Innovation in Mathematics

The first of four articles on innovation in four central fields. The mathematician seeks a new logical relationship, a new proof of an old relationship, or a new synthesis of many relationships

by Paul R. Halmos

Everybody knows that for the past 300 years innovations in science and technology have come at a steadily increasing pace. Practically everyone is aware that mathematics has played a central role in this advance. Yet, strangely, many people think of mathematics itself as a static art—a body of eternal truth that was discovered by a few ancient, shadowy figures, and upon which engineers and scientists can draw as needed.

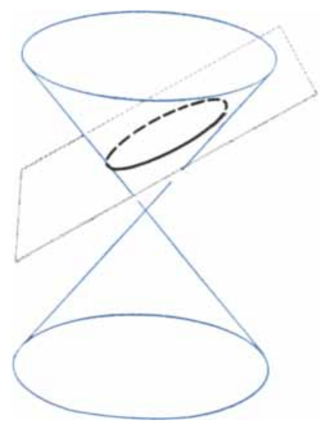
Of course nothing could be further from the truth. Mathematics is improving, changing and growing every day. On its growth depends not only progress in all the other fundamental investigations, but also progress in the crudest, bread-and-butter circumstances of our daily lives.

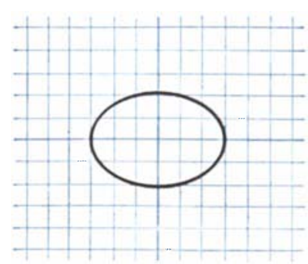
The late John von Neumann liked to cite this example of the relation between technological development and pure mathematics: A hundred and fifty years ago one of the most important problems of applied science—on which development in industry, commerce and government depended—was the problem of saving lives at sea. The statistics of the losses were frightful. The money and effort expended to solve the problem were frightful too—and sometimes ludicrous. No gadget, however complicated, was too ridiculous to consider—ocean-going passenger vessels fitted out like outrigger canoes may have looked funny, but they were worth a try. While leaders of government and industry were desperately encouraging such crank experiments, mathematicians were developing a tool that was to save more lives than all the crackpot inventors combined dared hope. That tool is what has come to be known as the theory of functions of a complex variable (a variable containing the "imaginary" number i, the square root of minus one). Among the many applications of this purely mathematical notion, one of the most fruitful is in the theory of radio communication. From the mathematician Karl Friedrich Gauss to the inventor Guglielmo Marconi it is only a few steps that almost any pair of geniuses such as James Clerk Maxwell and Heinrich Hertz can take in their stride.

The list of mathematical innovations could be continued almost endlessly. Here are just a few more. The theory of groups, for instance, was developed about 100 years ago. It probably would have seemed an ugly and useless invention to the contemporaries of Gauss. Today it is part of the mathematical repertory of every physicist. As recently as 50 years ago there was not a single professor of statistics in the U. S. Now statistical methods are an imperative tool in such sciences as genetics and experimental psychology. Von Neumann's theory of games, first published in 1928 and revived 20 years later, seems to be finding important applications in economics and operations research. Finally, lest anyone suppose that mathematical ideas spring full-blown and perfect from the brows of their creators, we may recall that Euclid's celebrated geometric proofs were found, after 2,000 years, to contain serious gaps. The holes in his reasoning were finally plugged by the great German mathematician David Hilbert, around the turn of the century.

Admitting, then, that there are innovations in mathematics, let us try to see just what they consist of, and, insofar as it is possible, how they come about. One way to classify a mathematical contribution is this: It may be a new proof of an old fact, it may be a new fact, or it may be a new approach to several facts at the same time.

A large part of the activity of professional mathematicians is a search for new proofs of old facts. One reason for this is pure pleasure: there is esthetic enjoyment in getting a fresh point of view on a familiar landmark. Another is that the original creator hardly ever





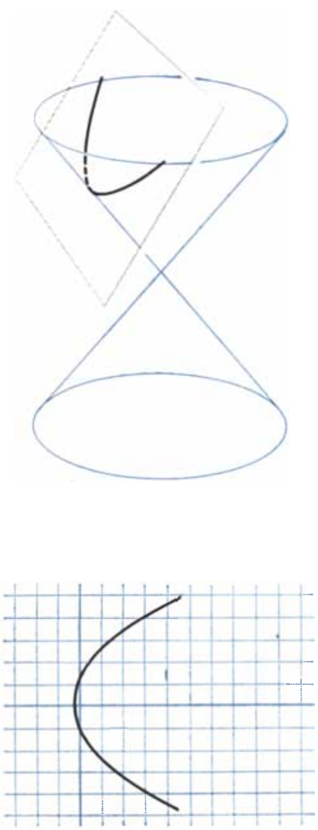
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

GEOMETRY AND ALGEBRA were connected by Descartes's invention of analytic geometry. The first three figures, from left reached his goal by the shortest, neatest, most efficient route, nor fully appreciated the connections between his brain-child and all other fields of mathematics. This is connected with a third and very practical motive. Mathematics has grown so luxuriantly in the past 2,000 years that it must be continually polished, simplified, systematized, unified and condensed. Otherwise the problem of handing the torch to each new generation would become completely unmanageable. No man alive today can know, even sketchily, all the mathematics published in the last 10 years. In order to give workers in the field enough understanding so that they can move ahead intelligently, it is absolutely imperative to find ever newer, shorter and simpler proofs that at the same time are more illuminating and provide more insight than their predecessors.

