CHEZ FERMAT A.D. 1637*

Erkka Maula and Eero Kasanen

Abstract

To set the stage, imagine a dinner party on a bright day of early September, 1637 in Toulouse. The main course was roast lamb with red wine and black olives from Cnidus. The select guests came from ancient Greece, contemporary France and modern Europe. For dessert, the host had chosen what later turned out to be his most famous creation. Today, its flavour is know to every mathematician and connoisseur. The aim of this essay is to study to what extent Fermat was justified in claiming that he also had a proof to his Last Theorem (FLT). The aim is not to offer yet another modern "proof".

Fermat's justification is studied by outlining first an historical scenario of the antecedents of FLT as a working hypothesis. It consists of three propositions and lemmas (*Props. 1-3* and *Lemmas 1-3*, with sketches of the proofs). These are elementary statements well within Fermat's reach and yet give a geometrical illustration of FLT. Their novelty is *Prop. 1* first suggested in [13:153-154]. These antecedents are called Fermat's heuristics.

The power of the outlined antecedents is measured by means of conclusions drawn from them (*Prop. 4* and *Lemma 4*), comparing the conclusions with modern results. In drawing these conclusions, only methods known from Fermat's own or his predecessors' works are employed. The comparisons indicate, however, that Fermat anticipated (granting his heuristics consisted of *Props. 1-3* and *Lemmas 1-3*) much later results. In particular, *Prop. 4* is more general than Terjanian's result in 1977 at *C.R.Acad.Sci. Paris 285*, and *Lemma 1* gives a better bound than M. Perisastri in 1969 at *Amer.Math.Monthly 76*. *Lemma 4*, in turn, offers a more promising way to an estimate of the exponent (n=p an odd prime) than Grünert's lower bound for an eventual solution to Fermat's equation in 1856 at *Archiv Math.Phys. 27*. These are the first mathematical results.

Further conclusions and comparisons are made possible by

Lemmas 5–6. They transform the problem and set the question of Fermat's justification into a new light. *Lemma 7* gathers together some results depending on *Prop. 4.* But that is only a watershed. A definitive answer is possible only if the final *Prop. 5*, FLT with odd exponents in one version, can be proved by Fermat's methods. Aiming at the proof, *Porisms 1–3* and *Lemma 8* are given. Enter Proof Reconstruction.

In the philosophical part, the implications of the foregoing heuristic, historical and mathematical considerations are outlined. They constitute, in our opinion, Fermat's true legacy with an impact on modern philosophy of mathematics and philosophical cosmology. In fact, this philosophical legacy runs parallel to Hamilton's research program which he gave up in favour of the quaternions (1844). Although Fermat's FLT and his Principle of the Least Time in optics are parts of his legacy, they are but the tip of an iceberg.

Fermat's Heuristics

It is 350 years since Fermat scribbled his "Last Theorem" (FLT) in the margin of his copy of Diophantus [3]. Despite recent advances, esp. Gerd Faltings' result (1983) and Yoichi Miyaoka's near-proof (1988), neither the mathematical nor the logical efforts nor yet computer calculations have been sufficient to solve the problem [cf. 15:2–3]. In the beginning of our century, Hilbert believed that the solution will be found [8]. In mid-1930's, however, unsolvable problems were gathering and the Theory of Algorithms was developed by Church and Turing. After the works of Post, Markov and others (c. 1947–1952), a negative solution was suggested to Hilbert's Problem X by Davis, Davenport, Putnam and Robinson (1953–1960). In 1970 it was found by Ju. V. Matijasevic and G. V. Cudnovskij [13:136–7]. This negative solution to the decision problem of a general Diophantine equation, although it does not rule out the possibility that the particular Diophantine equation FLT could either be positively solved or proven impossible to solve, reduced much of the hope [13:153].

Today, especially in Analytical Philosophy, FLT is often quoted as an example of Gödel's "true but undecidable statements" [15:216–8]. This is intellectual laziness. Gödel's result is of existential character and must not be used as a problem-killer. It is not worthwhile to claim conceptual command of a particular problem that one cannot solve. There is no rational reason for believing that just FLT is undecidable.

