from *SIAM News*, Volume 31, Number 3

Fermat's Last Theorem: Still Awaiting a First-Rate Exposition

Fermat's Last Theorem. *By Amir D. Aczel, Four Walls Eight Windows, New York, 1996, 200 pages, $18.00.*
Fermat's Enigma. *By Simon Singh, Walker and Company, New York, 1997, 288 pages, $22.00.*
Notes on Fermat's Last Theorem. *By Alf van der Poorten, John Wiley, New York, 1996, 240 pages, $49.95.*

BOOK REVIEW

By James Case

The good news concerning Andrew Wiles's proof of Fermat's last theorem (FLT) is that a first-rate exposition of it will eventually be written, one comparable to Gleick's *Chaos* [2] or *Gödel's Proof* [3] by Nagel and Newman. The bad news is that none of the books under review is in that league.

Although addressed to "undergraduates," van der Poorten's notes seem too technical to hold the interest of any save the most advanced and determined members of even that select audience. The few who persevere will learn what an elliptic curve is, how to perform arithmetic on one, and how to navigate among modular functions, Galois representations, class number formulae, and other structures essential to Wiles's proof. The only serious omission is "horizontal Iwasawa theory," which did not figure in the (flawed) original proof presented by Wiles in the summer of 1993. The topic has been omitted presumably because the notes in question were written during the 15 months between the appearance of the original and corrected versions of Wiles's proof. A special appendix (Appendix A) is provided for "those who only want to pretend to have read the rest of the book."

Singh's book, the longer of the remaining two, is an outgrowth of the author's involvement with the BBC's award-winning documentary on FLT. That program evolved into *The Proof*, an hour-long made-for-TV special that aired several times in October 1997, as part of PBS's often fascinating Nova series. Both the book and the TV special were careful to acknowledge the giants on whose shoulders Wiles stood to achieve his result. Peter Sarnak, John Conway, Barry Mazur, Ken Ribet, Nick Katz, John Coates, Richard Taylor, and Goro Shimura were thus on screen almost as much as Wiles himself, his favorite Lake Carnegie walk, and his (exceedingly neat) attic study.

Back to the Roots: The Notion of Mathematical Proof

The drama begins with Wiles's 1963 encounter with E.T. Bell's *The Last Problem* [1]. Only ten years old at the time, the precocious youth found the book on the shelves of his local library, borrowed it, read it, succumbed to the problem's appeal, and resolved to attack it himself one day. By the time he entered graduate school—at Cambridge, in the autumn of 1975—he had mastered 19th-century mathematics, and concluded that more powerful tools were needed. He was fortunate enough to be taken on as a research student by Coates, who pointed him toward elliptic curves and horizontal Iwasawa theory, the anvils upon which Wiles was to forge his early reputation.

Both Singh and Aczel begin not with Fermat himself, nor with the revival of mathematical discovery during the Renaissance, nor even with Diophantus, who initiated the search for integer solutions of simple algebraic equations in ancient times. Rather, they trace mathematics—particularly the notion of mathematical proof—back to its earliest roots among the Egyptians, Babylonians, and pre-Socratic Greeks. As a result, the early chapters of both books constitute selective histories of mathematics in the ancient and medieval worlds. Such an approach permits the authors to remind nontechnical readers of long-forgotten facts in an unobtrusive manner, while retaining the interest of more experienced folk with entertaining anecdotes.

The drama of Andrew Wiles's 1993 proof, says the author, begins with his 1963 encounter, at the age of ten, with E.T. Bell's The Last Problem.

Both authors then describe the basic facts of Fermat's life as a prominent jurist during the reign of Louis XIV, though in nothing like the detail supplied by Bell in *The Last Problem*. Each speculates at some length concerning the proof Fermat claimed to have discovered in or about 1637, but was unable to fit into the margin of his personal copy of Diophantus's *Arithmetica*. Singh points out that other copies of the original (1621) Latin translation, still to be found in various Cambridge libraries, are notable for the width of their margins.

The fact that Fermat never again mentioned his "marvelous demonstration," even in the course of outlining a valid proof for the special case $n = 4$, suggests that the lost proof was fallacious and that Fermat soon discovered his own error. This reviewer, for one, cannot agree with Singh that, ". . . there are plenty of mathematicians who believe that they can

still achieve fame and glory by discovering Fermat's original proof."

The Legacy of the "Marvelous Demonstration"

Fermat's proof for the case of $n = 4$ made use of his "method of descent," which deduced a contradiction from the fact that if a, b, c is a solution in positive integers of the equation $x^4 + y^4 = z^4$, then there exists a second such solution, a', b', c', for which $c' < c$. It was not until Euler took up the problem more than a century later that further progress was made. Euler began by observing that, by means of (then still suspect) imaginary numbers, the method of descent could be brought to bear on the case of $n = 3$ as well. Thus, more than a hundred years after Fermat's discovery, the equation $x^n + y^n = z^n$ was known to be insoluble in positive integers only for n divisible by either 3 or 4.