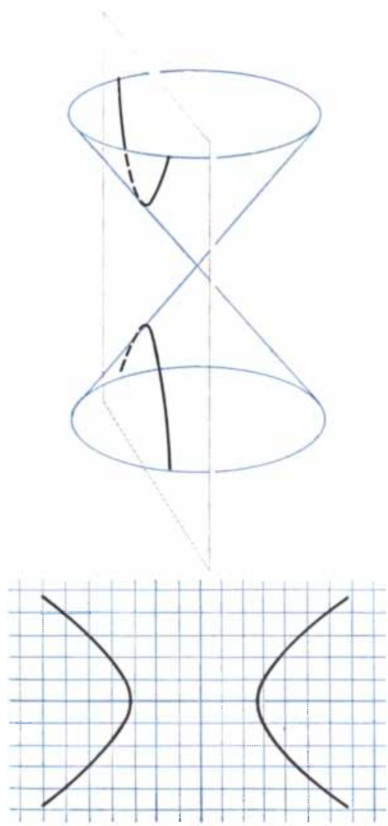
Curiously, it is sometimes good to find a new proof more complicated than an old one. If the new proof establishes some previously unsuspected connections between two ideas, it often leads to a generalization that makes the task of future learners far easier than it was for their teachers. René Descartes's coordinate, or "analytic," geometry is a good example. One consequence of Descartes's innovation is that it is possible to prove every proposition in Euclid's geometry by algebraic means [*see illustrations on page 70*]. The virtues of analytic geometry are many and they are great, but the simplicity of analytic proofs, when compared with Euclidean ones, is definitely not one of them. In most cases the analytic proof of a Euclidean fact about triangles or circles is a messy calculation that teaches us nothing.

The value of analytic geometry is that it reveals a connection between two branches of mathematics—algebra and geometry—which had been thought to be entirely separate. One of the main concerns of the early geometers was conic sections, the curves that are formed by cutting a cone with a plane [*see illustrations below*]. This is obviously a purely spatial way of thinking about the figures. When the conic sections (ellipse, parabola and hyperbola) are plotted on Cartesian coordinates and their algebraic equations are written down, it turns out that all the equations contain the squares of x and y, but no higher powers. Here is a new fact that provides deeper insight into the nature of the curves. And it also suggests a new question: What about the geometric picture of curves that do involve higher powers of x and y? Now we are led to consider a whole new class of geometric figures to which spatial intuition alone would never have led us. Furthermore, by picturing the equations geometrically, we gain new insight into their algebraic structure.

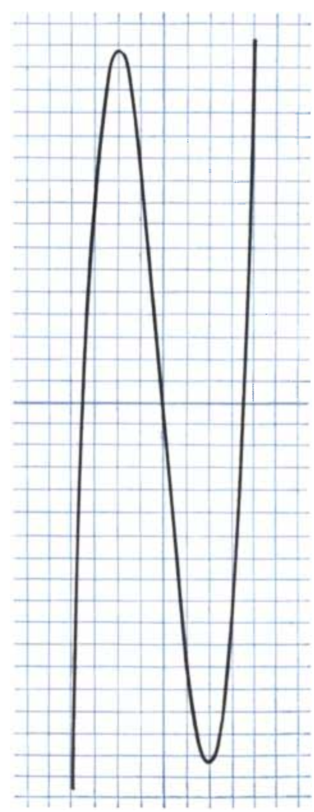
Of course most new proofs represent a gain in simplicity as well as in insight. Consider, for example, the problem of finding the area of the plane region bounded by the parabolic trajectory of a projectile and the line segment joining the gun and the target. This is a problem