Other attempts having failed (so far), we suggest an additional study of Fermat's antecedents [13:153-4]. For it is fairly sure that he did not invent anything like the abstractions of modern Number Theory, and definitely did not anticipate the latest results of Theoretical Physics (which Miyaoka made use of). Ours is, therefore, a Requiem to Fermat's predecessors, in particular to the Pythagoreans and Euclid, Diophantus and Pappus.

There are two separate problems: (i) to prove FLT using concepts and methods available to Fermat, and (ii) to prove FLT by whatever means. The present day is inclined to the latter approach. The former one is more demanding, probably more elegant, and certainly closer to rules of fair play.

Fermat's words about having "invented a truly remarkable proof" may indicate that he indeed had a proof never published, or he had come up with what today is called a proof-idea but was lacking either the technical means or the motive to develop it, or his proof contained unorthodox methods not revealed. Among these, however, his "method of descent" hardly could be included as he had made use of it e.g. in proving the case $p(x,4) + p(y,4) \neq p(z,4)$. Finally, Fermat may have erred, just as his intuition had failed with the numbers F^n [6:14-15].

As Fermat's note faced Diophantus' proof of the Pythagorean equation $p(x,2) + p(y,2) = p(z,2)$ (Hardy and Wright say he "proved the substance of Theorem 255", the same case; [6:190-1]) concerned with right-angled triangles with integer sides, Fermat may well have studied whether triangles occur in other cases also. We show that they do.

It can be taken for granted (cf. Pascal's words, [15:3]) that Fermat knew not only Diophantus but even his predecessors in and out, and in particular Euclid. Needless to say that Euclid was the geometer whose rigour and strategies gave the very paradigm for Fermat's predecessors.

We begin with Euclid too. Of particular interest, then, is the Pythagorean theory of triangles in Euclid's *Elementa*. It retains relics like I.1 (construction of an equilateral triangle as if it were the foundation of all the rest) and I.21 (which Euclid proves but never uses afterwards, [7:I:377]), and construction as a means of existence proof. But an actual construction has a double role: In proofs of existence [7:I:377;20] and as a culmination of the heuristic method of Analysis and Synthesis [cf. 12:120-2] in "something already known and being first in order" (Pappus) These roles, especially if not clearly distinguished from one another, may well account for the element of surprise in Fermat's characterization "a truly remarkable proof".

For geometrical figures are concrete and constitute the primary objects of understanding for Fermat's Greek predecessors. Higher potencies are abstract (the Greeks favoured squares and cubes), unless made concrete by means of geometrical constructions.

That again is what we have learned to expect in Greek contexts: a surprising concreteness, as it were, amidst the most abstract thought, such as the *gnomon* about which a whole world-view is turning [17], or a combination of sliding calipers and a plane sun-dial in the hub of Eudoxus's astronomy [11]. It is compatible with early Greek ideas of all scientific demonstration [cf. 16] and, we submit, also with Fermat.

In particular, geometric proof and intuition were legitime tools for Fermat just as they had been for the Greeks. That is why we direct our search for Fermat's heuristics to geometry. FLT, however, has implications which reach other fields of mathematics and physics as well.

The antecedents of FLT consist of three propositions and lemmas.

> *Prop. 1.* If $p(x,n) + p(y,n) = p(z,n)$ has a solution in positive integers and $n > 2$, then (x,y,z) are sides of a scalene triangle $(0 < x < y < z)$.

If $z \leq x$ or $z \leq y$, then $p(z,n) < p(x,n) + p(y,n)$. Hence $z > x$ and $z > y$. If $z > x + y$, then $p(z,n) > p(x+y,n) > p(x,n) + p(y,n)$. Hence $z < x + y$. But if further $x = y$, then $p(z,n) = 2p(x,n) = 2p(y,n)$ and $z = r(2,n)x = r(2,n)y$ where x, y and z cannot all be integers. Hence $x \neq y$. Let $x < y$. Thus $z > y > x$. From this and from $z < x + y$ it follows that $y < z + x$, $x < z + y$, $z - y < x$, $y - x < z$ and $z - x < y$. Therefore (x,y,z) are integer sides of a scalene triangle by Euclid's *Elementa* I.20. QED

