# Fermat's Last Stand

*His most notorious theorem baffled the greatest minds for more than three centuries. But after 10 years of work, one mathematician cracked it*

by Simon Singh and Kenneth A. Ribet


PIERRE DE FERMAT, a 17th-century master of number theory, often wrote to other mathematicians, asking if they had the ingenuity to match his solutions. He devised his most famous challenge, his so-called last theorem, while studying *Arithmetica*, by Diophantus of Alexandria. Fermat asserted that there are no nontrivial solutions for the equation $a^n + b^n = c^n$, where $n$ represents any whole number greater than 2. In the margin of *Arithmetica*, Fermat jotted a comment that tormented three centuries of mathematicians: "I have a truly marvelous demonstration of this proposition, which this margin is too narrow to contain."

This past June, 500 mathematicians gathered in the Great Hall of Göttingen University in Germany to watch Andrew J. Wiles of Princeton University collect the prestigious Wolfskehl Prize. The reward—established in 1908 for whoever proved Pierre de Fermat's famed last theorem—was originally worth $2 million (in today's dollars). By the summer of 1997, hyperinflation and the devaluation of the mark had reduced it to a mere $50,000. But no one cared. For Wiles, proving Fermat's 17th-century conundrum had realized a childhood dream and ended a decade of intense effort. For the assembled guests, Wiles's proof promised to revolutionize the future of mathematics.

Indeed, to complete his 100-page calculation, Wiles needed to draw on and further develop many modern ideas in mathematics. In particular, he had to tackle the Shimura-Taniyama conjecture, an important 20th-century insight into both algebraic geometry and complex analysis. In doing so, Wiles forged a link between these major branches of mathematics. Henceforth, insights from either field are certain to inspire new results in the other. Moreover, now that this bridge has been built, other connections between distant mathematical realms may emerge.

## The Prince of Amateurs

Pierre de Fermat was born on August 20, 1601, in Beaumont-de-Lomagne, a small town in southwest France. He pursued a career in local government and the judiciary. To ensure impartiality, judges were discouraged from socializing, and so each evening Fermat would retreat to his study and concentrate on his hobby, mathematics. Although an amateur, Fermat was highly accomplished and was largely responsible for probability theory and the foundations of calculus. Isaac Newton, the father of modern calculus, stated that he had based his work on "Monsieur Fermat's method of drawing tangents."

Above all, Fermat was a master of number theory—the study of whole numbers and their relationships. He would often write to other mathematicians about his work on a particular problem and ask if they had the ingenuity to match his solution. These challenges, and the fact that he would never reveal his own calculations, caused others a great deal of frustration. René Descartes, perhaps most noted for inventing coordinate geometry, called Fermat a braggart, and the English mathematician John Wallis once referred to him as "that damned Frenchman."

Fermat penned his most famous challenge, his so-called last theorem, while studying the ancient Greek mathematical text *Arithmetica*, by Diophantus of Alexandria. The book discussed positive whole-number solutions to the equation $a^2 + b^2 = c^2$, Pythagoras's formula de-

GAMMA LIAISON; SLIM FILMS

ANDREW J. WILES of Princeton University proved Fermat's famed last theorem in 1994, after a decade of concentrated effort. To complete his 100-page calculation, Wiles needed to draw on and further develop many modern ideas in mathematics. In particular, he had to prove the Shimura-Taniyama conjecture for a subset of elliptic curves, objects described by cubic equations such as $y^2 = x^3 + ax^2 + bx + c$.

scribing the relation between the sides of a right triangle. This equation has infinitely many sets of integer solutions, such as $a = 3$, $b = 4$, $c = 5$, which are known as Pythagorean triples. Fermat took the formula one step further and concluded that there are no nontrivial solutions for a whole family of similar equations, $a^n + b^n = c^n$, where $n$ represents any whole number greater than 2.