$y^2 = 2px$

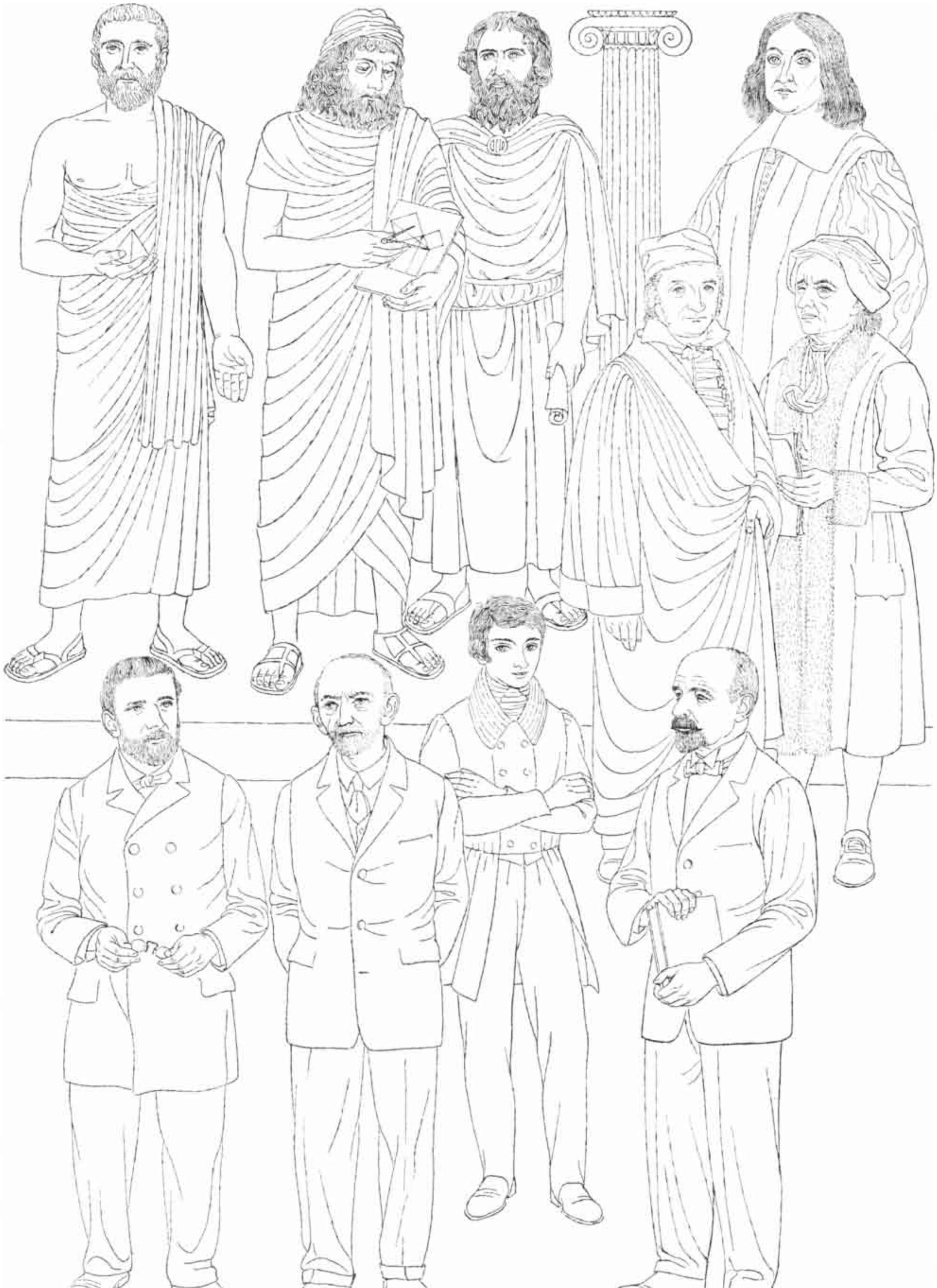


$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$



$y = ax^3 + bx$

to right, show at top the curves (ellipse, parabola and hyperbola) that can be cut from a cone by a plane. At bottom the same curves are represented as graphical pictures of algebraic equations. These equations turn out to contain the squares of the variables, but no higher powers of the variables. The figure at the right on this page is the graph of an equation in which one of the variables is cubed.



that Archimedes could and did solve; his solution depends on the celebrated "method of exhaustion," a way of finding complicated areas by adding together more and more simple ones. Archimedes' solution is both ingenious and long. The problem can also be solved, in one line of writing, by any mediocre sophomore calculus student. To be sure, the sophomore's efficient solution is the product of the profound thought of many mathematicians over many years.

Individual proofs do become shorter, but only because they become embedded in larger contexts from which it is easy to pluck them. The larger contexts spread and mix with other general concepts, forming a still-larger unified whole. After a couple of centuries 10 of the greatest discoveries of the era are likely to find themselves together between the covers of a slim volume in the pocket of a graduate student who, with luck, will absorb them all in two or three months.

So much for new proofs. What about new facts? In a trivial sense we have all discovered new mathematical facts. We see one every time we add a column of figures on our tax return. The chances are that no one has ever before observed that the sum of just those figures is what it is. A really interesting new mathematical fact has much more breadth and generality. Here is an example that is not exactly brand new (it was proved by Leonhard Euler some 200 years ago), but might be new to nonmathematicians: Every positive whole number is the sum of not more than four squares. The squares are of course 1, 4, 9, 16 and so on. Between them there are larger and larger gaps [*see illustration at top left on pages 72 and 73*]. If we add the squares two at a time (repetitions such as $4 + 4$ are allowed), we get a sequence with fewer gaps. If we fill in all the numbers that are sums of three squares, there will be still fewer gaps. Euler's theorem says that if we fill in all the numbers that are sums of four squares, there will be no more gaps left.

Another example of a mathematical fact, also no longer new, but which illustrates the powerful role that mathematical innovation can play, is the theory of the solvability of equations. It was created by the young French genius Evariste Galois in the early part of the 19th century. Galois's predecessors had found general formulas for solving equations up to the fourth degree—that is, equations in which the unknown is raised to no higher power than four. (One is the familiar "quadratic formula" of high-school algebra, which solves equations of the second degree.) Naturally they expected that there were also formulas that would solve equations of higher degree, and they spent an enormous amount of time and effort looking for them.

