Fermat's Last Theorem and Modern Arithmetic

Pierre de Fermat's famous conjecture may have been proved at last. Ironically, it turns out to be a marginal note in a body of work with wider significance

Kenneth A. Ribet and Brian Hayes

Eric Temple Bell, the mathematician and biographer of mathematicians, believed that Fermat's last theorem would be one of the questions left unresolved when human civilization destroyed itself in nuclear war. Bell made this prediction shortly before his own death in 1960. If he had lived a few decades longer, it is an interesting question whether he would have been more surprised at humanity's continuing survival or at the announcement, on June 23, 1993, of a proof of Fermat's last theorem.

The theorem itself is easily stated. Pierre de Fermat asserted that if a, b and c are integers greater than 0, and if n is an integer greater than 2, then there are no solutions to the equation:

$$a^n + b^n = c^n.$$

The simplicity of the statement is deceptive: The proposition resisted all attempts at proof for more than 350 years. And the recent proof, devised by Andrew Wiles of Princeton University, requires an extraordinary arsenal of mathematical tools and techniques to attack the problem. Wiles's proof is embodied in a dense and difficult manuscript, which incorporates by reference a vastly larger body of mathematical work developed over the past 30 years or more.

As this article is being written, the status of Wiles's proof is uncertain. Checking by referees revealed a few problems, most of which were quickly resolved, but one gap in the argument appears to be more serious. Wiles has said he is confident the gap can be filled, but until the proof is published and reviewed by the larger mathematical community, the issue will remain unsettled. Indeed, the proof is not a proof and the theorem is not a theorem until all the outstanding problems have been resolved. (Nevertheless, in this article we shall continue to speak of the proof and the theorem as a matter of convenience.)

It is important to understand the true place of Fermat's last theorem in modern mathematics: It is a much-celebrated puzzle, but it is hardly a central or crucial proposition. Having a proof of the theorem may not lead to much else of great interest. On the other hand, the pursuit of a proof has contributed to the development of much important mathematics. In particular, Wiles approached the problem by setting out to prove another proposition, called the Taniyama-Shimura conjecture, from which Fermat's last theorem follows as a corollary.

The Taniyama-Shimura conjecture is deeper and potentially more significant than Fermat's last theorem itself. It belongs to a realm of mathematics that has been developing rapidly over the past three decades without attracting much notice outside the mathematics profession. This realm is called "arithmetic algebraic geometry," or "modern arithmetic." It grew out of an attempt to apply the methods of modern mathematics to the study of problems, called Diophantine problems, in which the goal is to find all solutions in whole numbers to a family of equations. Modern arithmetic has a rich structure of its own, and it seems to be connected in one way or another to every other branch of mathematics. It is remarkable that the abstract machinery of this discipline has led to a new understanding of the most famous of all Diophantine problems—Fermat's last theorem.

Marginalia

The story of how Fermat proposed his "last theorem" has been told many times, but it is too good a story to forgo telling it again. Pierre de Fermat was born in the south of France in 1601 and spent most of his life in Toulouse, where he was a prominent jurist in the bureaucracy serving Louis XIV. As a mathematician he was an amateur, but a well-connected one; he carried on extensive correspondence with René Descartes, Blaise Pascal and other luminaries of the age. Indeed, the main source of knowledge about his mathematical work is his correspondence—and his annotations in the margins of books.

Sometime in the 1630s Fermat was reading the *Arithmetic* of Diophantus of Alexandria, a work probably written in the third century C.E., which discusses various problems to be solved in whole numbers or in rational numbers (ratios of whole numbers). Fermat made numerous notes in his copy of the *Arithmetic*; the particular marginal comment of interest here per-

Kenneth A. Ribet is professor of mathematics at the University of California at Berkeley. He received his A.B. and A.M. degrees from Brown University in 1969 and his Ph.D. from Harvard University in 1973. Ribet has worked on diverse questions in number theory and arithmetic algebraic geometry; his best-known result is his proof that the Taniyama-Shimura conjecture implies Fermat's last theorem. In 1989 he received the first Prix Fermat (jointly with Abbas Bahri). Brian Hayes is a science writer and former editor of American Scientist; *his contributions to this article concern the exposition, not the mathematics. Address for Ribet: Department of Mathematics, University of California, Berkeley, CA 94720. Internet: ribet@math.berkeley.edu.*

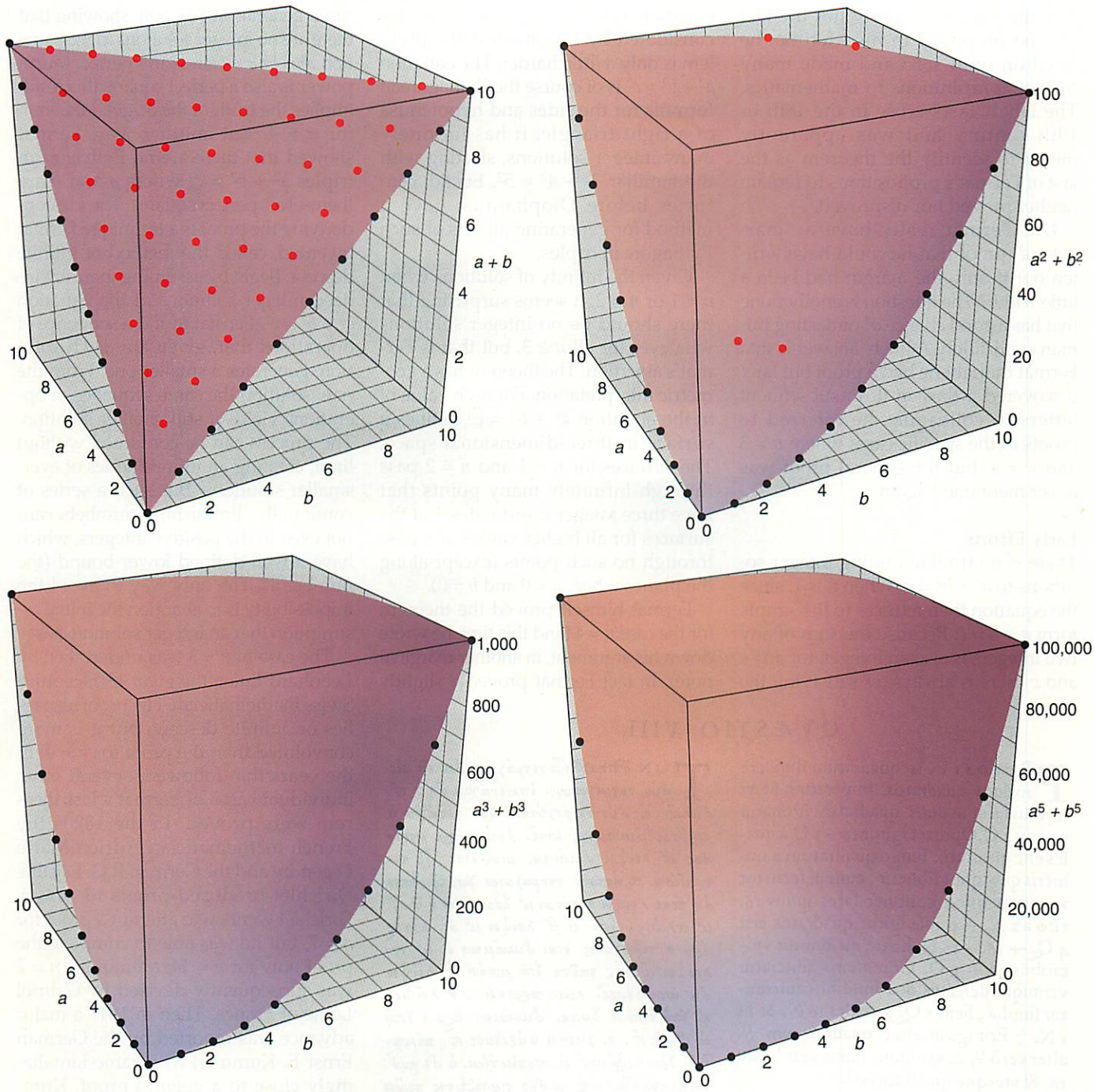


Figure 1. Fermat's last theorem, which asserts that the equation $a^n + b^n = c^n$ has no integer solutions when $n > 2$ and $abc \neq 0$, can be given a simple geometric interpretation. For any given value of n, the function $f(a,b) = a^n + b^n$ defines a smooth surface in three-dimensional space. In the case of $n = 1$ *(upper left)* the surface is a plane that passes through infinitely many points with integer coordinates; indeed, at every point where a and b are integers, the third coordinate $a + b$ is also an integer. Integer points where a or b is equal to zero are marked by black dots; other integer points are shown as red dots. For $n = 2$ *(upper right)* the surface is a paraboloid, and the only integer points are those that satisfy the Pythagorean equation $a^2 + b^2 = c^2$. Apart from points along the axes where $a = 0$ or $b = 0$, there are only four such points on the small section of the surface visible here; they correspond to the equations $3^2 + 4^2 = 5^2$ and $6^2 + 8^2 = 10^2$. On the entire surface, the number of integer points is infinite. Sections of the surfaces for $n = 3$ *(lower left)* and $n = 5$ *(lower right)* have no integer points (except along the $a = 0$ and $b = 0$ axes). Fermat's last theorem implies these surfaces can be extended to infinity without ever intersecting a point with three integer coordinates. Moreover, the theorem states that the same is true of the surfaces corresponding to all other values of n greater than 2.

tained to Question 8 in Book 2, where Diophantus asked, "Given a number that is a square, write it as a sum of two other squares." Fermat's note, translated from the Latin, reads: "It is impossible to separate a cube into two cubes or a fourth power into two fourth powers or, in general, any power greater than the second into powers of like degree. I have discovered a truly marvelous demonstration, which this margin is too narrow to contain."

The tantalizing suggestion of a proof once known and then lost to posterity has doubtless contributed to the popular romance of Fermat's last theorem. So has the designation "last," although Fermat had nothing to do with that.

The theorem was surely not the last one he proposed in his lifetime; he lived on until 1665 and made many further contributions to mathematics. The label "last" arose in the 18th or 19th century and was apparently meant to identify the theorem as the last of Fermat's propositions to remain neither proved nor disproved.