It seems remarkable that although there are infinitely many Pythagorean triples, there are no Fermat triples. Even so, Fermat believed he could support his claim with a rigorous proof. In the margin of *Arithmetica*, the mischievous genius jotted a comment that taunted generations of mathematicians: "I have a truly marvelous demonstration of this proposition, which this margin is too narrow to contain." Fermat made many such infuriating notes, and after his death his son published an edition of *Arithmetica* that included these teases. All the theorems were proved, one by one, until only Fermat's last remained.

Numerous mathematicians battled the last theorem and failed. In 1742 Leonhard Euler, the greatest number theorist of the 18th century, became so frustrated by his inability to prove the last theorem that he asked a friend to search Fermat's house in case some vital scrap of paper was left behind. In the 19th century Sophie Germain—who, because of prejudice against women mathematicians, pursued her studies under the name of Monsieur Leblanc—made the first significant breakthrough. Germain proved a general theorem that went a long way toward solving Fermat's equation for values of $n$ that are prime numbers greater than 2 and for which $2n + 1$ is also prime. (Recall that a prime number is divisible only by 1 and itself.) But a complete proof for these exponents, or any others, remained out of her reach.

At the start of the 20th century Paul Wolfskehl, a German industrialist, bequeathed 100,000 marks to whoever could meet Fermat's challenge. According to some historians, Wolfskehl was at one time almost at the point of suicide, but he became so obsessed with trying to prove the last theorem that his death wish disappeared. In light of what had happened, Wolfskehl rewrote his will. The prize was his way of repaying a debt to the puzzle that saved his life.

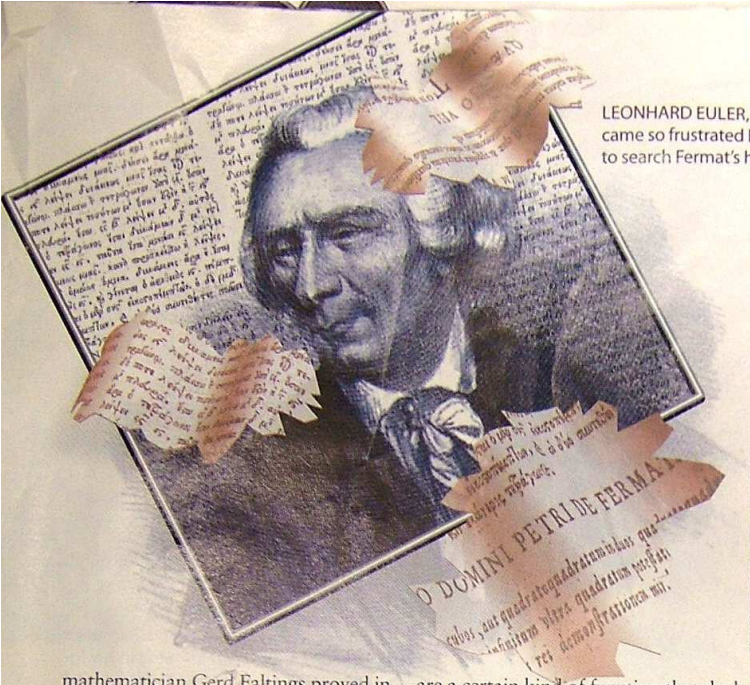
Ironically, just as the Wolfskehl Prize was encouraging enthusiastic amateurs to attempt a proof, professional mathematicians were losing hope. When the great German logician David Hilbert was asked why he never attempted a proof of Fermat's last theorem, he replied, "Before beginning I should have to put in three years of intensive study, and I haven't that much time to squander on a probable failure." The problem still held a special place in the hearts of number theorists, but they regarded Fermat's last theorem in the same way that chemists regarded alchemy. It was a foolish romantic dream from a past age.

