enabled Robinson to make preci. Leibniz' claim that one could without error reason as if infinitesimals existed.

Leibniz had thought of infinitesimals 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.

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 structu 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 relati "thinner than" is interpreted in the standard way. Th "thinner than" has to be reinterpreted, as a pseudorelation, between pseudostudents (pictures of students). We could define the pseudorelation "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. Sko-