Did Fermat really have a "marvelous" proof that he could have written out if only the margin had been a little wider? The question is another one that has a good chance of outlasting human civilization. A likely answer is that Fermat thought he had a proof but later discovered a flaw in it. In subsequent letters to colleagues he referred to proofs of the specific cases where $n = 3$ and $n = 4$, but the general proof was never mentioned again.

Early Efforts

There is no trouble finding integer solutions to $a^n + b^n = c^n$ when n is 1, since the equation then reduces to the simple form $a + b = c$. Because the sum of any two integers is also an integer, for any a and b there is always a c satisfying the equation. When n is equal to 2 (the case considered by Diophantus) the problem is only a little harder. The equation $a^2 + b^2 = c^2$ is of course the Pythagorean formula for the sides and hypotenuse of a right triangle; it has infinitely many integer solutions, starting with the familiar $3^2 + 4^2 = 5^2$. Euclid, centuries before Diophantus, gave a method for generating all sets of such Pythagorean triples.

Given the infinity of solutions when $n = 1$ or $n = 2$, it seems surprising that there should be no integer solutions whatever for all $n \geq 3$, but that is Fermat's assertion. The theorem has a geometric interpretation. For each value of n, the equation $a^n + b^n = c^n$ defines a surface in three-dimensional space. The surfaces for $n = 1$ and $n = 2$ pass through infinitely many points that have three integer coordinates, but the surfaces for all higher values of n pass through no such points (except along the planes where $a = 0$ and $b = 0$).

Fermat himself proved the theorem for the case $n = 4$ (and this time he wrote down his argument, in another marginal note). In fact Fermat proved a slightly more general proposition, showing that there are no integer solutions to the equation $a^4 + b^4 = c^2$; since any perfect fourth power is also a perfect square, this result implies the truth of the original theorem for $n = 4$. Put another way, Fermat showed that there are no Pythagorean triples $a^2 + b^2 = c^2$ where a and b are themselves perfect squares. The idea underlying the proof is a technique Fermat invented, called the method of infinite descent. Begin by assuming there are indeed integer solutions of the equation $a^4 + b^4 = c^2$. Fermat found a sequence of operations that, given any such solution, generates a smaller one. From the new solution the same sequence of operations yields a still smaller solution. The process can be continued without limit, creating an infinite series of ever-smaller solutions. But such a series of continually diminishing numbers cannot exist in the positive integers, which have a well-defined lower bound (the number 1). The only way to avoid the impossibility is to abandon the initial assumption that an integer solution exists.

The case of $n = 3$ was undertaken by Leonhard Euler, the great 18th-century Swiss mathematician. His proof also relies on infinite descent, but it is more convoluted than the proof for $n = 4$. In the years that followed, several more individual cases of Fermat's last theorem were proved. In the 1820s the French mathematician Adrien-Marie Legendre and the German P. G. Lejeune Dirichlet produced proofs for $n = 5$. Dirichlet went on to attempt a proof for $n = 7$, but he was able to complete the proof only for $n = 14$; a proof for $n = 7$ was subsequently devised by Gabriel Lamé of France. Then in 1847 a major advance was reported by the German Ernst E. Kummer, who came tantalizingly close to a general proof. Kummer's work implies that Fermat's last theorem is true for an infinite class of exponents, namely all values of n that are divisible by "regular" primes, which are a subset of the prime numbers (positive integers divisible only by themselves and by 1). The only non-regular primes less than 100 are 37, 59 and 67, and Kummer was subsequently able to supply a proof for these specific values as well. Thus Fermat's last theorem was proved for all $n < 100$.

In recent years computer-aided studies have pushed the lower limit of any possible counterexample successively higher. A result published last July (Buhler, Crandall, Ernvall

QVÆSTIO VIII.

PROPOSITVM quadratum diuidere in duos quadratos. Imperatum fit vt 16. diuidatur in duos quadratos. Ponatur primus 1 Q. Oportet igitur 16 − 1 Q. æquales eſſe quadrato. Fingo quadratum à numeris quotquot iibuerit, cum defectu tot vnitatum quod continet latus ipſius 16. eſto à 2 N. − 4. ipſe igitur quadratus erit 4 Q. + 16. − 16 N. hæc æquabuntur vnitatibus 16 − 1 Q. Communis adiiciatur vtrimque defectus, & à ſimilibus auferantur ſimilia, fient 5 Q. æquales 16 N. & fit 1 N. $\frac{4}{5}$. Erit igitur alter quadratorum $\frac{256}{25}$. alter verò $\frac{144}{25}$ & vtriuſque ſumma eſt $\frac{400}{25}$ ſeu 16. & vterque quadratus eſt.

ΤΟΝ δὲ ὑπαρχθέντα τετράγωνον διελεῖν εἰς δύο τετραγώνους. ἐπιτετάχθω δὴ τὸν ιϛ̄ διελεῖν εἰς δύο τετραγώνους. καὶ τετάχθω ὁ πρῶτος δυναμεως μιᾶς. δεήσει ἄρα μονάδας ιϛ̄ λείψει δυνάμεως μιᾶς ἴσας ἔσῃ) τετραγώνῳ. πλάσω τὸ τετράγωνον ἀπὸ ϛ̄ς. ὅσων δή ποτε λείψει τοσούτων μ̄ ὅσων ἔχει ἡ τοῦ ιϛ̄ μ̄ πλευρά. ἔστω ϛ̄ς β̄ λείψει μ̄ δ̄. αὐτὰς ἄρα ὁ πρῶτος ἔσται δυνάμεων δ̄ μ̄ ιϛ̄ λείψει ϛ̄ς ιϛ̄. ταῦτα ἴσα μονάσι ιϛ̄ λείψει δυνάμεως μιᾶς. κοινὴ προσκείσθω ἡ λεῖψις, καὶ ἀπὸ ὁμοίων ὁμοῖα. δυνάμεις ἄρα ἕ ἴσαι ἀριθμοῖς ιϛ̄. καὶ γίνεται ὁ ἀριθμὸς ιϛ̄. πέμπτων. ἔστω ὁ μὲν σνς̄ εἰκοσοπέμπτων. ὁ δὲ ρμδ̄ εἰκοσοπέμπτων, ἃ οἱ δύο συναμφότερος ποιοῦσι τὼ εἰκοσόπεμπτα, ἤτοι μονάδας ιϛ̄. καὶ ἔστιν ἑκάτερος τετράγωνος.

OBSERVATIO DOMINI PETRI DE FERMAT.

CVbum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos & generaliter nullam in infinitum vltra quadratum poteſtatem in duos eiuſdem nominis-fas eſt diuidere cuius rei demonſtrationem mirabilem ſane detexi. Hanc marginis exiguitas non caperet.

Figure 2. Famous marginal note in which Pierre de Fermat stated his "last theorem" is preserved in an edition of the *Arithmetic* of Diophantus published by Fermat's son Samuel. Fermat read the *Arithmetic* in the first modern edition, published in 1621 by Claude-Gaspar Bachet, but the copy with his annotations has not been found. In Samuel Fermat's edition the note is transcribed below the Latin and Greek texts of Question 8 in Book 2. It reads: "It is impossible to separate a cube into two cubes or a fourth power into two fourth powers or, in general, any power greater than the second into powers of like degree. I have discovered a truly marvelous demonstration, which this margin is too narrow to contain." (Photograph courtesy of the Duke University Special Collections Library.)

and Metsänkylä 1993) implies that Fermat's last theorem must be true for all exponents n less than 4 million, and so any integer solution to $a^n + b^n = c^n$ would have to consist of astronomically large integers. (It turns out the smallest possible value of c^n would be a number of more than 26 million decimal digits.) Still, no mathematician could consider the question settled just because a finite range of cases has been dismissed. There are infinitely many of Kummer's nonregular primes, and so no extension of the case-by-case analysis can ever be complete.

What Needs to Be Proved

The modern approach to Fermat's last theorem is an indirect one. It does not attack Fermat's equation $a^n + b^n = c^n$ itself but instead analyzes a new equation of a different form, in which the numbers a^n and b^n have an essential role.

In broad outline, the argument goes as follows. Suppose there is a counterexample to Fermat's last theorem, or in other words a pair of integers a^n and b^n whose sum is also a perfect nth power. Then there must exist a mathematical object called an elliptic curve specified by an equation whose coefficients are determined by a^n and b^n. Call the elliptic curve E. One of us (Ribet) proved in 1986 that the curve E cannot have a certain property called modularity. What Wiles announced last June is that all elliptic curves in a class that includes E are modular. From this contradiction it follows that E cannot exist, and neither can the supposed counterexample to Fermat's last theorem.

In this article we shall fill in a few details of this argument. In particular we shall explain what an elliptic curve is and what it means to be modular. A rigorous account of the entire proof and its background would be very demanding indeed, and so we shall only illuminate the main points.

It is best to begin by refining somewhat the question of what is to be proved. Specifically, several constraints can be put on the values of a, b, c and n in the Fermat equation $a^n + b^n = c^n$. First, we can confine our attention to cases where n is an odd prime number. Only prime exponents need be considered because any counterexample to the theorem in which n is composite would imply the existence of a smaller counterexample with a prime exponent. In other words, if $a^{pq} + b^{pq} = c^{pq}$ has a solution in integers, then so must both $a^p + b^p = c^p$ and $a^q + b^q = c^q$. The only composite values of n that escape this reasoning are powers of 2, since they have no odd factors. However, all powers of 2 greater than the first power are divisible by 4, and Fermat's own proof dispenses with this case. Indeed, strictly speaking there is no need to worry about any multiples of 3, 4, 5, 7 and all primes less than 4 million, but exploiting these known results offers no advantage in devising a general proof. Wiles's proof applies to all prime values of n greater than or equal to 5.

A similar line of argument eliminates all cases except those where a, b and c are relatively prime—that is to say, they have no factors in common. Again, if you knew a counterexample in which a, b and c had a common factor, you could divide both sides of the equation by that factor to produce a smaller solution.