## The Childhood Dream

Children, of course, love romantic dreams. And in 1963, at age 10, Wiles became enamored with Fermat's last theorem. He read about it in his local library in Cambridge, England, and promised himself that he would find a proof. His schoolteachers discouraged him from wasting time on the impossible. His college lecturers also tried to dissuade him. Eventually his graduate supervisor at the University of Cambridge steered him toward more mainstream mathematics, namely into the fruitful research area surrounding objects called elliptic curves. The ancient Greeks originally studied elliptic curves, and they appear in *Arithmetica*. Little did Wiles know that this training would lead him back to Fermat's last theorem.

Elliptic curves are not ellipses. Instead they are named as such because they are described by cubic equations, like those used for calculating the perimeter of an ellipse. In general, cubic equations for elliptical curves take the form $y^2 = x^3 + ax^2 + bx + c$, where $a$, $b$ and $c$ are whole numbers that satisfy some simple conditions. Such equations are said to be of degree 3, because the highest exponent they contain is a cube.

Number theorists regularly try to ascertain the number of so-called rational solutions, those that are whole numbers or fractions, for various equations. Linear or quadratic equations, of degree 1 and 2, respectively, have either no rational solutions or infinitely many, and it is simple to decide which is the case. For complicated equations, typically of degree 4 or higher, the number of solutions is always finite—a fact called Mordell's conjecture, which the German




LEONHARD EULER, the greatest number theorist of the 18th century, be came so frustrated by Fermat's last theorem that in 1742 he asked a friend to search Fermat's house for any scrap of paper left behind.

GRANGER COLLECTION; BROWN UNIVERSITY LIBRARY; SLIM FILMS

mathematician Gerd Faltings proved in 1983. But elliptic curves present a unique challenge. They may have a finite or infinite number of solutions, and there is no easy way of telling.

To simplify problems concerning elliptic curves, mathematicians often reexamine them using modular arithmetic. They divide $x$ and $y$ in the cubic equation by a prime number $p$ and keep only the remainder. This modified version of the equation is its "mod $p$" equivalent. Next, they repeat these divisions with another prime number, then another, and as they go, they note the number of solutions for each prime modulus. Eventually these calculations generate a series of simpler problems that are analogous to the original.

The great advantage of modular arithmetic is that the maximum values of $x$ and $y$ are effectively limited to $p$, and so the problem is reduced to something finite. To grasp some understanding of the original infinite problem, mathematicians observe how the number of solutions changes as $p$ varies. And using that information, they generate a so-called L-series for the elliptic curve. In essence, an L-series is an infinite series in powers, where the value of the coefficient for each $p$th power is determined by the number of solutions in modulo $p$.

In fact, other mathematical objects, called modular forms, also have L-series. Modular forms should not be confused with modular arithmetic. They are a certain kind of function that deals with complex numbers of the form $(x + iy)$, where $x$ and $y$ are real numbers, and $i$ is the imaginary number (equal to the square root of $-1$).

What makes modular forms special is that one can transform a complex number in many ways, and yet the function yields virtually the same result. In this respect, modular forms are quite remarkable. Trigonometric functions are similar inasmuch as an angle, $q$, can be transformed by adding $\pi$, and yet the answer is constant: $\sin q = \sin (q + \pi)$. This property is termed symmetry, and trigonometric functions display it to a limited extent. In contrast, modular forms exhibit an immense level of symmetry. So much so that when the French polymath Henri Poincaré discovered the first modular forms in the late 19th century, he struggled to come to terms with their symmetry. He described to his colleagues how every day for two weeks he would wake up and search for an error in his calculations. On the 15th day he finally gave up, accepting that modular forms are symmetrical in the extreme.

A decade or so before Wiles learned about Fermat, two young Japanese mathematicians, Goro Shimura and Yutaka Taniyama, developed an idea involving modular forms that would ultimately serve as a cornerstone in Wiles's proof. They believed that modular forms and elliptic curves were fundamentally related—even though elliptic curves apparently belonged to a totally different area of mathematics. In particular, because modular forms have an L-series—although derived by a different prescription than that for elliptic curves—the two men proposed that every elliptic curve could be paired with a modular form, such that the two L-series would match.