Galois dared to doubt the existence of such generalized formulas. He attacked the problem from a fresh point of view, looking not for tricks that would yield the supposedly hidden formulas, but trying to find more general properties of equations and their solutions. His work led to the important and fruitful concept of groups—sets of entities for which an operation similar to multiplication is defined. An immediate result of his beautiful and deep intellectual construction was to confirm his doubt: there are no general formulas for the solution of all algebraic equations. Thus he discovered a new fact.

The work of Galois exemplifies the third type of innovation also: the new approach. Group theory has proved an invaluable tool for attacking an extremely wide range of mathematical problems. Furthermore, Galois's ideas turned up another surprising connection between algebra and geometry. They show that the famous ancient puzzles of squaring the circle, duplicating the cube and trisecting the angle also have no general solutions. The reverberations of Galois's new approach can still be felt in modern algebra.

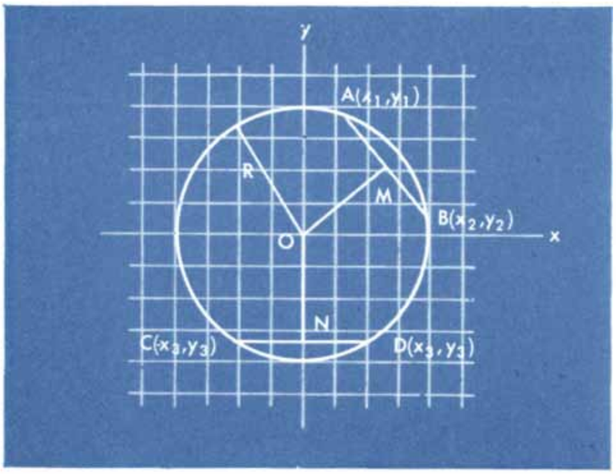
Where does a mathematical innovation come from? Sometimes, but by no means always, the source lies outside mathematics. Just as mathematics can make contributions to engineering, physics, psychology, genetics, economics and other disciplines, these other disciplines can keep mathematical creativity alive by asking stimulating questions, pointing to fresh lines of development, and, at the very least, providing suggestive language for the expression of mathematical ideas. It has happened that when a physicist needed a mathematical theory, it was already sitting there, waiting to be picked up and used. More often it happens that when something new is needed for such applications, the news percolates up to the ivory tower in a few decades (or more), and the answer comes down after a comparable time interval.

New mathematics often comes from plain curiosity. The right kind of mathematical curiosity is a precious possession that usually belongs only to professionals of the highest rank. The hardest problem of a young mathematician is to find a problem. The right question, well asked, is more than half the battle, and often the only part that requires inspiration. The answer itself may be difficult, and it may require ingenuity in the use of known techniques, but it often happens that all the thrill of creation and insight is concentrated in the question.

It should perhaps be mentioned that after the question is formulated, the mathematician does not proceed (as it is often supposed) like a scientific Sherlock Holmes. A mathematician is not a deduction machine, but a human being. New mathematics comes to him not by pure thought and deduction, but by sweat, experiment, induction and, if he is lucky, inspiration. Of course a mathematical experiment does not involve wires, tubes and bubbling liquids; it consists, rather, of a detailed examination of some particular cases or analogues of the desired result. (Example: Write down the first 10 squares, and write down systematically all the numbers that can be obtained as sums of two, or three, or four of them.) On the basis of such experiments the mathematician jumps inductively to bold conclusions. It may then be a difficult task to prove them, but often the purely deductive arrangement of the work serves more to communicate facts than to establish them.

To return to the source of innovations, it should be said that mathematicians share their curiosity, as well as their knowledge, with one another. At the beginning of his career a student often battens off the curiosity of his teachers. Great men do essentially the same thing when they decide to attack a problem that their predecessors could not solve. Galois himself was solving problems that he did not create. There will always be unsolved problems from bygone days: two famous ones that are not quite dead are the four-color map problem and Fermat's last theorem

SOME INNOVATORS whose mathematical contributions are discussed in this article are depicted on the opposite page. In the top row, from left to right, are Pythagoras, Euclid, Archimedes and Pierre de Fermat. In front of Fermat are Karl Friedrich Gauss and Leonhard Euler. The figures in the bottom row are Henri Poincaré, David Hilbert, Evariste Galois and Georg Cantor.