We shall state two further facts about the values of a^n, b^n and c^n without explaining them in detail. Exactly one of a, b and c must be even; we shall assume that the even number is b. Since n is at least 5, b^n must be divisible not only by 2 but also by 2^5, or 32. Among the two odd values, one must be congruent to 1 modulo 4 (that is, it must leave a remainder of 1 when divided by 4), and the other must be congruent to 3 modulo 4. Here we shall assume that a^n is 3 modulo 4.

From here on we can essentially leave behind Fermat's equation and work with new variables A, B and C, which represent a^n, b^n and c^n respec-

The Bettmann Archive

Figure 3. Pierre de Fermat (1601–65) was a successful lawyer and judge, but he is known today almost entirely for his activities in a realm where he was an amateur: mathematics. Although he did not publish his mathematical work, he carried on an extensive correspondence with other savants, and made major contributions to number theory and the study of probability. This engraving appears in his son's edition of Diophantus.

tively. The new variables must satisfy all the constraints we have just established. In particular:

$A + B = C$
$ABC \neq 0$
A, B and C are relatively prime
B is divisible by 32
$A = 3$ modulo 4 and $C = 1$ modulo 4.

A final property of A, B and C should not be left unstated: To serve as a counterexample to Fermat's last theorem, A, B and C must be perfect nth powers, where n is a prime ≥5. Hence the product ABC is also a perfect nth power, since $a^n b^n c^n = (abc)^n$.

It is at this point that elliptic curves enter the story. The curve of interest is described by the equation

$$y^2 = x(x - A)(x + B)$$

where the numbers A and B are derived from the hypothetical counterexample to Fermat's last theorem introduced above. (Although C does not appear in the equation, no information is lost because C can be represented by the sum $A + B$.) The strategy of the proof will be to show that the curve specified by this equation has the property that it is not modular, and then to show that all elliptic curves in a class that includes this one must be modular. The only escape from the contradiction is to acknowledge that numbers with all of the properties attributed to A, B and $A + B$ cannot exist.

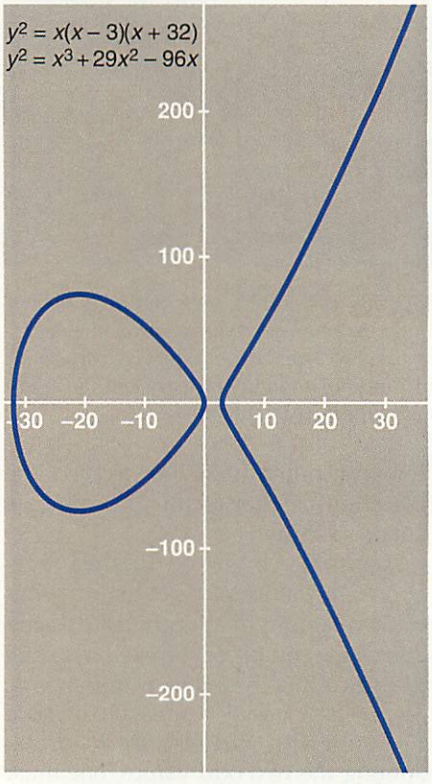


Figure 4. Elliptic curve is the locus of points that satisfy a certain cubic equation. Such curves have a deep connection with Fermat's last theorem. Specifically, if there were a counterexample to the theorem, it would imply the existence of an elliptic curve with some highly distinctive properties. The curve shown here is defined by the equation $y^2 = x^3 + 29x^2 - 96x$, or in factored form $y^2 = x(x - 3)(x + 32)$.

Elliptic Curves

Before proceeding with our account of the proof, we must pause to introduce these interesting mathematical objects called elliptic curves. To begin with, an elliptic curve is *not* an ellipse. The name reflects a connection with elliptic functions (which were devised to help calculate the perimeter of an ellipse, but which have turned out to have other uses as well). Elliptic curves are plane curves defined by a certain class of cubic equations, and their shape is not even vaguely elliptical.

We can construct a specific elliptic curve by choosing values for the numbers A and B in the equation given above. The constraints on these numbers require that A be congruent to 3 modulo 4, and so a straightforward assignment is $A = 3$. Similarly, B must be divisible by 32, and so a logical choice is $B = 32$. (Of course $A = 3$ and $B = 32$ do not actually constitute a counterexample to Fermat's last theorem; they are merely numbers that satisfy some of the criteria that would have to be met by any such counterexample.) With these substitutions, the equation becomes:

$$y^2 = x(x - 3)(x + 32).$$

Multiplying the three factors on the right-hand side of the equation yields an equivalent expression:

$$y^2 = x^3 + 29x^2 - 96x.$$

Note that this is a cubic equation, or an equation of degree 3, since the highest power of any variable is a cube; more specifically the equation relates the square of y to a cubic polynomial in x. Also note that the coefficients of the equation are integers; in general the coefficients of an elliptic curve can be numbers of any kind, but the curves we shall consider here will be assumed to have integer coefficients unless stated otherwise.

The equation we have just constructed defines a curve in the x, y plane. The curve is the locus of all points whose x and y coordinates satisfy the equation. For example, the point (0, 0) is clearly on the curve, since the substitution $x = 0$, $y = 0$ makes the equation a true statement. The curve defined by the equation is sketched in Figure 4. It consists of two disconnected pieces: a closed loop to the left of the y axis and an infinitely long open segment to the right. The precise form of the curve depends on the values of the coefficients, and in some cases it is a single connected piece.

Not every cubic equation generates an elliptic curve. To qualify, a curve must be smooth, or nonsingular, a concept that can be made more precise by saying that the curve must have a well-defined tangent at every point along its length. The curve may not have any cusps, where there is no defined tangent; nor may it cross itself, creating a nodal point with two or more tangents. Figure 5 shows a few examples of smooth elliptic curves and of other cubic curves having singularities.

From an algebraic point of view, requiring that the curve be nonsingular is equivalent to requiring that the equation have three distinct roots, or in other words that three different values of x make the expression $x(x - A)(x + B)$ equal to zero. Obviously one of the roots is $x = 0$, and the others are $x = A$ and $x = -B$. Hence to generate a proper elliptic curve, the values of A and B must satisfy three constraints: $A \neq 0$, $B \neq 0$ and $A \neq -B$. The last constraint can be rephrased as $A + B \neq 0$, or equivalently $C \neq 0$, and so the overall requirement is $ABC \neq 0$.

Why have mathematicians focused so much attention on this one family of curves? After all, there are infinitely many polynomial equations in x and y, generating an endless variety of curves in the plane. What is so special about the elliptic curves? One answer is that elliptic curves are the first nontrivial family of curves from the Diophantine point of view.

All plane curves—or the equations that generate them—can be classified according to their *genus*, which is a number closely related to the equation's degree. Specifically, a nonsingular curve defined by an equation of degree d has a genus equal to $(d - 1)(d - 2)/2$. Lines and conic sections—ellipses, parabolas and hyperbolas—are defined by linear equations ($d = 1$) or quadratic equations ($d = 2$) and necessarily have genus 0. Elliptic curves, which are nonsingular by

definition and have degree 3, are curves of genus 1. Nonsingular equations of the fourth, fifth or greater degree yield curves of higher genus. In 1922 Louis J. Mordell made an intriguing observation relating the genus of an equation to the number of rational solutions it has, or equivalently the number of points with rational coordinates the curve passes through. It was already known that curves of genus 0 always have either no rational solutions or infinitely many, and the infinite cases are always easy to describe. Mordell conjectured that all curves of genus 2 or greater have at most finitely many rational solutions. In 1983 Mordell's conjecture was proved (to the surprise of the mathematical community) by Gerd Faltings, a young mathematician then at the University of Wuppertal in Germany. Curves of genus 1—namely the elliptic curves—remain as an intermediate case with no simple way of telling whether the number of solutions is finite or infinite.