Shimura and Taniyama knew that if they were right, the consequences would be extraordinary. First, mathematicians generally know more about the L-series of a modular form than that of an elliptic curve. Hence, it would be unnecessary to compile the L-series for an elliptic curve, because it would be identical to that of the corresponding modular form. More generally, building such a bridge between two hitherto unrelated branches of mathematics could benefit both: potentially each discipline could become enriched by knowledge already gathered in the other.

The Shimura-Taniyama conjecture, as it was formulated by Shimura in the early 1960s, states that every elliptic curve can be paired with a modular form; in other words, all elliptic curves are modular. Even though no one could find a way to prove it, as the decades passed the hypothesis became increasingly influential. By the 1970s, for instance, mathematicians would often assume that the Shimura-Taniyama conjecture was true and then derive some new result from it. In due course, many major findings came to rely on the conjecture, although few scholars expected it would be proved in this century. Tragically, one of the men who inspired it did not live to see its ultimate importance. On November 17, 1958, Yutaka Taniyama committed suicide.

### The Missing Link

In the fall of 1984, at a symposium in Oberwolfach, Germany, Gerhard Frey of the University of Saarland gave a lecture that hinted at a new strategy for attacking Fermat's last theorem. The theorem asserts that Fermat's equation has no positive whole-number solutions. To test a statement of this type, mathematicians frequently assume that it is false and then explore the consequences. To

COURTESY OF PRINCETON UNIVERSITY; SLIM FILMS

...ORO SHIMURA AND YUTAKA TANIYAMA (*top* and *bottom*, respectively) developed an idea during the 1950s that ultimately served in Wiles's proof. Their conjecture involved modular forms—functions that deal with complex numbers of the form $(x + iy)$, where $x$ and $y$ are real numbers, and $i$ is the imaginary number (equal to the square root of $-1$). The two men proposed that every elliptic curve could be paired with a modular form, such that the L-series associated with each would match. Tragically, Taniyama did not live to see Wiles's success. On November 17, 1958, he killed himself.

say that Fermat's last theorem is false is to say that there are two perfect $n$th powers whose sum is a third $n$th power.

Frey's idea proceeded as follows: Suppose that $A$ and $B$ are perfect $n$th powers of two numbers such that $A + B$ is again an $n$th power—that is, they are a solution to Fermat's equation. $A$ and $B$ can then be used as coefficients in a special elliptic curve: $y^2 = x(x - A)(x + B)$. A quantity that is routinely calculated whenever one studies elliptic curves is the "discriminant" of the elliptic curve, $A^2B^2(A + B)^2$. Because $A$ and $B$ are solutions to the Fermat equation, the discriminant is a perfect $n$th power.

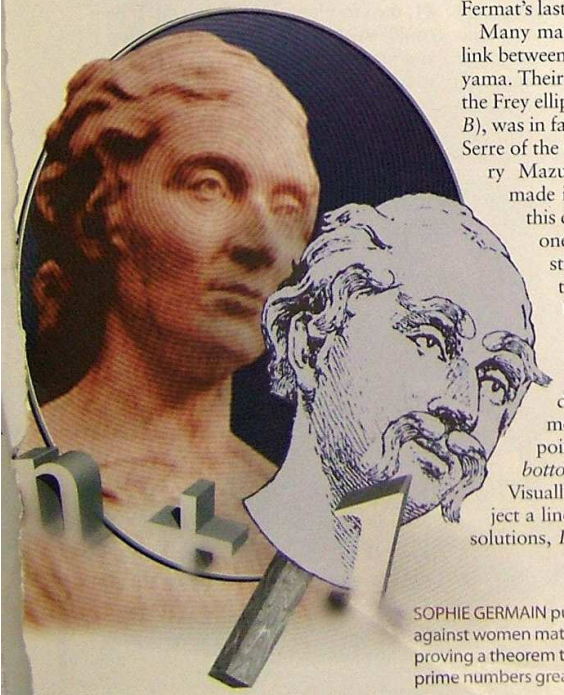
The crucial point in Frey's tactic is that if Fermat's last theorem is false, then whole-number solutions such as $A$ and $B$ can be used to construct an elliptic curve whose discriminant is a perfect $n$th power. So a proof that the discriminant of an elliptic curve can never be an $n$th power would contain, implicitly, a proof of Fermat's last theorem. Frey saw no way to construct that proof. He did, however, suspect that an elliptic curve whose discriminant was a perfect $n$th power—if it existed—could not be modular. In other words, such an elliptic curve would defy the Shimura-Taniyama conjecture. Running the argument backwards, Frey pointed out that if someone proved that the Shimura-Taniyama conjecture is true and that the elliptic equation $y^2 = x(x - A)(x + B)$ is not modular, then they would have shown that the elliptic equation cannot exist. In that case, the solution to Fermat's equation cannot exist, and Fermat's last theorem is proved true.