In what follows we subdivide *positive-side* scalene triangles, say (a,b,c); $a > b > c > 0$, into acute-angled, right-angled and obtuse-angled scalene triangles (with positive sides) and define the equivalences:

> *Def. 1.* A scalene triangle (a,b,c) is acute-angled iff $p(a,2) < p(b,2) + p(c,2)$; $a > b > c$.
>
> *Def. 2.* A scalene triangle (a,b,c) is right-angled iff $p(a,2) = p(b,2) + p(c,2)$; $a > b > c$.
>
> *Def. 3.* A scalene triangle (a,b,c) is obtuse-angled iff $p(a,2) > p(b,2) + p(c,2)$; $a > b > c$.

As a matter of fact, Euclid's *Elementa* distinguishes (I:Defs. 10,11,12) acute, right and obtuse angles [17:173]; for *Defs. 1–3* see Heron's *Metrica* I, ch.4. In Euclid, the right angle and its simple parts suffice [17:173–7].

Lemma 1. If $p(x,n) + p(y,n) = p(z,n)$ $(x < y < z)$ has a solution in positive integers and $n > 2$, then x, y, and z are integer sides of an acute-angled scalene triangle.

As to proof, use *Prop. 1*, set $m = 2$ below in *Prop. 2* and refer to *Def. 1*. QED

Obviously it holds good also that

Lemma 2. If $p(x,n) + p(y,n) = p(z,n)$ has a solution in positive integers and $n > 2$, then there is the smallest acute-angled triangle (a,b,c) such that $\gcd(a,b,c) = 1$ and $a > b > c > 0$ and the integers a, b, c, n also constitute a solution, the acute-angled scalene triangle (z,y,x) from *Lemma 1* being a multiple of it.

In the sequel we focus on the smallest integer solution (a,b,c,n).

Prop. 2. If $p(a,n) = p(b,n) + p(c,n)$ has a solution in positive integers and $n > 2$ $(a > b > c)$, then $(p(a,n-1),p(b,n-1),p(c,n-1))$ ... $(p(a,2),p(b,2),p(c,2))$ are triangles and $(p(a,n-1) > p(b,n-1) > p(c,n-1) > 0)$ ... $(p(a,2) > p(b,2) > p(c,2) > 0)$ integers.

If $p(a,n-1) \geq p(b,n-1) + p(c,n-1)$, then $p(a,n) = ap(a,n-1) \geq ap(b,n-1) + ap(c,n-1) > p(b,n) + p(c,n)$ as $a > b > c$. Hence $p(a,n-1) < p(b,n-1) + p(c,n-1)$. By similar proofs, if $p(a,m) \geq p(b,m) + p(c,m)$ where $n-1 > m > 1$, then $p(a,n) > p(b,n) + p(c,n)$ as $a > b > c$. Hence $p(a,m) < p(b,m) + p(c,m)$. Furthermore, $p(a,n-1) > p(b,n-1) > p(c,n-1)$ and $p(a,m) > p(b,m) > p(c,m)$, because $a > b > c$. From these it follows (quite as in *Prop. 1*) that also all integer powers of a,b,c up to the power n-1, are integer sides of triangles (by *Elem.* I.20). QED

We say that the triangle (a,b,c) in *Prop. 2* is $(n-1)$-*potent* and define :

Def. 4. A scalene triangle (a,b,c) is n-potent iff $a > b > c$ and $p(a,n) < p(b,n) + p(c,n)$.

Equilateral triangles are (∞)-potent; right-angled and obtuse triangles are (1)-potent. Within a given perimeter a+b+c an acute-angled triangle (a,b,c) has its highest potency when $a = b+1$ and $c = b-1$, but in that case $p(a,n) = p(b,n) + p(c,n)$ $(n>2)$ has no integer solutions [4]. The outcome of all this is that (n) is not an independent variable. Its value depends both on $2h = a+b+c$ and on how far apart a,b,c are from one another. We know that if FLT fails, $a-b < c$ and $b-c < a$ (*Prop. 1*). Inkeri and Hyyrö [9] in fact came very close to this in their restriction $z-y$, $y-x < M$, where $M > 0$ is given in advance [15:25]. Under this restriction, [9] anticipates Gerd Falting's result (1983): there are at most finitely many integer solutions to $p(a,n) = p(b,n) + p(c,n)$ when $n > 2$, if any solutions at all.