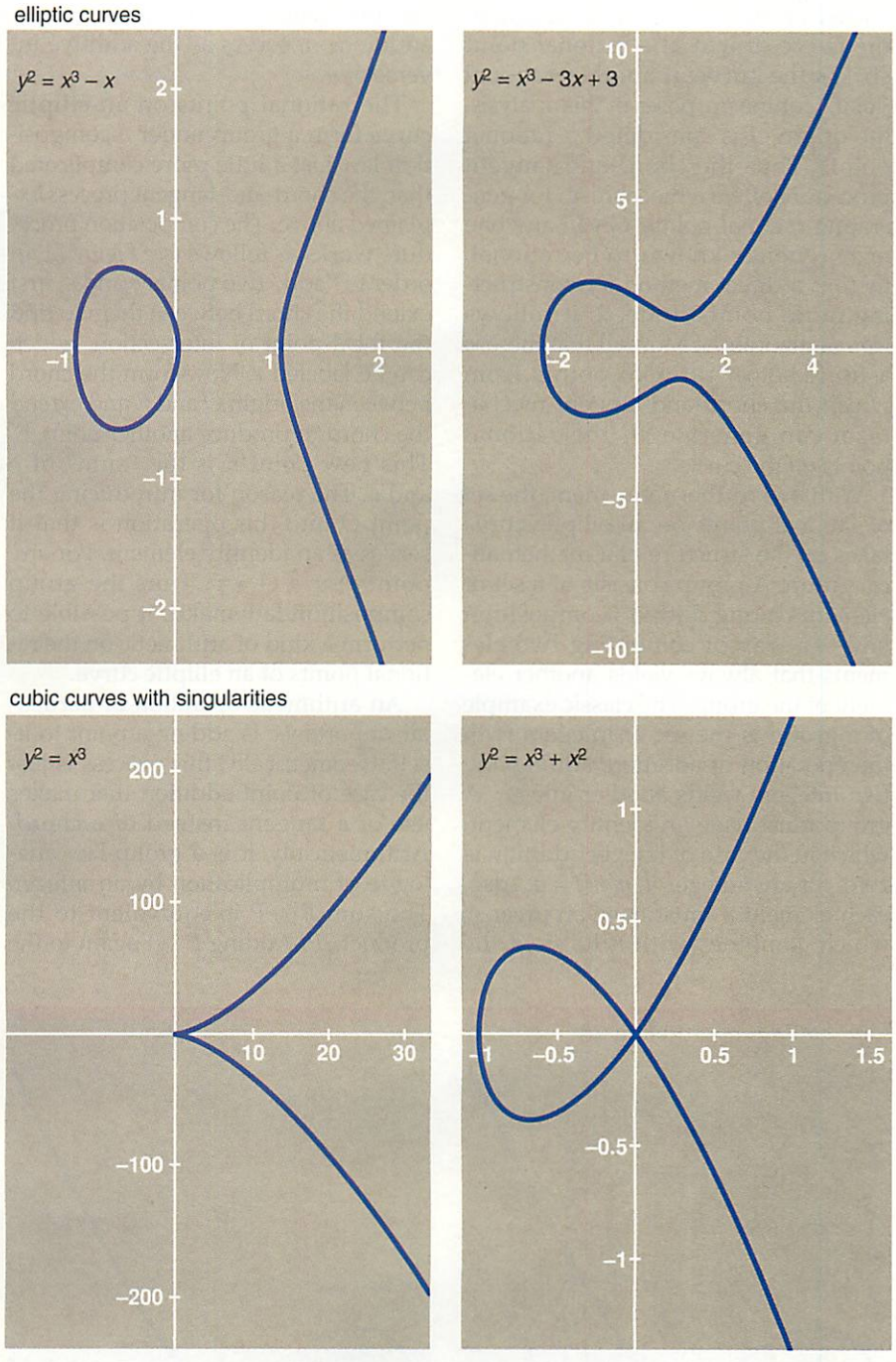


Figure 5. Cubic equations define curves of various forms, only some of which qualify as elliptic curves. The curves shown in both of the upper graphs are elliptic curves; one of them consists of two separate pieces, whereas the other is a single connected strand. The curves in the lower graphs are not elliptic because they have singularities, or points where the curve does not have a unique tangent. Both curves are singular at the point (0, 0).

Chords and Tangents

On an elliptic curve, the number of rational points can be either finite or infinite, depending on the coefficients of the specific equation. In all cases, however, the set of rational points has a rich structure, allowing them to be explored systematically.

If you draw a line, or chord, between any two points on an elliptic curve—whether the points are rational or irrational—the line can be extended to intersect the curve at a third point. Some obvious exceptions to this rule are chords drawn between points that have the same x coordinate, so that the chords are parallel to the y axis. These "vertical" chords extend to infinity without striking the curve again. The exceptional nature of the vertical chords can be eliminated by adding to the curve a single extra point at infinity, where all the vertical chords can be imagined to converge. This extra point is called the origin and is designated by the symbol O; any vertical chord must pass through O. Thus when the origin is made a part of the curve, every chord intersects the elliptic curve at exactly three points.

There is a similar construction for lines drawn tangent to an elliptic curve: Every tangent intersects the curve at one point in addition to the point of tangency. In the special case of vertical tangents, the additional point of intersection is the origin O. A tangent can be thought of as the limiting case of a chord that gets shorter and shorter until finally its end points coincide. In effect, then, the tangent is a chord that passes through the same point twice. According to this line of reasoning, every chord or tangent drawn on an elliptic curve has exactly three points of intersection. It is to preserve this property that cubic curves with singularities are excluded from the category of elliptic curves, since cusps or other singularities are points that do not have a unique tangent.

The geometry that makes possible the chord-and-tangent construction on elliptic curves becomes more remarkable when attention turns to the rational points on the curve—the points whose x and y coordinates are both rational numbers. When a chord is drawn through any two rational points, the third point of intersection is also ratio-

nal (*see Figure 6*). Similarly, a tangent to the curve drawn at a rational point strikes the curve at another rational point. (For the purposes of this analysis, the origin O is considered a rational point.) Thus the chord-and-tangent procedure offers a mechanism for generating rational points: Given any one or two points known to be rational, there is a direct method for constructing more points. Indeed, it follows from a theorem of Mordell that there is a finite set of rational points from which the chord-and-tangent mechanism can generate *all* the rational points on the curve.

With one further refinement, the set of rational points on an elliptic curve takes on the structure of a mathematical group. A group consists of a set of elements along with a "composition law"—a way of combining two elements that always yields another element of the group. The classic example of a group is the set of integers with the operation of addition; adding any two integers yields another integer. A group must have an identity element, which in the case of integer addition is zero; for any integer n, $n + 0 = n$. Also, each element n must have an inverse, which combines with n to yield the identity element; in the case of integer addition, $-n$ serves as the additive inverse of n.

The rational points on an elliptic curve form a group under a composition law just a little more complicated than the chord-and-tangent process explained above. The composition procedure works as follows (*see Figure 7*). In order to "add" two points p and q, first extend the chord between them to find the third point of intersection, which can be labeled r. Now form the chord between the origin O and r, and extend the chord to produce another point, r'. This new point r' is the "sum" of p and q. The reason for introducing the point O into this operation is that it serves as an identity element. For any point p, $p + O = p$. Thus the group composition law makes it possible to perform a kind of arithmetic on the rational points of an elliptic curve.

An arithmetic operation of particular importance is adding a point to itself. Geometrically, this process is just the case of point addition that makes use of a tangent instead of a chord. Arithmetically, it is a group-law analogue of multiplication by an integer. The sum $P + P$ is equivalent to the product $2P$. Adding P yet again to the result of these operations yields the product $3P$, and so on. For some points this process can be continued indefinitely without ever returning to a point already visited; such points are said to be of infinite order. Other points, of finite order, can be added to themselves only a finite number of times before they yield O as the product; after that, adding P to O yields P again by definition, and the finite sequence of points repeats. Figures 8 and 9 show some examples of finite-order and infinite-order points.

What does the arithmetic of elliptic curves have to do with Fermat's last theorem? The connection is through the concept of a curve's being modular, or associated with a modular form, which we turn to next.

What It Means to Be Modular

The study of elliptic curves can be traced back to Fermat and even to Diophantus, and modular forms have their origins in the 19th century, but the two areas have been deeply linked only since 1955. In that year Yutaka Taniyama, a young Japanese mathematician, made a bold conjecture, which initially took the form of a series of problems presented at a conference.

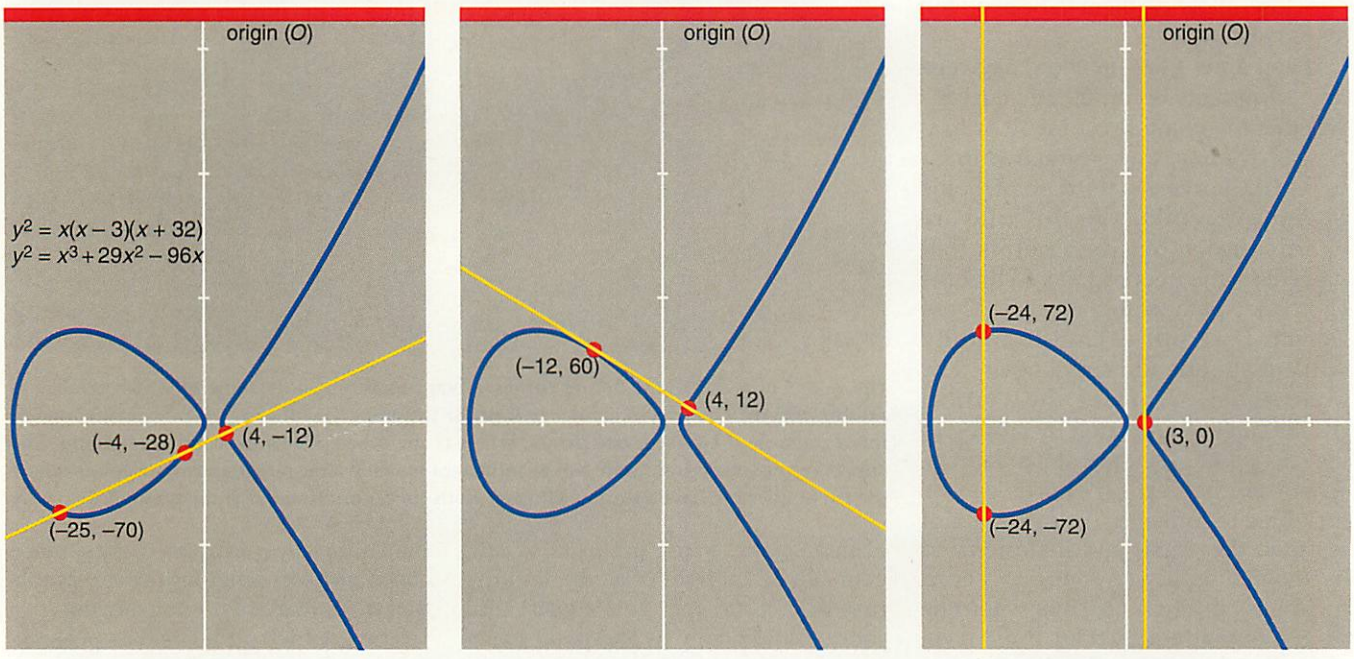


Figure 6. Chord-and-tangent process illuminates the relations of rational points on an elliptic curve. The procedure works as follows. Given any two points on the curve with rational coordinates, construct a line connecting them and extend it to infinity in both directions. The line will either be a chord that intersects the curve at a third rational point *(left diagram)*, or it will be tangent to the curve at one of the selected points *(middle diagram)*. A tangent can be thought of as a kind of degenerate chord that intersects the curve twice at the same point, so that the tangent, too, effectively touches the curve at three rational points. The only apparent exceptions to this rule are chords and tangents parallel to the y axis *(right diagram)*. To cope with these exceptions, an extra point, called the origin, is added to the curve; the origin is a point at infinity, but it is shown here as a line at the top of each diagram. The origin becomes a third point on all vertical chords and tangents, which accordingly have the same geometry as other chords and tangents.