Many mathematicians explored this link between Fermat and Shimura-Taniyama. Their first goal was to show that the Frey elliptic curve, $y^2 = x(x - A)(x + B)$, was in fact not modular. Jean-Pierre Serre of the College of France and Barry Mazur of Harvard University made important contributions in this direction. And in June 1986 one of us (Ribet) at last constructed a complete proof of the assertion. It is not possible to describe the full argument in this article, but we will give a few hints.

To begin, Ribet's proof depends on a geometric method for "adding" two points on an elliptic curve [*see bottom illustration on next page*]. Visually, the idea is that if you project a line through a pair of distinct solutions, $P_1$ and $P_2$, the line cuts the curve at a third point, which we might provisionally call the sum of $P_1$ and $P_2$. A slightly more complicated but more valuable version of this addition is as follows: first add two points and derive a new point, $P_3$, as already described, and then reflect this point through the $x$ axis to get the final sum, $Q$.

This special form of addition can be applied to any pair of points within the infinite set of all points on an elliptic curve, but this operation is particularly interesting because there are finite sets of points having the crucial property that the sum of any two points in the set is again in the set. These finite sets of points form a group: a set of points that obeys a handful of simple axioms. It turns out that if the elliptic curve is modular, so are the points in each finite group of the elliptic curve. What Ribet proved is that a specific finite group of Frey's curve cannot be modular, ruling out the modularity of the whole curve.

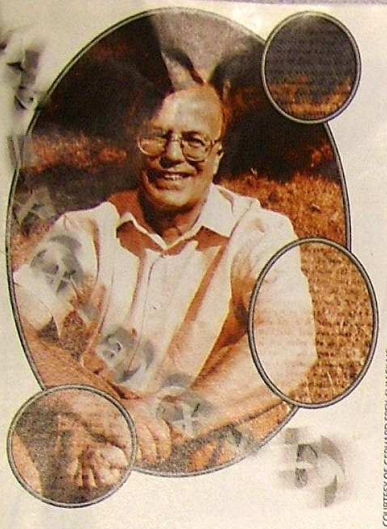
For three and half centuries, the last theorem had been an isolated problem, a curious and impossible riddle on the edge of mathematics. In 1986 Ribet, building on Frey's work, had brought it



STEVE MURREZ *Black Star*; SLIM FILMS

SOPHIE GERMAIN pursued her studies under the name of Monsieur Leblanc because of prejudice against women mathematicians. She made the first significant breakthrough in the 19th century, proving a theorem that went a long way toward solving Fermat's equation for values of $n$ that are prime numbers greater than 2 and for which $2n + 1$ is also prime.