Prop. 3. For a (n-1)-potent scalene triangle (a,b,c) where a > b > c > 0 and n > 2 and gcd(a,b,c) = 1, none of its integer powers k ≤ n-1 is a right-angled triangle with integer sides.

The proof follows from Diopohantus's proof. Recall the characterization of right-angled triangles with p,q. Now suppose the claim does not hold and we find a right-angled triangle (p(a,k),p(b,k),p(c,k)), where 1 < k ≤ n-1, in the sequel. Hence a = r(p(p,2)+p(q,2),k), b = r(2pq,k) or b = r(p(p,2)-p(q,2),k) and c = r(p(p,2)-p(q,2),k) or c = r(2pq,k), where p,q are relatively prime integers not both odd and $p > q$. Thus a,b,c cannot all be integers. A contradiction; hence the claim is proven. QED

Lemma 3. For a (n-1)-potent scalene triangle (a,b,c) where a > b > c > 0 and n > 2 and gcd(a,b,c) = 1, the highest integer potency (n-1) is an obtuse-angled triangle with integer sides.

The proof follows from *Prop. 3* and from the truth that if we have an acute-angled triangle (p(a,k),p(b,k),p(c,k)) which is a potency of (a,b,c) and where 1 ≤ k < n-1, then also (p(a,2k),p(b,2k),p(c,2k)) where 2 ≤ 2k ≤ n-1, is a triangle (by *Def. 1*). If k is the highest potency that belongs to an acute-angled triangle, then 2k ≤ n-1 and n-1 cannot belong to an acute-angled triangle nor to a right-angled triangle. This proves the claim. QED

By way of a summary, then, if FLT fails and p(a,n) = p(b,n) + p(c,n) where n > 2 has a solution in positive integers, then there is a scalene acute-angled (n-1)-potent triangle (a,b,c) where a > b > c > 0 and gcd(a,b,c) = 1. The rising potencies are represented by scalene triangles of increasing apex angle and change from acute-angled to obtuse-angled, but none of the rising integer potencies is represented by a right-angled triangle with integer sides.

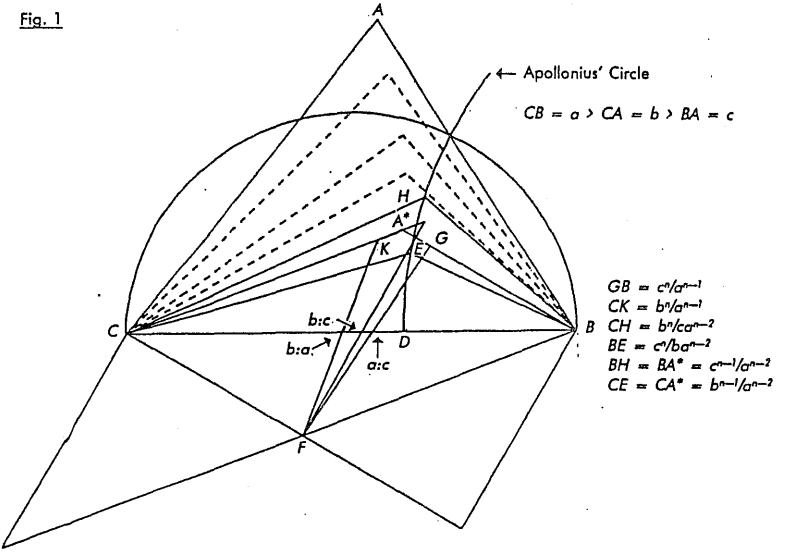
Anticipating our subsequent discussion, this must have been Fermat's pivotal heuristic vision. Because there is no right-angled integer-side triangle with an integer exponent, but it is possible that some equal powers of (a,b,c) are sides of a right-angled triangle, one must consider non-integer potencies also. On the other hand, if we postulate a non-integer potency of (a,b,c) that is a right-angled triangle, then there cannot be any other non-integer potency of (a,b,c) that is a right-angled triangle. Here we have a seed of *reductio ad absurdum*, awaiting the proof.