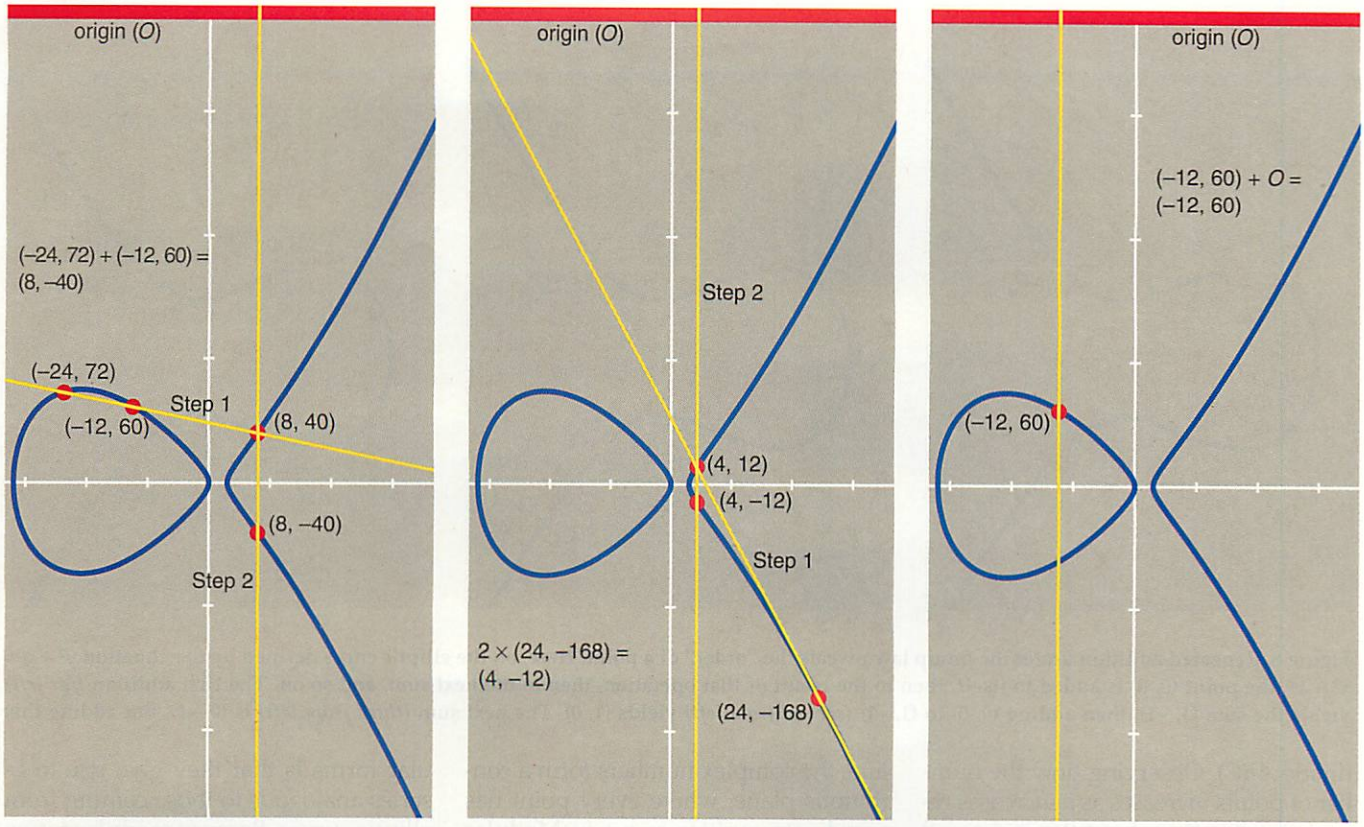


Figure 7. Rational points on an elliptic curve form a mathematical group under a slightly more elaborate version of the chord-and-tangent process. The group composition law provides a means of "adding" two points *(left diagram)*: First draw a chord connecting them and find the third point of intersection; then construct another chord between this third point and the origin. The third point on this latter chord is the sum of the original two points. To add a point to itself *(middle diagram)*, construct a tangent rather than a chord in the first step, then again draw a line from the origin through the third point identified by the tangent. The importance of the origin in these procedures is that it serves as the identity element of the group. Adding any point to the origin *(right diagram)* yields the same point again.

The conjecture was made more precise by Goro Shimura of Princeton University, and it is now known as the Taniyama-Shimura conjecture. It contends that all elliptic curves with rational coefficients are modular. At the time, this proposal was viewed with a certain skepticism, but it has grown in credibility over the years. Even before Wiles undertook to prove the Taniyama-Shimura conjecture, many mathematicians had come to believe that it is probably true.

One reason the Taniyama-Shimura conjecture seemed so unlikely at first is that elliptic curves and modular forms are very different kinds of objects. To see how they are connected, consider again the set of rational points on an elliptic curve. There are various questions one might ask about these points for any given curve. How many are there? If the number is finite, is there a method for counting them? Are there any patterns that govern where they appear along the curve? Can they be classified?

A fruitful approach to such questions is to think of the equation defining an elliptic curve not as an equivalence but as a congruence modulo some prime number p. In other words, "reduce" the equation by dividing all values of x and y by p, saving only the remainder. We can illustrate this process for the elliptic curve defined by the equation $y^2 + y = x^3 - x^2$. The curve has just five rational points, namely (0, 0), (0, –1), (1, 0), (1, –1) and the origin. Now consider the equation as a congruence modulo 7. All of the five points listed above remain solutions of the equation modulo 7. What is more, additional points, which do not lie on the curve itself, become solutions when the equation is reduced modulo 7. For example, the point (5, 1) becomes a solution because $1^2 + 1$ modulo 7 is congruent to $5^3 - 5^2$ modulo 7 (they both leave a remainder of 2). Reducing modulo 5 yields a different set of solution points, and reducing by 13 generates still another set.

In general, reduction is not possible for all primes. After the reduction, the equation must still specify a nonsingular curve, which means that the three roots must remain distinct modulo p. For an equation of the form $y^2 = x(x - A)(x + B)$, with A and B meeting the various criteria set forth above, this condition is satisfied for all p that do not divide the product $AB(A + B)$, or equivalently ABC. For the specific curve $y^2 = x(x - 3)(x + 32)$ the admissible primes are those that do not divide $3 \times 32 \times 35 = 3{,}360$. Hence the curve cannot be reduced by the primes 2, 3, 5 and 7, since they all divide 3,360. The product of the primes that divide ABC—in this case $2 \times 3 \times 5 \times 7 = 210$—is called the *conductor* of the elliptic curve; it specifies the set of primes that give such "bad reductions."

What is gained by reducing an elliptic curve modulo p? There is one immediate benefit: The number of rational points on the reduced curve is guaranteed to be finite. Exploring a finite object is often easier than studying an infinite one. Beyond this consideration, there is the hope that "local" solutions, relative to a specific prime, may reveal something about the "global" solutions to the original equation. In particular, one can study an elliptic curve by counting the number of solutions modulo p for many primes p (always excluding those that

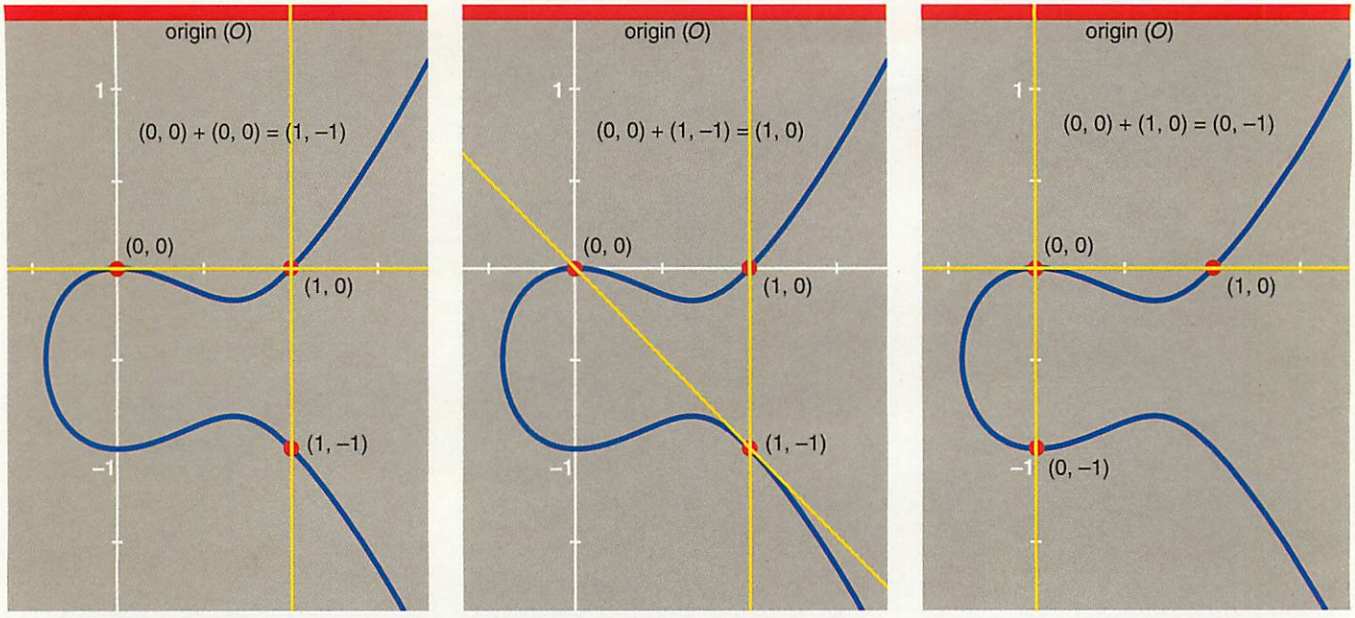


Figure 8. Repeated addition under the group law reveals the "order" of a point. Here, on the elliptic curve defined by the equation $y^2 + y = x^3 + x^2$, the point (0, 0) is added to itself, then to the result of that operation, then to the next sum, and so on. The first addition *(far left)* yields the sum (1, –1), then adding (0, 0) to (1, –1) *(second from left)* yields (1, 0). The next sum *(third from left)* is (0, –1), but adding this

divide ABC). Observing how the number of points increases as p increases reveals information about the curve and especially about the group law for the rational points. The information is encoded in a mathematical object called an L-series, made up from a recipe that begins with numbers designated a_p, which measure how many points modulo p there are for each prime p. The exact relation between the L-series and the size of the group of rational points is the subject of an unproved conjecture; the basic idea is that a curve with many rational points should have many points modulo p for various p. The converse should also be true.