*Fermat's Last Stand*



COURTESY OF GERHARD FREY; SLIM FILMS

GERHARD FREY suggested a new strategy for attacking Fermat's last theorem in 1984: Suppose that $A$ and $B$ are perfect $n$th powers such that $A + B$ is again an $n$th power—that is, they are a solution to Fermat's equation. $A$ and $B$ can then be used as coefficients in a special elliptic curve: $y^2 = x(x - A)(x + B)$; the "discriminant" of this elliptic curve, $A^2B^2(A + B)^2$, is also a perfect $n$th power. Frey suspected that such an elliptic curve could not be modular. In other words, Frey pointed out that if someone proved that the Shimura-Taniyama conjecture is true or that all elliptic curves are modular, then they might be able to show that the elliptic equation $y^2 = x(x - A)(x + B)$ cannot exist—in which case, the solution to Fermat's equation cannot exist, and Fermat's last theorem is proved true.

center stage. It was possible to prove Fermat's last theorem by proving the Shimura-Taniyama conjecture. Wiles, who was by now a professor at Princeton, wasted no time. For seven years, he worked in complete secrecy. Not only did he want to avoid the pressure of public attention, but he hoped to keep others from copying his ideas. During this period, only his wife learned of his obsession—on their honeymoon.

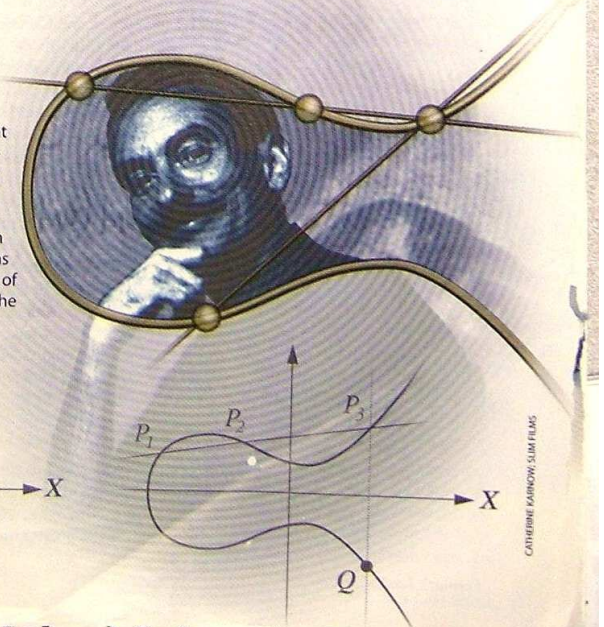
### Seven Years of Secrecy

Wiles had to pull together many of the major findings of 20th-century number theory. When those ideas were inadequate, he was forced to create other tools and techniques. He describes his experience of doing mathematics as a journey through a dark, unexplored mansion: "You enter the first room of the mansion, and it's completely dark. You stumble around bumping into the furniture, but gradually you learn where each piece of furniture is. Finally, after six months or so, you find the light switch. You turn it on, and suddenly it's all illuminated. You can see exactly where you were. Then you move into the next room and spend another six months in the dark. So each of these breakthroughs, while sometimes they're momentary, sometimes over a period of a day or two, they are the culmination of, and couldn't exist without, the many months of stumbling around in the dark that precede them."

As it turned out, Wiles did not have to prove the full Shimura-Taniyama conjecture. Instead he had to show only that a particular subset of elliptic curves—one that would include the hypothetical elliptic curve Frey proposed, should it exist—is modular. It wasn't really much of a simplification. This subset is still infinite in size and includes the majority of interesting cases. Wiles's strategy used the same techniques employed by Ribet, plus many more. And as with Ribet's argument, it is possible to give only a hint of the main points involved.