By the early years of the 19th century, FLT had become firmly ensconced as number theory's most notorious puzzle. It was then that a young French woman named Sophie Germain made an important contribution by proving that if a, b, c were a solution of Fermat's equation for $n = 5$, then $a, b,$ and c would all be divisible by 5. Cognizant that it would suffice to prove FLT for prime exponents $n > 2$, she then focused her attention on those primes p (like 5 and 11, but unlike 7 and 13) for which $2p + 1$ is again a prime, showing that if $a, b,$ and c constitute a solution of Fermat's equation for any such prime exponent, then $a, b,$ and c must all be divisible by that same prime. Her theorem allowed Legendre and Dirichlet—working independently—to dispose of the case of $n = 5$ in 1825, and Lamé to eliminate $n = 7$ a few years later. But no general proof was forthcoming. Both Aczel and Singh hasten to point out that, although Germain employed the pseudonym A. Le Blanc in her early correspondence with both Gauss and Lagrange, each was delighted to learn her secret and hastened to offer continuing friendship and support.

Encouraged by these developments, the French Academy of Sciences offered an array of prizes, including a gold medal and 3000 francs, to the mathematician who should be first to resolve the mystery of Fermat's last theorem. It was thus a dramatic moment when Gabriel Lamé rose, at the March 1, 1847, meeting of the Academy, to outline the technique that seemed to have brought him to the verge of proving FLT. Lamé's technique involved the rewriting of Fermat's equation in the form $y^n = z^n - x^n$ and expressing the right-hand side as a product of linear factors involving the complex nth roots of unity. The plot thickened when Cauchy rose from the audience and announced that he was working along similar lines and expected to announce a complete proof in the near future.

But the hopes of the assembled savants were dashed on May 24, when Liouville read a letter from Kummer in Berlin to the effect that the proposed technique was doomed to fail for prime values of n such that the key factorization was nonunique. He identified $n = 37$, 59, and 67 as the only two-digit primes for which the attendant difficulty was unavoidable but established that such "irregular primes" are infinite in number. In 1856, the Academy withdrew the question from competition and awarded the medal (if not the cash) to Kummer for "his beautiful researches on the complex numbers composed of roots of unity and integers."

A new prize was established in 1908, through a bequest from a wealthy German industrialist named Paul Wolfskehl. Having studied mathematics as a young man, and never having lost interest in the subject, he was keenly aware that progress toward a proof of FLT had all but ceased with Kummer's revelations. So it was that, on the very eve of his planned midnight suicide—the woman of his dreams had rejected him—he discovered a gap in Kummer's logic. By dawn he had filled the gap, regained his will to live, and resolved to revive interest in FLT by bequeathing 100,000 Marks to the first person to submit a satisfactory proof to the *Königliche Gesellschaft der Wissenschaften*, Göttingen. More than 600 unsatisfactory "proofs" were submitted in the first year of the competition, followed by thousands more in subsequent years. The flow continued even after hyperinflation in the Weimar Republic had rendered Wolfskehl's bequest valueless. Fortunately, new sources of funding were tapped in the years following World War II, so that the prize (awarded to Wiles on June 27, 1997) was again worth $50,000.

Takayama–Shimura

All these events and more are duly reported by Aczel and Singh. Both fall short, however, in documenting more recent developments, such as those involving the famed Takayama–Shimura conjecture. Wiles proved a restricted version of that conjecture—after a mere eight years of frequently intense effort—beginning soon after Ken Ribet demonstrated that FLT would follow from just such a proof.

The conjecture, as originally stated by Yutaka Taniyama at a Tokyo meeting in September 1955, asserted that any "elliptic curve" in the xy-plane could be uniformized by "modular functions" f and g, much as the ellipse $(x/a - 1)^2 + (y/b)^2 = 1$ is uniformized (parameterized) by the rational functions $x = f(t) = 2ab(b^2 + t^2a^2)$ and $y = g(t) = 2abt(b^2 + t^2a^2)$. Later, in conversation with André Weil, Taniyama was overheard to amend his assertion to include more general "automorphic functions" as well. But his good friend Goro Shimura later concluded (some years after Taniyama's suicide, in 1958) that the additional generality was superfluous—if the statement were true at all, modular functions alone would suffice.