The complete L-series is an infinite product, incorporating information from infinitely many primes, but finite approximations can be calculated to any desired accuracy for any specific elliptic curve by the direct method of counting rational points modulo many primes. This process is an arduous one, however, and the resulting L-series is in a format that makes it hard to use in further calculations. Modular forms offer a remarkable shortcut—at least for certain elliptic curves, and possibly for all of them.

Modular forms come from a rather different realm of mathematics: They are analytic functions defined on the complex numbers. (A complex number has both a real and an imaginary part; the imaginary part is a multiple of i, the square root of –1.) Just as the real numbers can be arranged on a continuous line, the complex numbers form a continuous plane, where every point has coordinates written as $x + iy$. Modular forms are defined on the complex upper half-plane, the part of the complex plane with $y > 0$. In other words, a modular form is a function that takes each complex number from the upper half-plane and returns as its value another complex number (or possibly the same number).

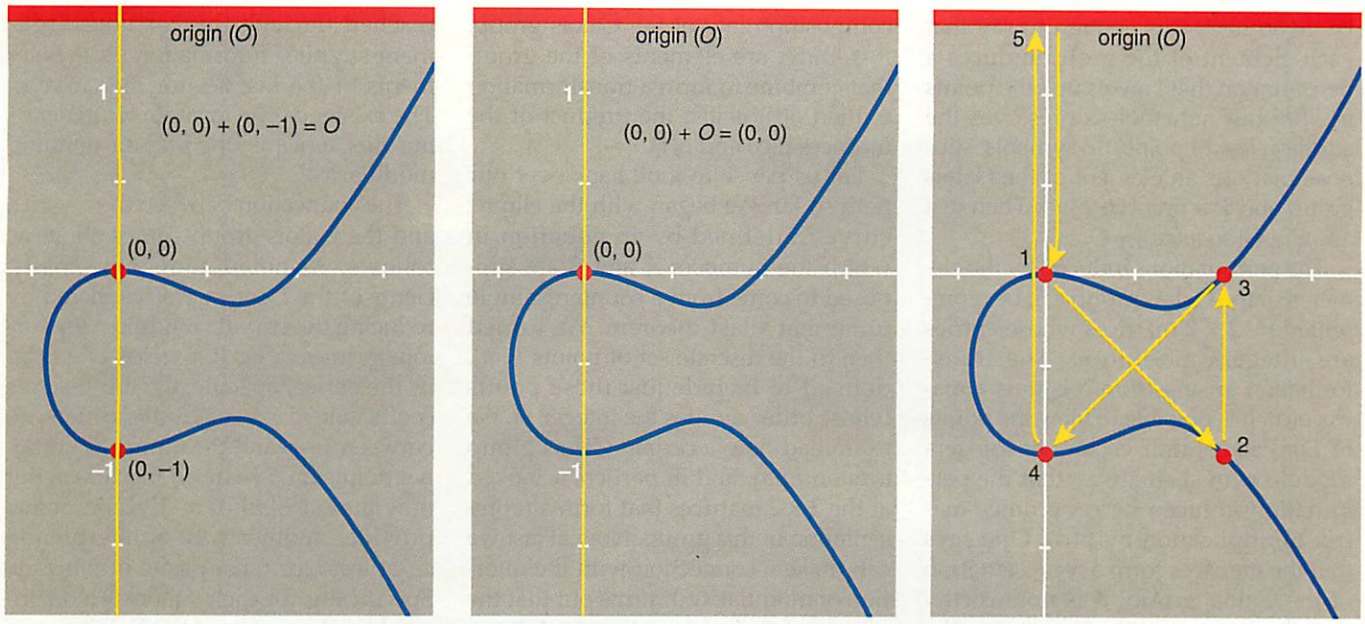
A salient characteristic that distinguishes modular forms among the functions of complex analysis is that they are essentially invariant under certain transformations of the upper half-plane. These transformations, defined by square integer matrices, are known as fractional-linear transformations. For example, each modular form f is invariant under integer translations: The value of f at the complex number z is the same as its value at $z + 1$. In other cases the modular form is not strictly invariant but is multiplied by a simple factor. The "level" of a modular form f is a positive integer that determines the set of fractional-linear transformations that leave f invariant. Roughly speaking, the space of modular forms of level N grows as N grows. For example, there are no nonzero modular forms of level 12 or of level smaller than 10, but there is a nonzero form of level 11; it is unique up to multiplication by a constant.

In connection with Fermat's last theorem and the Taniyama-Shimura conjecture, an important property of modular forms is that they give rise to L-series analogous to those coming from elliptic curves. Because modular forms and L-series both belong to the realm of complex analysis, studying the L-series attached to an elliptic curve becomes easier once it is known that the same L-series is also attached to a modular form. And it is the content of the Taniyama-Shimura conjecture that for each elliptic curve there is a modular form whose L-series is the same as that of the elliptic curve.

Since elliptic curves are algebraic objects, the Taniyama-Shimura conjecture represents a deep connection between algebra and complex analysis. Implausible as the conjecture seemed at first, it is now supported by a vast amount of numerical and philosophical evidence. Wiles's proof of the conjecture for a major class of elliptic curves can be viewed as further support for this remarkable connection.

Proving E Is Not Modular

The sequence of events leading up to Wiles's announcement last summer was set in motion in 1985 with a conjecture by Gerhard Frey of the University of the Saarlands in Germany. It was Frey who drew attention to equations of the form $y^2 = x(x - A)(x + B)$, where A and B are supposed to come from a counterexample to Fermat's last theorem. The corresponding elliptic curve is now often called the Frey curve. Because of constraints on the



point to (0, 0) *(fourth from left)* leads to the origin. Since the origin is the identity element, adding (0, 0) to the origin *(fifth from left)* necessarily returns the process to (0, 0) again. The point (0, 0) is said to be of order five, since five additions bring it back to the starting position. The sequence of operations is summarized in the diagram at far right.

numerical values of A and B, Frey understood that the elliptic curve $y^2 = x(x - A)(x + B)$ cannot be modular. He was unable to give a rigorous proof, but Jean-Pierre Serre of the Collège de France soon clarified precisely what was needed to justify Frey's insight: He formulated an explicit conjecture about modular forms whose truth would imply that Frey was correct. A year later one of us (Ribet) supplied the proof. This result established a direct link between elliptic curves and Fermat's last theorem, since the Taniyama-Shimura conjecture asserts that *all* elliptic curves are modular.

How can one demonstrate that an elliptic curve is or is not modular? For a specific equation with known, numerical coefficients, there are computational methods of answering such questions; the methods are arduous but reliable. The computational methods cannot be employed in this instance, however, because the object of study is a curve whose existence is hypothetical. We can write down Frey's equation as $y^2 = x(x - A)(x + B)$, but unless we know of a counterexample to Fermat's last theorem, we cannot fill in numerical values for A and B; and if the theorem is true, of course, there *are* no such values. Since there is no hope of calculating the properties of a curve we cannot exhibit, the only recourse is to choose a less direct strategy.

The first stage in this process is to examine a special subgroup of the points on an elliptic curve. For a curve E and a chosen integer m, this subgroup is designated $E[m]$, and it consists of those points of order dividing m; such points are called the m-division points. Recall that the order of a point is the number of times it must be added to itself before the result is equal to the origin O. Thus the group $E[m]$ is made up of all points that after multiplication by m (or equivalently after being added to themselves m times) yield the origin. There is a reason for scrutinizing this rather curious set of points. If the elliptic curve E is modular, then a study of $E[m]$ reveals information about the modular form associated with E. Moreover, there is a sense in which $E[m]$ itself can be modular, and if you can show that $E[m]$ is modular for an infinite number of m, then you can show that E is modular. Conversely, if you can show that $E[m]$ is *not* modular for some value of m, then you know that E cannot be modular either.

A couple of caveats need to be mentioned here. First, on some elliptic curves there are points of infinite order, which will not be included in the set $E[m]$ for any finite value of m. Second, the points in $E[m]$ do not necessarily have integer coordinates or even rational coordinates. The most that can be said of the coordinates is that they are algebraic numbers: solutions of algebraic equations with rational coefficients.

How can one prove that $E[m]$ is not modular for some m? The key is to set m equal to n, the exponent in the hypothetical counterexample to Fermat's last theorem. As noted above, in any such counterexample the product ABC is necessarily a perfect nth power. But making ABC an nth power gives the group $E[n]$ some unusual properties, akin to those of the n-division points on an elliptic curve of conductor 2. It is already known, however, that no elliptic curve of conductor 2 can possibly exist; the smallest conductor is 11. The suggested connection with curves of conductor 2 comes very near to being a contradiction that would prove Fermat's last theorem directly, without reference to the Taniyama-Shimura conjecture. Unfortunately, no one has yet found a way to turn the hint into a rigorous demonstration. The actual proof that $E[n]$ is not modular follows a more circuitous path. On the hypothesis that $E[n]$ is modular, it must be associated with a modular form of some minimal level. The essential point of the proof is to show that this level is 2, which is impossible since there are no nonzero modular forms of level 2.

The argument supporting this conclusion is an intricate one, which wanders into still denser thickets of modern arithmetic. The place to begin is to look at the action of a group of transformations, called a Galois group, on the elements of $E[m]$, for each value of m. The definition of the Galois group will not

be given here; it is enough to say that each element of the group induces a permutation that "mixes up" the points in $E[m]$ but nonetheless preserves the addition law of points. In symbols, suppose that σ is an element of the Galois group and P a point in $E[m]$: Then just as $mP = O$, so $m(\sigma P) = O$.

The permutations induced by the elements of the Galois group can be represented by 2×2 matrices whose entries are integers modulo m. The transformation arising from σ is thus a matrix $\rho(\sigma)$. It is possible to view the points of $E[m]$ as column vectors of integers modulo m in such a way that the permutation induced by σ becomes matrix multiplication by $\rho(\sigma)$. One says that the matrices form a representation of the Galois group. It is noteworthy that the representation preserves the composition law of the Galois group; if σ and τ are elements of the group that combine to form a transformation υ, then $\rho(\upsilon)$ is just the product of the matrices $\rho(\sigma)$ and $\rho(\tau)$.