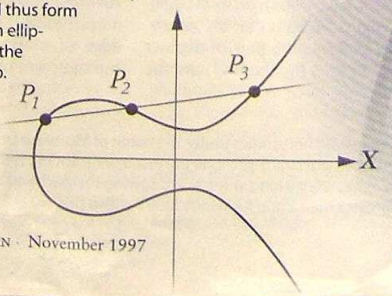
The difficulty was to show that every elliptic curve in Wiles's subset is modular. To do so, Wiles exploited the group property of points on the elliptic curves and applied a theorem of Robert P. Langlands of the Institute for Advanced Study in Princeton, N.J., and Jerrold Tunnell of Rutgers University. The theorem shows, for each elliptic curve in Wiles's set, that a specific group of points inside the elliptic curve is modular. This requirement is necessary but not sufficient to demonstrate that the elliptic curve as a whole is modular.

The group in question has only nine elements, so one might imagine that its modularity represents an extremely small first step toward complete modularity. To close this gap, Wiles wanted to examine increasingly larger groups, stepping from groups of size 9 to $9^2$, or 81, then to $9^3$, or 729, and so on. If he could reach an infinitely large group and prove that it, too, is modular, that would be equivalent to proving that the entire curve is modular.



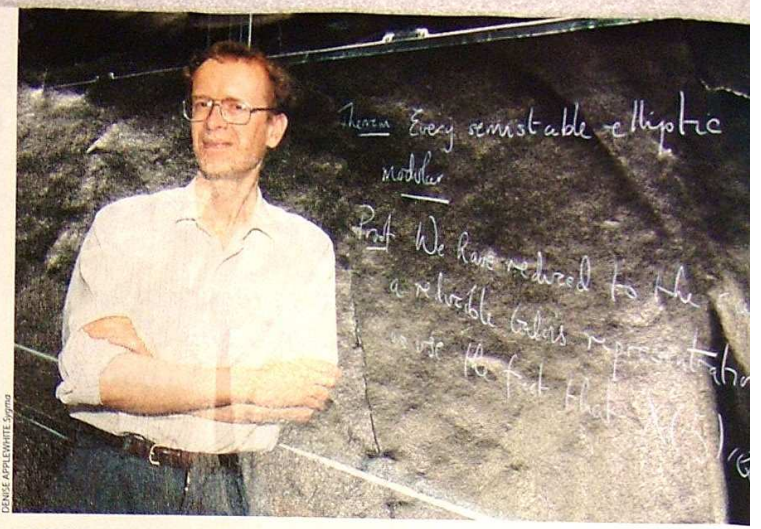
CATHERINE KARNOW; SLIM FILMS

KENNETH A. RIBET followed Frey's lead and in June 1986 proved that any elliptic curve could not be modular if its discriminant were a perfect $n$th power. Ribet's proof depends on a geometric method for "adding" points on an elliptic curve. Visually the idea is that it is possible to project a line through a pair of points on the elliptic curve, $P_1$ and $P_2$, to obtain a third point, $P_3$. This new point is then reflected in the $x$ axis to obtain $Q$, which is said to be the sum of $P_1$ and $P_2$. Whereas the set of all points on an elliptic curve is infinite, there are finite sets of points having the crucial property that the sum of any two points in the set is again in the set. Such finite sets obey certain special axioms and thus form so-called finite groups. If an elliptic curve is modular, so are the points in each finite group. Ribet proved that a specific finite group of Frey's curve cannot be modular, ruling out the modularity of the whole curve.



Wiles accomplished this task via a process loosely based on induction. He had to show that if one group was modular, then so must be the next larger group. This approach is similar to toppling dominoes: to knock down an infinite number of dominoes, one merely has to ensure that knocking down any one domino will always topple the next. Eventually Wiles felt confident that his proof was complete, and on June 23, 1993, he announced his result at a conference at the Isaac Newton Mathematical Sciences Institute in Cambridge. His secret research program had been a success, and the mathematical community and the world's press were surprised and delighted by his proof. The front page of the *New York Times* exclaimed, "At Last, Shout of 'Eureka!' in Age-Old Math Mystery."