Both Aczel and Singh begin their discussions of this all-important topic by defining an elliptic curve to be the graph of an equation of the form

$$y^2 = ax^3 + bx^2 + cx + d, \tag{1}$$

a, b, c, d integers, and exhibiting a few examples. Aczel chooses the normalization $d = 0$, while Singh selects $a = 1$. But both immediately muddy the waters by discussing elliptic curves of the form $y^2 + y = ax^3 + bx^2 + cx + d$, which are not of the

given form. This would be a minor transgression in a book written for an audience of trained mathematicians but could easily discourage beginners like the youthful Wiles, presumably the most important members of the target audience.

Worse still are the two authors' attempts to explain modular functions. It may be true, as John Conway conceded on TV, that modular functions cannot be explained in a single sentence. But they can be explained in a single page, and should be in any coherent discussion of the Taniyama–Shimura conjecture. One way is to observe that mathematics relies heavily on certain "special functions," which arise over and over again, in a variety of seemingly unrelated contexts, and that many such functions obey simple "functional equations," such as

$$\begin{aligned} &f(xy) = f(x) + f(y), \\ &f(x + y) = f(x)f(y), \\ &\text{and} \\ &f(x + 1) = xf(x). \end{aligned} \tag{2}$$

Moreover, if g is any function defined on the extended complex plane, then

$$\begin{aligned} &f(z) = g(z) + g(1 - z) + g(1/z) + \\ &g(1/(1 - z)) + g(z/(z - 1)) + g((z - 1)/z) \end{aligned} \tag{3}$$

satisfies all six of the functional equations

$$f(z) = f((az + b)/(cz + d)) \tag{4}$$

obtained by allowing the associated matrix $A=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ to range over the group

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}. \tag{5}$$

So it need come as no great surprise that there exist certain "extra special" functions that satisfy the entire infinite family of equations (4) obtained by allowing A to vary over an infinite group G of matrices. Such functions are called "automorphic" if G is any discrete subgroup of the special linear group $SL_2(\mathbb{C})$, and "modular" if G is either the "modular subgroup" of $SL_2(\mathbb{C})$ for which a, b, c, and d are (real) integers such that $ad - bc = 1$, or any of the many subgroups thereof.

Inspiring Generations of Potential Mathematicians

Such explanations need to be fleshed out a bit for the novice reader, but can (with a little effort) be made comprehensible to a surprisingly wide audience. On the other hand, it serves no conceivable purpose to display equation (4) without a word of explanation (as does Aczel on page 94) or to exhibit a reproduction of the M.C. Escher painting *Circle Limit IV* (as does Singh on page 180) in the hope that "the interested reader" will somehow figure out what modular functions are and how Wiles exploited them. Even the most gifted of novices need more guidance than that. E.T. Bell, and others like him, have inspired generations of potential mathematicians—the ten-year-old Wiles among them—by convincing their readers (either rightly or wrongly) that they already understood the gist of the explanations offered and could in time master the details. The books by Aczel and Singh seem more likely to leave the opposite impression! The British edition of Singh's book is rumored to be more satisfactory in its treatment of matters mathematical.

These books can be recommended to the reader who wants to know who (besides Wiles himself) contributed to the proof of FLT, and is likely to be entertained by accounts of Fermat's years on the bench, Galois' fatal duel, Cauchy's legendary arrogance, Poincaré's discovery of automorphic functions, and the deeds of Gödel, Turing, von Neumann, and Paul Cohen. Singh invokes all four of the latter in the process of explaining that FLT might have turned out to be undecidable. Aczel's slender volume is particularly concerned with questions of priority, recounting in great and unseemly detail the circumstances surrounding the original enunciation of the Takayama–Shimura conjecture, including alleged attempts by certain members of Bourbaki to steal some or all of the credit, as well as Serge Lang's efforts to reflect the credit back onto the actual formulators of the pregnant (and still only partially proven) conjecture. Neither book can be recommended to anyone wishing a coherent nontechnical account of what Wiles et al. actually did.

Perhaps the most remarkable aspect of a proof like Wiles's, or of the classification of the simple groups during the late 1970s and the work of Feit and Thompson that pointed the way to that landmark result, is that they bring so astonishing a variety of seemingly unrelated results to bear on a single problem. In so doing, they confirm that the profession has (on the whole) been spending its time wisely—by steady accretion, mankind's collective problem-solving abilities have been elevated to previously unattained levels!

References

[1] E.T. Bell, *The Last Problem* (as updated by Underwood Dudley), MAA, Washington, 1990.

[2] J. Gleick, *Chaos*, Penguin Books, New York, 1987.

[3] E. Nagel and J.R. Newman, *Gödel's Proof*, NYU Press, New York, 1958.

James Case is an independent consultant who lives in Baltimore, Maryland.