Given the curve $x^2+y^2=R^2$; AB = CD

To prove OM = ON

M is midpoint of AB and its coordinates are

$$\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$$

$$OM = \sqrt{\left(\frac{x_1+x_2}{2}\right)^2+\left(\frac{y_1+y_2}{2}\right)^2}$$

Expanding and substituting R^2 for $x_1^2+y_1^2$ and $x_2^2+y_2^2$

$$OM = \sqrt{\frac{R^2+x_1x_2+y_1y_2}{2}}$$

But ON = y_3, so we must prove

$$y_3 = \sqrt{\frac{R^2+x_1x_2+y_1y_2}{2}}$$

$$AB = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

Expanding and substituting as above

$$AB = \sqrt{2R^2-2x_1x_2-2y_1y_2}$$

CD = $2x_3$, and since AB = CD,

$$\sqrt{2R^2-2x_1x_2-2y_1y_2} = 2x_3$$

$$2R^2-2x_1x_2-2y_1y_2 = 4x_3^2$$

$$x_1x_2+y_1y_2 = R^2-2x_3^2$$

Substituting in the expression for OM,

$$OM = \sqrt{\frac{R^2+R^2-2x_3^2}{2}}$$

$$OM = \sqrt{R^2-x_3^2}$$

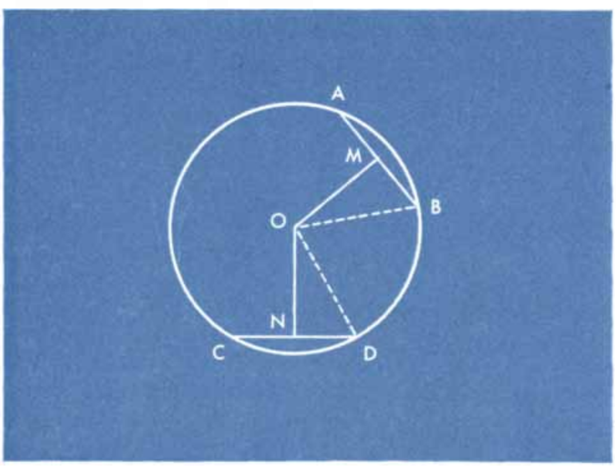
But $x_3^2+y_3^2=R^2$

$\therefore R^2-x_3^2=y_3^2$

$\therefore OM = \sqrt{y_3^2} = y_3$

$\therefore$ OM = ON

ANALYTIC PROOF of a Euclidean theorem is outlined above. The theorem states that equal chords of a circle are equidistant from the center. Proving the statement requires no ingenuity, but the algebraic computations are long, tedious and unilluminating.



Draw OB and OD

OB = OD	Radii of the same ⊙ are equal
MB = ½ AB and ND = ½ CD	A perpendicular from the center of a circle to a chord bisects the chord
$\therefore$ MB = ND	Halves of equals are equal
$\therefore \triangle OND \cong \triangle OMB$	Hypotenuse and leg
$\therefore$ OM = ON	

Q.E.D.

EUCLIDEAN PROOF of the proposition that equal chords are equidistant from the center is much shorter and neater than the analytic proof. Analytic geometry is valuable not as a method of grinding out proofs, but as a link between algebra and geometry.

(more properly, since it has not been proved, Fermat's last conjecture). Mathematicians of this century have been singularly fortunate in having a ready-made and inspiring list of problems to work on. It consists of 23 searching questions put together by David Hilbert and presented at the International Congress of Mathematicians in Paris in 1900. Several of the problems have been solved (and have made the reputations of their solvers); many of them are still open.

Practical application, curiosity and history are the main wellsprings of innovation, but two others should also be mentioned. One is failure; the other is error. Everyone who has so far tried to prove Fermat's conjecture has failed, but some of the efforts have produced the most fruitful concepts in modern algebra and in number theory. As for error, whole books full of brilliant mathematics have been inspired by it. An oversight or a misstatement by mathematician X is often just what mathematician Y needs to find the truth; if X and Y happen to be the same person, so much the better.

Now that we have examined some of the sources of mathematical creation, let us return to the creations themselves. Perhaps the greatest single innovation in the last 100 years is set theory, invented by the German mathematician Georg Cantor. In it we find some new proofs of old facts, and we also find many, many new facts. Most important of all, we find a new approach that has completely changed the methods and the spirit of all mathematics, from the most philosophical questions about its foundations to the most intricate problems of classical algebra, geometry and analysis.

Unlike many modern scientific theories, this one is based on an extremely simple and familiar notion. The word "set" means just what it says—a collection of objects or abstract entities (such as numbers or points). The letters on this page make up a set; all odd numbers make up another. Out of this apparently trivial concept flows an astonishing stream of mathematical riches.

An old fact for which Cantor found an important new proof by means of set theory is the existence of two types of numbers, known as algebraic and transcendental. To appreciate the distinction, consider the numbers we deal with in algebra. In a standard problem we start with an equation, for example $x^3 - 2x^2 + 3 = 0$, and look for the value or values of x that satisfy it. To define an algebraic number we go in the other direction: start with a number and look for an equation. To put it more precisely, a number is called algebraic if it is a solution of an equation such as $a + bx = 0$, or $a + bx + cx^2 = 0$, or $a + bx + cx^2 + dx^3 = 0$, etc., where a, b, c, d, *etc.* stand for ordinary whole numbers (possibly negative and possibly even zero). A number that is not algebraic is called transcendental.

Now there are infinitely many equations with infinitely many different solutions. So there are infinitely many algebraic numbers. The question arises: Are there any transcendental numbers? Cantor knew the answer (it is yes). But his proof of it proceeded along completely original lines. He considered the set of all numbers and the set of algebraic numbers. He then found a way to compare the sizes of these infinitely large collections.