Let us pause to look back over our path so far. We began with the elliptic curve E, defined by an equation in which the numbers A and B are supposed to come from a counterexample to Fermat's last theorem. We turned then to the discrete set of points $E[m]$, defined to include just those points whose order divides the integer m. We examined how a certain Galois group acts on $E[m]$, and in particular looked at the 2×2 matrices that form a representation of this group. Now at last we can make a connection with the question of modularity. It turns out that the group of transformations we have reached through this tortuous argument supplies information on the elements of the L-series for the curve E. The existence of a formula for generating this series is one way of defining modularity.

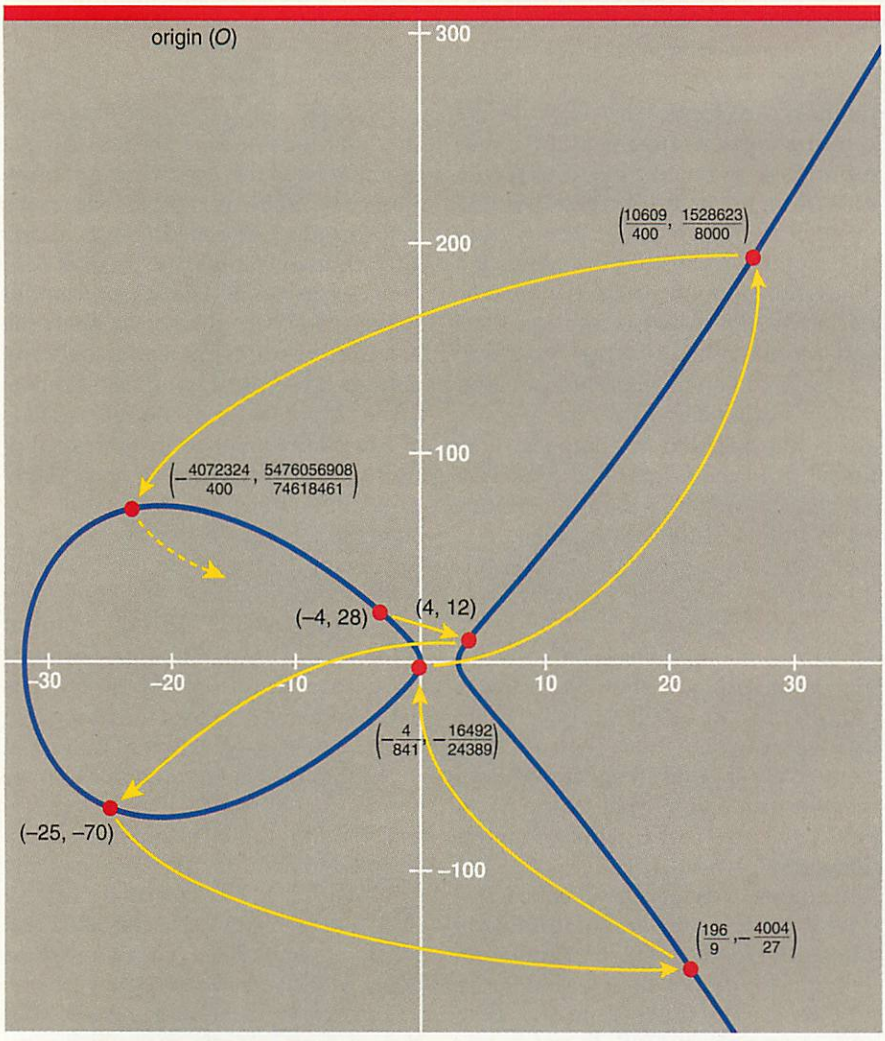


Figure 9. Some points of certain elliptic curves have infinite order: They can be added an unlimited number of times without ever reaching the origin. On the curve defined by $y^2 = x(x - 3)(x + 32)$ all rational points except the three points on the x axis—(0, −32), (0, 0) and (0, 3)—are of infinite order. The diagram shows the first few points in the sequence of points generated by repeated addition of (−4, 28). The first sum is (4, 12); then adding (−4, 28) to this point yields (−25, −70). The progression of sums can be continued indefinitely without ever reaching the origin and thereby entering a cycle.

The connection between the L-series and the Galois group comes about as follows. As noted above, the coefficients of the L-series are calculated by reducing the curve E modulo p for various primes p. Each p yields one term in the series; specifically, the L-series coefficient a_p is equal to the difference between $1 + p$ and the number of rational points on E reduced modulo p. But now the set $E[m]$ of m-division points provides another way of interpreting a_p, at least for most prime numbers p. Specifically, for each p there is a distinguished element σ_p of the Galois group and a corresponding matrix $\rho(\sigma_p)$ that depends on m and p. The sum of the two diagonal elements of this matrix is a number modulo m that is congruent to the integer a_p.

The ability to recapture a_p modulo m via Galois theory is the starting point of Ribet's proof that Frey's elliptic curve E cannot be modular. If you adopt the hypothesis that E is indeed modular, then the unusual properties of $E[n]$ mentioned above allow you to find a nonzero modular form of level 2 that is related modulo n to the form associated with E. But there is no such form of level 2, contradicting the original supposition that E is modular. In this way the argument reaches its final conclusion: If there is a counterexample to Fermat's last theorem, then there must be at least one elliptic curve that is not modular, contrary to the Taniyama-Shimura conjecture.

Proving *E* Is Modular

Wiles has said that he began work on a proof of the Taniyama-Shimura conjecture as soon as he learned that it would imply a proof of Fermat's last theorem. His campaign to solve the problem was to last seven years—and perhaps it has not ended yet.

Wiles's proof does not quite encompass the full Taniyama-Shimura conjecture; it excludes certain cases. When a curve is reduced modulo p, it is possible for all three roots to coalesce into a single numerical value. An example is the equation $y^2 = x(x - 10)(x + 15)$, where the three roots 0, 10 and −15 are all congruent to zero modulo 5. Wiles's proof does not apply to curves of this

kind. It is restricted to *semistable* elliptic curves, which are those with the property that whenever two of the roots coalesce modulo p, the third root remains distinct. For an equation of the form $y^2 = x(x - A)(x + B)$, the curve is semistable if no prime p divides both A and B. This condition clearly holds for the example $y^2 = x(x - 3)(x + 32)$. In fact it holds for any equation derived from a counterexample to Fermat's last theorem, because of the congruences modulo 4 and modulo 32 imposed on A and B. Hence the elliptic curve E derived from the counterexample must be semistable.

In setting out to prove that all semistable elliptic curves are modular, Wiles worked with the same mathematical tools employed in Ribet's proof, and many more in addition—including some that did not yet exist when Wiles began in 1986. Like Ribet, Wiles considers the set of points $E[m]$ for which $mP = O$, and the resulting representations of the Galois group. But in a crucial respect Wiles's task is a harder one. Where exhibiting a single counterexample was enough to complete Ribet's proof, Wiles must establish that $E[m]$ is modular for an infinite set of integers m.

Wiles's basic strategy is to study the family of sets $E[3]$, $E[9]$, $E[27]$, etc., or in other words the family defined by $E[m^\nu]$, where ν is any positive integer. There is a good reason for choosing this particular series: In about 1980 Robert P. Langlands of the Institute for Advanced Study and Jerrold B. Tunnell of Rutgers University proved that $E[3]$ itself is modular (Langlands 1980; Tunnell 1981). What the Langlands-Tunnell theorem means is that for any elliptic curve E, the set of points of order 3 forms a group that has an associated modular form. The trick is to extend that result to the entire family of sets $E[3^\nu]$.

Wiles effects this extension via arguments that hinge on representations of the Galois group. But there is a further complication, albeit a minor one. For the proof to work, the representation defined by the 3-division points of E must be irreducible, in the sense that it cannot be built out of smaller representations. Wiles sidesteps this problem through a clever tactic. He shows that if E is semistable, then either the representation coming from the 3-division points of E is irreducible, or else the representation arising from the 5-division points of E is irreducible. He then

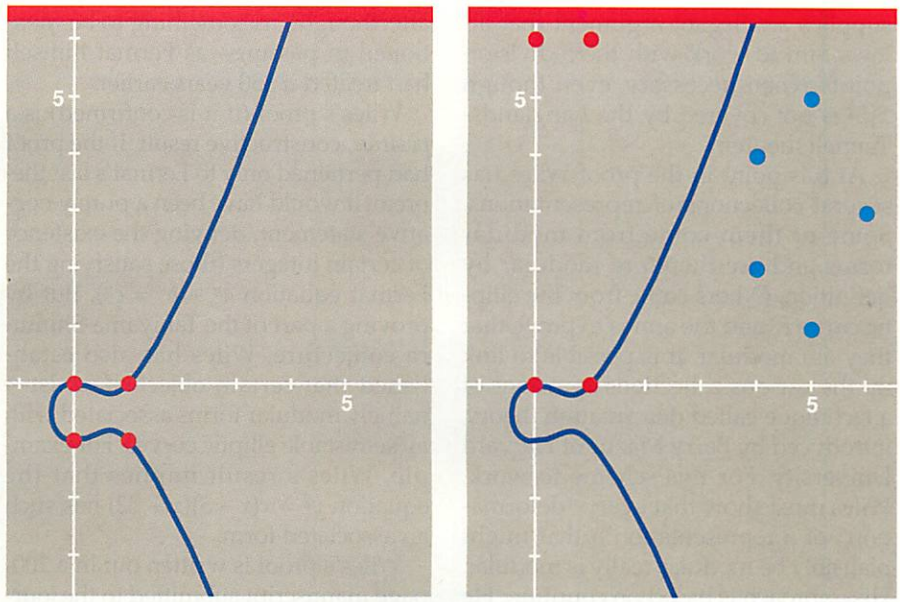


Figure 10. "Reducing" the equation of an elliptic curve modulo a prime number p yields a new set of points that can be regarded as solutions of the equation modulo p. For the equation $y^2 + y = x^3 + x^2$ there are only five rational solutions: (0, 0), (1, 0), (0, –1), (1, –1) and the origin *(left diagram)*. When the equation is reduced modulo 7 *(right diagram)*, five more points satisfy the resulting congruence *(blue dots)*. The original five points also remain solutions *(red dots)*, although (0, –1) and (1, –1) are transformed into (0, 6) and (1, 6) by the modulo operation. Reduction modulo p is allowed only if the equation does not become singular.