As the media circus intensified, the official peer-review process began. Almost immediately, Nicholas M. Katz of Princeton uncovered a fundamental and devastating flaw in one stage of Wiles's argument. In his induction process, Wiles had borrowed a method from Victor A. Kolyvagin of Johns Hopkins University and Matthias Flach of the California Institute of Technology to show that the group is modular. But it now seemed that this method could not be relied on in this particular instance. Wiles's childhood dream had turned into a nightmare.




DENISE APPLEWHITE *Sygma*

"EUREKA!" read a *New York Times* headline after Wiles revealed his first proof of Fermat's last theorem at a lecture in June 1993. Soon thereafter, though, reviewers found a serious flaw. Wiles discussed the error only with his former student Richard Taylor. Together they tried to patch up the method Wiles had used and applied tools that he had previously rejected. At last, on September 19, 1994, they found the vital fix.

### Finding the Fix

For the next 14 months, Wiles hid himself away, discussing the error only with his former student Richard Taylor. Together they wrestled with the problem, trying to patch up the method Wiles had already used and applying other tools that he had previously rejected. They were at the point of admitting defeat and releasing the flawed proof so that others could try to correct it, when, on September 19, 1994, they found the vital fix. Many years earlier Wiles had considered using an alternative approach based on so-called Iwasawa theory, but it floundered, and he abandoned it. Now he realized that what was causing the Kolyvagin-Flach method to fail was exactly what would make the Iwasawa theory approach succeed.

Wiles recalls his reaction to the discovery: "It was so indescribably beautiful; it was so simple and so elegant. The first night I went back home and slept on it. I checked through it again the next morning, and I went down and told my wife, 'I've got it. I think I've found it.' And it was so unexpected that she thought I was talking about a children's toy or something, and she said, 'Got what?' I said, 'I've fixed my proof. I've got it.'"

For Wiles, the award of the Wolfskehl Prize marks the end of an obsession that lasted more than 30 years: "Having solved this problem, there's certainly a sense of freedom. I was so obsessed by this problem that for eight years I was thinking about it all of the time—when I woke up in the morning to when I went to sleep at night. That particular odyssey is now over. My mind is at rest." For other mathematicians, though, major questions remain. In particular, all agree that Wiles's proof is far too complicated and modern to be the one that Fermat had in mind when he wrote his marginal note. Either Fermat was mistaken, and his proof, if it existed, was flawed, or a simple and cunning proof awaits discovery.


### *The Authors*

SIMON SINGH and KENNETH A. RIBET share a keen interest in Fermat's last theorem. Singh is a particle physicist turned television science journalist, who wrote *Fermat's Enigma* and co-produced a documentary on the subject. Ribet is a professor of mathematics at the University of California, Berkeley, where his work focuses on number theory and arithmetic algebraic geometry. For his proof that the Shimura-Taniyama conjecture implies Fermat's last theorem, Ribet and his colleague Abbas Bahri won the first Prix Fermat.


### *Further Reading*


YUTAKA TANIYAMA AND HIS TIME: VERY PERSONAL RECOLLECTIONS FROM SHIMURA. Goro Shimura in *Bulletin of the London Mathematical Society*, Vol. 21, pages 186–196; 1989.

FROM THE TANIYAMA-SHIMURA CONJECTURE TO FERMAT'S LAST THEOREM. Kenneth A. Ribet in *Annales de la Faculté des Sciences de L'Université de Toulouse*, Vol. 11, No. 1, pages 115–139; 1990.

MODULAR ELLIPTIC CURVES AND FERMAT'S LAST THEOREM. Andrew Wiles in *Annals of Mathematics*, Vol. 141, No. 3, pages 443–551; May 1995.

RING THEORETIC PROPERTIES OF CERTAIN HECKE ALGEBRAS. Richard Taylor and Andrew Wiles in *Annals of Mathematics*, Vol. 141, No. 3, pages 553–572; May 1995.

NOTES ON FERMAT'S LAST THEOREM. A. J. van der Poorten. Wiley Interscience, 1996.

FERMAT'S ENIGMA. Simon Singh. Walker and Company, 1997.