To compare finite sets we simply count their members. Thus there is no difficulty in showing that the set of all English consonants is larger than the set of all vowels. Cantor invented a kind of generalized counting that can be applied to infinite sets [*see illustrations at bottom left on next two pages*]. He was then able to show that the set of all numbers is unmistakably larger than the set of algebraic numbers. The matter is settled; transcendental numbers must indeed exist in profusion.

Many new facts about numbers and other mathematical systems have been uncovered by Cantor's methods. His most impressive contribution, however, was his new point of view, to which almost the entire mathematical world has now been converted. Instead of considering individual numbers or points or functions, the post-Cantor mathematician considers large sets of numbers or points or functions. These have properties that cannot be ascribed to the individual elements, but that nevertheless shed some light on the elements. A set of two (or more) people can walk arm in arm, but one person cannot. And we can learn something about a person from the company he keeps.

As a rather far-fetched example, imagine that you are a European attending the Princeton-Yale football game. Between the halves one of the bands marches out on the field. You have a pair of high-powered binoculars through which you can inspect the musicians one by one. The color of their blazers means nothing to you, so you cannot tell which college they belong to. But if you put down your glasses and notice that the entire band is forming the letter P as it

$x^n + y^n = z^n$

where

x, y, z, n

are positive integers

If $n = 2$

the equation reads

$x^2 + y^2 = z^2$

and there are

many solutions

For example

$3^2 + 4^2 = 5^2$

$5^2 + 12^2 = 13^2$

etc.

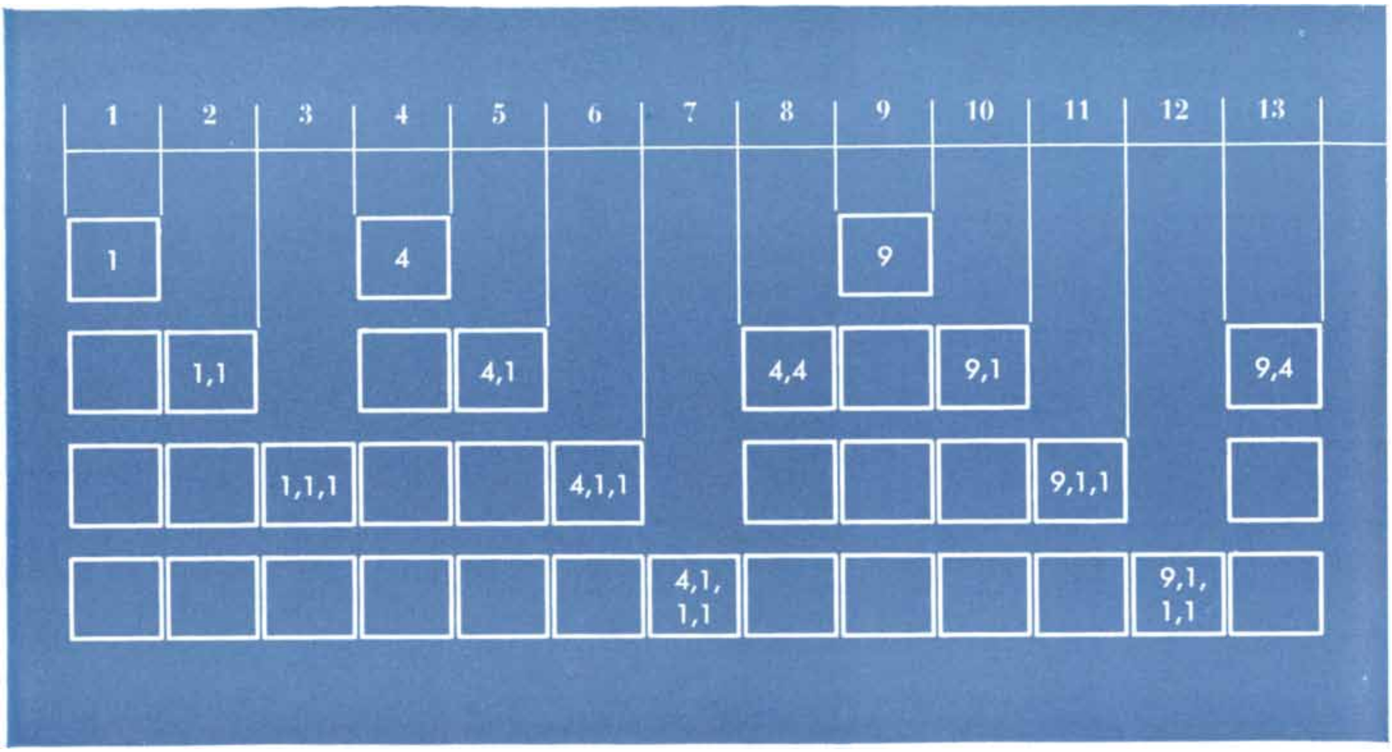
But if $n > 2$, as for instance in

$x^3 + y^3 = z^3$

there is no integral solution

for x, y, z

FERMAT'S LAST THEOREM states that the equation $x^n + y^n = z^n$ has no solutions in which x, y and z are positive whole numbers, if n is a whole number greater than two. Fermat noted in a book that he had "discovered a truly remarkable proof which this margin is too small to contain," but no one has been able to prove the theorem.