In the last part of the heuristic analysis, we make use of *Elem.I.21* and put all the triangles that are potencies of a (n-1)-potent scalene acute-angled triangle (a,b,c) on the same

scale with (a) as their common base (a > b > c). For a general view, integer potencies suffice.

Next, by means of elementary Euclidean constructions (taking one side in turn as the base and changing the other two sides, bisecting an angle, and drawing transversals parallel with the sides) we construe on the base CB = a the points of division (see *Fig. 1*) a:b, a:c, b:c, c:b, c:a and b:a.

Fig. 1



Complementing the triangles (a,b,c) ... (p(a,n-1),p(b,n-1), p(c,n-1)), now put on the same scale, one by one into parallelograms, we have made an escalator. It brings us algoristically from any potency of (a,b,c) to the next, and even up to the (nth) potency. Conversely, in the heuristic synthesis, it brings us algoristically downwards, down to the first potencies again. In all these operations, rational lines remain rational.

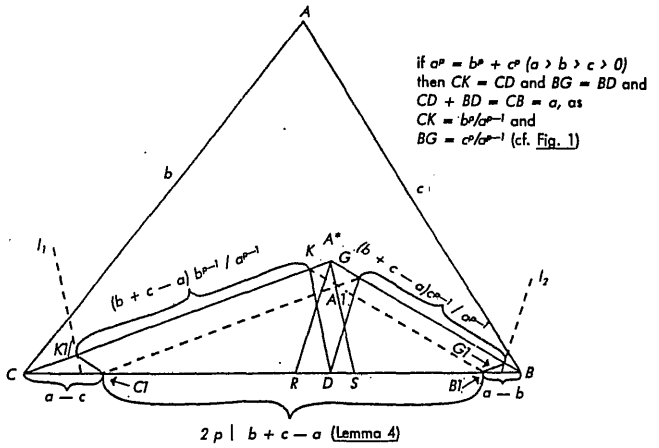
We now focus on the highest power (n-1) represented by an obtuse-angled triangle (*Lemma 3*) CA*B = (a,p(b,n-1)/p(a,n-2), p(c,n-1)/p(a,n-2)), and complement it into the parallelogram CA*BF by construing another, identical triangle. The line FG through the point of division a:c cuts the portion GB = p(c,n)/p(a,n-1) off BA*, and the line FK through the point of division b:a cuts the portion CK = p(b,n)/p(a,n-1) off CA*. The line through the point of division b:c and F cuts the portions (p(c,n)/bp(a,n-2)) off BA* and (p(b,n)/cp(a,n-2)) off the extension of CA*.

Finally, we construe the triangles CHB = (a,p(b,n)/cp(a,n-2), p(c,n-1)/p(a,n-2)) and CEB = (a,p(b,n-1)/p(a,n-2), p(c,n)/bp(a,n-

2)) and bisect their vertex angles thus obtaining (by *Elem.*VI.3) twice the point of division D = (p(b,n):p(c,n)). Apollonius' Circle, therefore, runs via the points D, E and H.

It remains to show that the triangles CKD and BGD are not both isosceles (*Fig.2*), and FLT is proven.

Fig. 2



$l_1 \,//\, KD \,//\, A^*S$ and $l_2 \,//\, GD \,//\, A^*R$;
$CK \,//\, C1A1 \,//\, B1G1$ and
$BG \,//\, B1A1 \,//\, C1K1$

triangles $CA^*B \equiv C1A1B1 \equiv CK1C1 \equiv B1G1B$
and $CA^*S \equiv CKD$
and $BA^*R \equiv BGD$

We postpone the question of a proof toward the end of the essay. In the meantime, we hope we have been able to clarify Fermat's way of reasoning, as indeed all the clues into it extracted from the original source, have been utilized [cf.1:289–306 and 2]. In case we have correctly interpreted Fermat's heuristics, however, we have gained inside information, too. Let us use it.