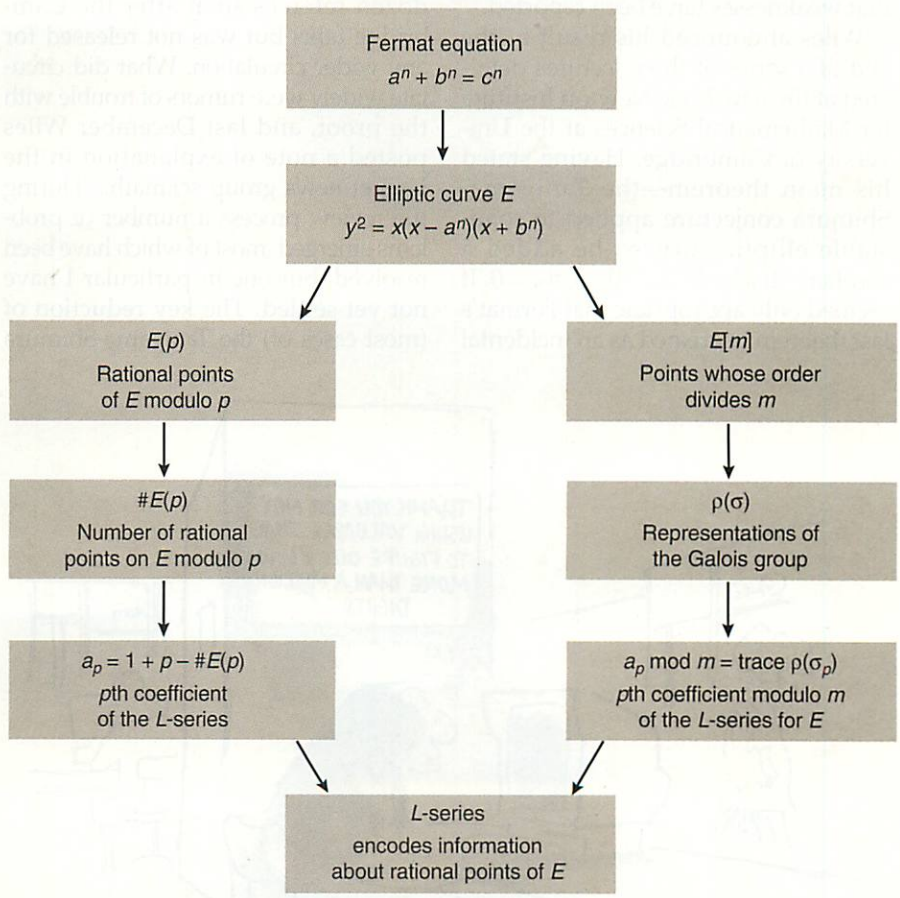


Figure 11. L-series is a mathematical object that has a crucial role in the proof of Fermat's last theorem. One route to calculating the L-series *(left branch of diagram)* is to reduce the elliptic curve E modulo p for many primes p; a formula relates the number of rational points on E modulo p to the pth coefficient of the L-series. The other route *(right branch of diagram)* examines the set $E[m]$ of points whose order divides m for various integers m; a group of transformations that act on this set yields information modulo m about the coefficients of the L-series.

supplies an elegant argument that allows him to work with the 5-division points when necessary, even though $E[5]$ is not covered by the Langlands-Tunnell theorem.

At this point in the proof Wiles has several collections of representations. Some of them come from modular forms and are therefore modular by definition. Others come from the elliptic curve E, and the aim is to prove that they are modular. It is possible to link up the various collections by means of a technique called deformation theory, introduced by Barry Mazur of Harvard University. For this scheme to work, Wiles must show that every "deformation" of a representation ρ that might plausibly be modular really is modular. His approach is based on counting: He endeavors to show that there are no more deformations than there are modular forms. This is the most difficult and technical part of the proof. It entails calculating an upper bound on the size of an object called a Selmer group. And it is in this part of the proof that weaknesses have been reported.

Wiles announced his result at the end of a series of three lectures delivered at the new Isaac Newton Institute for Mathematical Sciences at the University of Cambridge. Having stated his main theorem—the Taniyama-Shimura conjecture applied to semistable elliptic curves—he added a corollary: If $a^n + b^n = c^n$, then $abc = 0$. It seemed only appropriate that Fermat's last theorem be treated as an incidental afterthought, as something to be mentioned in passing—as Fermat himself had treated it 350 years earlier.

Wiles's proof (if it is confirmed) is a *positive*, constructive result. If the proof had pertained only to Fermat's last theorem, it would have been a purely negative statement, denying the existence of certain integers (those satisfying the Fermat equation $a^n + b^n = c^n$). But by proving a part of the Taniyama-Shimura conjecture, Wiles has also established that certain objects *do* exist—namely modular forms associated with all semistable elliptic curves. For example, Wiles's result implies that the equation $y^2 = x(x - 3)(x + 32)$ has such an associated form.

Wiles's proof is written out in a 200-page manuscript submitted to the journal *Inventiones Mathematicae*. The paper has five chapters, each of which has ample content to stand up as a journal article on its own. Most of the major results in arithmetic algebraic geometry of the past 25 years are cited.

The manuscript was sent to half a dozen referees soon after the Cambridge talks but was not released for any wider circulation. What did circulate widely were rumors of trouble with the proof, and last December Wiles posted a note of explanation in the Usenet news group sci.math: "During the review process a number of problems emerged, most of which have been resolved, but one in particular I have not yet settled. The key reduction of (most cases of) the Taniyama-Shimura conjecture to the calculation of the Selmer group is correct. However the final calculation of a precise upper bound for the Selmer group... is not yet complete as it stands. I believe that I will be able to finish this in the near future using the ideas explained in my Cambridge lectures." In February, when Wiles began a series of lectures at Princeton, the gap had not been filled.

The setback is naturally disappointing, but there is every reason to share in Wiles's optimism. Moreover, even in the worst-case scenario, if the flaw cannot be repaired, the line of inquiry is surely not exhausted. There is more yet to come from modern arithmetic.

Bibliography

Bell, E. T. 1961, 1990. *The Last Problem*. Introduction and notes by Underwood Dudley. Washington, D.C.: Mathematical Association of America.

Buhler, J., R. Crandall, R. Ernvall and T. Metsänkylä. 1993. Irregular primes and cyclotomic invariants to four million. *Mathematics of Computation* 61:151–153.

Cox, David A. 1994. Introduction to Fermat's last theorem. *American Mathematical Monthly* 101(1):3–14.

Edwards, Harold M. 1977. *Fermat's Last Theorem: A Genetic Introduction to Algebraic Number Theory*. New York: Springer-Verlag.

Husemöller, Dale. 1987. *Elliptic Curves*. With an appendix by Ruth Lawrence. New York: Springer-Verlag.

Knapp, Anthony W. 1992. *Elliptic Curves*. Princeton: Princeton University Press.

Langlands, R. P. 1980. *Base Change for GL(2)*. Annals of Mathematics Studies Vol. 96. Princeton: Princeton University Press.

Ribenboim, Paulo. 1979. *13 Lectures on Fermat's Last Theorem*. New York: Springer-Verlag.

Ribet, Kenneth A. 1990a. On modular representations of Gal($\overline{\mathbf{Q}}/\mathbf{Q}$) arising from modular forms. *Inventiones Mathematicae* 100:431–476.

Ribet, Kenneth A. 1990b. From the Taniyama-Shimura conjecture to Fermat's last theorem. *Annales de la Faculté des Sciences de l'Université de Toulouse* 11:116–139.

Ribet, Kenneth A. 1993. *Modular Elliptic Curves and Fermat's Last Theorem*. Selected Lectures in Mathematics. Providence, R.I.: American Mathematical Society. Video recording.

Rubin, Karl, and Alice Silverberg. 1994. A report on Wiles' Cambridge lectures. *Bulletin of the American Mathematical Society*, to appear.

Silverman, Joseph H. 1986. *The Arithmetic of Elliptic Curves*. New York: Springer-Verlag.

Silverman, Joseph H., and John H. Tate. 1989, 1992. *Rational Points on Elliptic Curves*. New York: Springer-Verlag.

Silverman, Joseph H., and Paul van Mulbregt. 1992. Elliptic curve calculator: A collection of Mathematica routines to perform calculations on elliptic curves defined over the rational numbers. Available from the Internet mathematics archive at wuarchive.wustl.edu.

Tunnell, J. 1981. Artin's conjecture for representations of octahedral type. *Bulletin of the American Mathematical Society* 5:173–175.

Errata and Addendum

In the article "Fermat's Last Theorem and Modern Arithmetic" (March–April) certain constraints imposed on the equation $A + B = C$ were expressed incorrectly. On page 148 the article states that $A = 3$ modulo 4, $C = 1$ modulo 4 and B is divisible by 32. These conditions are incompatible. They should have been applied to the symmetrical variant of the equation, $A + B + C = 0$, in which C is replaced by $-C$. (The replacement is permissible because C represents c^n in Fermat's equation, and n is odd.) One of the variables is divisible by 32; since all three variables play identical roles in the equation, no generality is lost by supposing that this variable is B. Exactly one of the remaining variables is 3 modulo 4, and one can suppose that this variable is A. The remaining variable, C, is 1 modulo 4.

In the article "The Predatory Behavior of the White Shark" (March–April) the number of white-shark attacks analyzed on videotape recordings should have been 129.

The authors of the article "Behaviorism, Congitivism and the Neuropsychology of Memory" (January–February), Herbert L. Petri and Mortimer Mishkin, wish to credit the seminal work on instrumental conditioning by Ruth W. Colwill (see especially material in *The Psychology of Learning and Motivation*, Academic Press, 1986).

Am. Sci. Vol 82 No 3 May-June 1994 p. 205