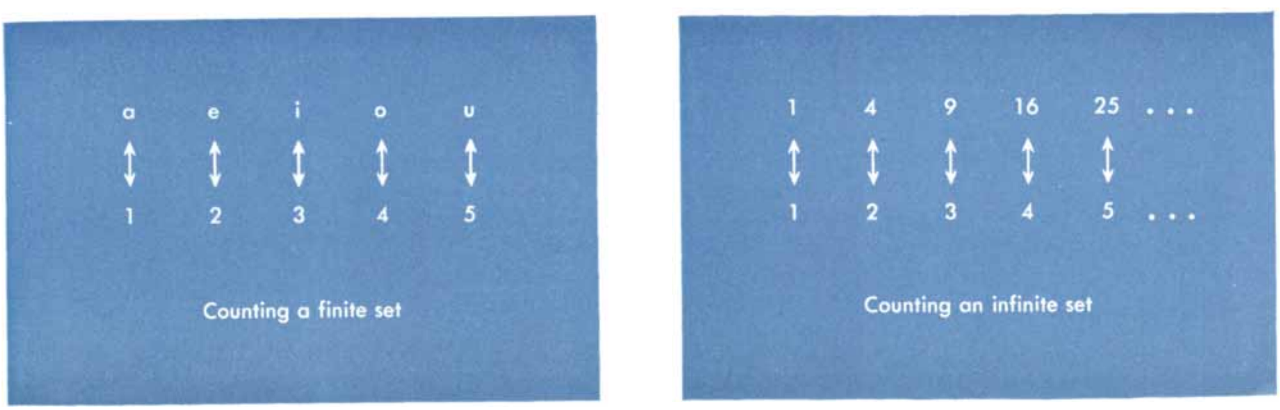


A MATHEMATICAL DISCOVERY of Euler is illustrated in this table. The theorem states that every positive whole number can be expressed as the sum of no more than four squares of other whole numbers. Rectangles in top row represent the squares themselves. Those in succeeding rows represent numbers that can be formed by adding two, three and four squares respectively. Figures

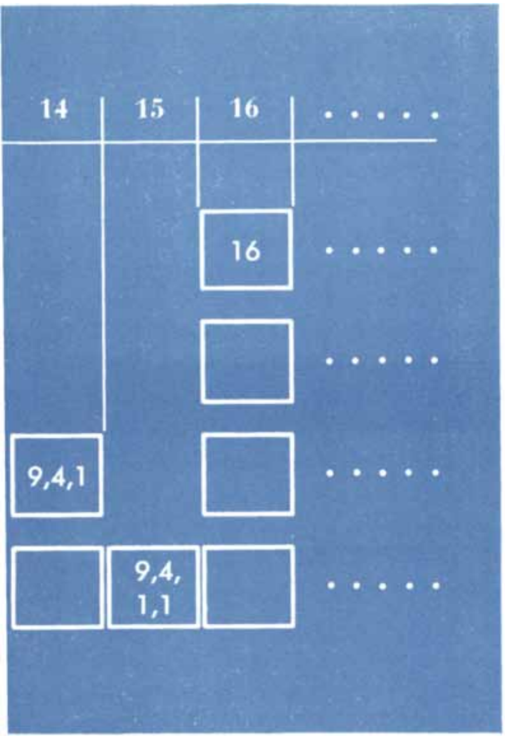
marches out, you can probably decide that they are Princetonians. Of course, it might be the Yale band making a courteous gesture, but the general idea is clear. The structure of a set can tell you something about its members.

The classical mathematician was interested in individual problems. Confronted with a system of equations, he asked: Do they have solutions? If so, what does each solution look like? The modern mathematician also wants to know the answers to these questions, but he approaches the problem differently. He might begin, for instance, by asking: "Is the sum or the product of two solutions also a solution?" This is a question about the structure of the set of all possible solutions. If the answer is yes, he knows that he is dealing with a particular type of set (for instance, a group), and this gives him important information about the individual solutions.

Some problems lead to complicated and difficult sets. Consider a few sets of points that can be chosen from a particular straight line. (Assume that the line is graduated like a thermometer. There is a point marked zero, with positive numbers on one side of it and negative numbers on the other.) Let us start with some simple sets: the set of all points above zero (the positive numbers); the set of all points (we might as well say numbers) between 2 and 7; the set of all points below — 2. All these are easily conceived, everyday sets, and can be readily visualized in geometrical dia-

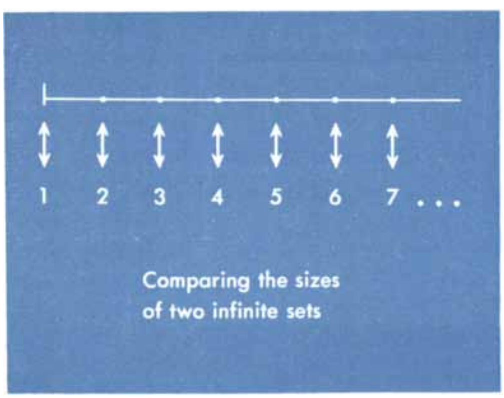


SETS are counted and compared by the method illustrated here. Counting a finite set, such as the vowels (*left*), means matching its members to positive integers. Counting an infinite set, such as the squares of the whole numbers (*middle*), also means matching each of its members to a positive integer. The infinite set of points in a line (*right*) is known to be larger than the infinite set of positive



in rectangles are the squares which, when added together, give the number. The theorem is true for all of the positive integers.