First results

> *Prop.4.* If p(a,n) = p(b,n) + p(c,n) has a solution in positive integers and n > 2, then the exponent n is odd and (b+c) | p(a,n).

The proof follows from *Prop. 3* in a few steps. (a,b,c) is an (n-1)-potent acute-angled scalene triangle (a > b > c > 0) by *Prop. 2*, and by *Lemma 2* gcd(a,b,c) = 1. Let k ≥ 1 be the exponent of the highest acute-angled potency of (a,b,c), represented by the triangle (p(a,k),p(b,k),p(c,k)). Hence (p(a,2k),p(b,2k), p(c,2k)) is still a triangle (and obtuse-angled

by *Lemma 3*) and $p(a,2k) < p(b,2k) + p(c,2k)$ by *Def. 1*. The next potency of (a,b,c) with the exponent k+1, is represented by an obtuse-angled triangle $(p(a,k+1),p(b,k+1),p(c,k+1))$ by *Prop. 3*. Hence $p(a,2k+2) > p(b,2k+2) + p(c,2k+2)$ by *Def. 3*. Thus if $p(a,n) = p(b,n) + p(c,n)$, then $2k < n=2k+1 < 2k+2$, and $n = 2k+1$ is odd. Hence $p(b,n) + p(c,n) = (b + c)(p(b,n-1) - p(b,n-2)c + \ldots - bp(c,n-2) + p(c,n-1))$, and $(b + c) \mid p(a,n)$ (where $b+c > a > b-c$ by *Prop. 1*). QED.

In December 1977 Terjanian proved an equivalent result for the first case of FLT (it holds for the exponent p when there do not exist integers x,y,z all different from zero, such that $p \nmid xyz$ and $p(x,p) + p(y,p) = p(z,p)$), viz. if x,y,z are nonzero integers such that $p(x,2p) + p(y,2p) = p(z,2p)$, where p is an odd prime, then $2p$ divides x or y [18:973-975; cf.15:65]. *Prop. 4*, however, holds also for the second case of FLT (it holds for the exponent p when there do not exist integers x,y,z all different from zero, such that $p \mid xyz$, $\gcd(x,y,z) = 1$ and $p(x,p)+p(y,p) \neq p(z,p)$).

According to Ribenboim [15:26-7,66], Terjanian's result (1977) was the best on even exponents. By *Prop. 4*, therefore, Fermat not only anticipated but also exceeded the most recent results in this direction. Fermat's Method of (Infinite) Descent, which he used in proving the separate case $p(x,4) + p(y,4) = p(z,4)$ and many other propositions with great success, provides an alternative to *Prop. 4* in attacking FLT (see below).

If FLT holds for an exponent n, then it holds also for any multiple of n. And since every integer $n \geq 3$ is a multiple of 4 or of a prime $p \neq 2$, it suffices to prove FLT for $n = 4$ and for every prime $p \neq 2$. These easy observations [cf. 15:3] must have been made by Fermat, too, because he actually proved the case $n = 4$ [cf. 3,II,Ch.xxii]. Hence we can restrict ourselves to exponents $n = p$, where p is an odd prime.

Considerations of parity show that if $p(a,n) = p(b,n) + p(c,n)$, then $2h = a+b+c$ is even, and as $\gcd(a,b,c) = 1$ by *Lemma 2*, one and only one of a,b,c is even. Without loss of generality [15:36], a,b,c are also pairwise relatively prime.

Lemma 4. If $p(a,p) = p(b,p) + p(c,p)$ has a solution in positive integers and $p > 2$ is a prime, then $2p \mid b+c-a$.