grams [*see top three illustrations at right*]. Now suppose we think of the points between the whole numbers as decimals, and construct the set of all those points with a positive odd number to the left of the decimal point. This is still not too bad; with a small mental effort we can imagine it and draw i s picture [*bottom illustration at right*]. But what about the set of points in whose decimal representation the digit 6 never appears? It is a perfectly reasonably defined set, and we know quite a



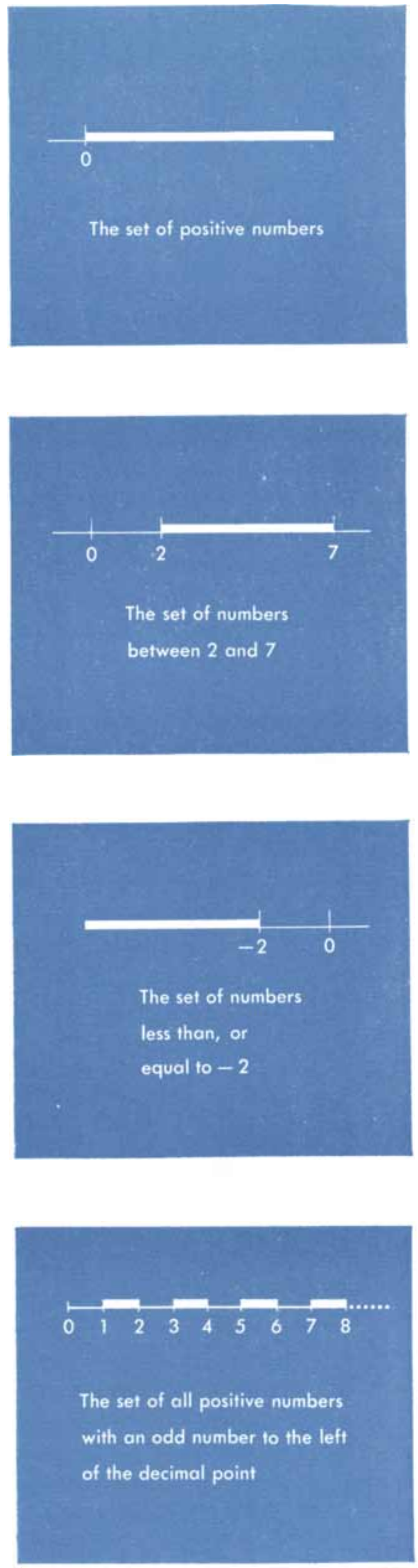
integers because, no matter how some of the points are matched up with integers, some other points will always be left unmatched.

lot about it. We know, for instance, that the point 11/20 (.55) and the point 8/7 (1.142857142857 . . .) belong to the set, and the point *pi* (3.14159265 . . .) does not. The geometric diagram of the set is probably impossible to visualize—and even so this set is much simpler than some which mathematicians are regularly forced to study. To show just one possible additional complication, consider the set of points in whose decimal representation the digit 6 may or may not occur, but never six times consecutively. Once again we know something about the set; we know, for instance, that 11/20 and 8/7 still belong to it. There is no one on the face of the earth who can decide whether *pi* belongs to the set or not.

Of course mathematicians long before Cantor had been studying certain kinds of sets (for example, lines, triangles, circles and the like), even if they did not think of them in those terms. In the early days of set theory, many mathematicians took up the idea with more enthusiasm than discretion. Any set was as good as any other. The result was mathematical anarchy. A kind of inverted snobbishness even led some workers to prefer the wildest and most unruly sets to the well-behaved, coherent sets of older days. This radicalism was not welcomed in all quarters. The great French mathematician Henri Poincaré remarked on one occasion: "Later generations will regard set theory as a disease from which one has recovered."

But after its youthful excesses post-Cantorian mathematics settled down to a mature and responsible evaluation of itself and its role in history. The set-theoretic approach is now instilled into young mathematicians virtually in the cradle, and, as a result, it is so much in their bloodstream as to have lost almost all its controversial character. It has proved to be one of the most powerful unifying themes in the history of mathematics, a theme which reveals connections between apparently remote regions of ideas.

What is the mathematics of today that will precipitate the controversies of tomorrow and become the orthodoxy of the next day? No one can say. It may take a decade, or a century, to see a mathematical innovation in its proper light. But of one thing we may be certain: As long as there is a world with mathematicians in it, innovation will continue. The new ideas will be studied, sometimes applied to practical problems, and always enjoyed.



SETS OF NUMBERS such as those in the illustrations above are easy to visualize and to represent diagrammatically. Others that are mentioned in the text are difficult to conceive and also impossible to diagram.