The proof follows from Fermat's Theorem (FT, 1640): if p is a prime and $p \nmid r$, then $p(r,p-1) \equiv 1 \pmod{p}$; Fermat, of course, spoke of divisibility. Multiplying both sides of the congruences $p(a,p-1) \equiv 1 \pmod{p}$, $p(b,p-1) \equiv 1 \pmod{p}$ and $p(c,p-1) \equiv 1 \pmod{p}$ where p is a prime and $p \nmid abc$, by a,b and c respectively, we obtain $p(a,p) \equiv a \pmod{p}$, $p(b,p) \equiv b \pmod{p}$ and $p(c,p) \equiv c \pmod{p}$. Between these two groups of three, the first and the second and the third congruences are pairwise equivalent when $p \nmid abc$

and the latter three are trivial when $p \mid abc$ [6:63]. From the latter three congruences it follows $p \mid p(a,p) - p(b,p) - p(c,p) + b + c - a$ and, if $p(a,p) = p(b,p) + p(c,p)$, $p \mid b + c - a$. Since $b + c - a$ is even because of parity when $p(a,p) = p(b,p) + p(c,p)$, and p odd by *Prop. 4*, it follows that $2p \mid b + c - a$. QED.

FT which Fermat stated in 1640 (*Oeuvres*,ii:209), was proved by Euler in 1736 and generalized in 1760 [3,I,Ch.iii]. We do not consider it anachronistic, however, to include FT into Fermat's arsenal when he claimed to have a proof of FLT: it is even a plausible candidate for an "unorthodox method not revealed"in 1637.

In 1856 Grünert proved that if $0 < x < y < z$ are integers and $p(z,n) = p(x,n) + p(y,n)$, then $x > n$ [5:119-120;cf. 15:226], where n is given. Since then, much better lower bounds for eventual solutions to Fermat's equation $p(z,n) = p(x,n) + p(y,n)$ have been found with elementary methods also. Yet, to quote Ribenboim [15:227], "conceptually, these results do not throw any more light onto the problem". In this respect, Fermat's $2p \mid b + c - a$ from *Lemma 4* is a more promising bound than Grünert's (see below).

From Grünert's proof it follows [15:226] $y < z < y + x/n < y(1 + 1/n)$ and hence z, y are relatively close together; therefore the size of x should be much smaller. On the other hand, M. Perisastri showed in 1969 that x cannot be much smaller than z: $z < p(x,2)$ [14:671-675]. But in an acute-angled triangle (z,y,x) where $z > y > x > 0$ (*Lemma 1*), $p(x,2) + p(y,2) > p(z,2)$ by *Def. 1*, and thus $p(x,2) > (z + y)(z - y) = (z + y)r$ where $x > z - y = r \geq 1$. Hence $z < p(x,2)/r - y < p(x,2)$ or a better bound still. Perisastri's bound $z < p(x,2)$ does not exlude all obtuse-angled triangles with integer sides. It is, therefore, less stringent than *Lemma 1*.

It is a natural continuation to proceed from *Prop. 4* and *Lemma 4* toward an analysis of the exponent p, using the methods of *Elementa* VII-IX. It is unnecessary to reproduce Euclid's notation, however.

As a,b,c are pairwise relatively prime (without loss of generality), the sum of any two and each summand are likewise (by *Elem.* VII.28). But (b+c) and $p(a,p)$ are not relatively prime, and neither (b+c) nor (a) a prime. For if (b+c) were a prime and since $(b+c) \mid p(a,p)$ (by *Prop. 4*), then $(b+c) \mid a$ (by *Elem.* VII.30 = [6:Theorem 3]), which is impossible because $b+c > a$ (by *Prop. 1*). Thus $(b+c) \mid p(a,p)$ but $(b+c) \nmid a$. On the other hand, (b+c) cannot be a multiple or a higher power of (a) either. Since $a > b > c$ are positive integers (by *Prop. 1*), within any perimeter $2h = a + b + c$, it holds that $b+c \leq b + (b-1)$ and $a \geq b+1$. Hence $a <$

b+c < 2a-2 < 2a. On all these grounds, then, (b+c) | p(a,*p*) implies that both (b+c) and (a) are composite, all prime factors of (b+c) are also prime factors of (a) but not conversely, and at least one prime factor of (b+c) has a higher exponent than the corresponding prime factor of (a).

According to Euclid, "any number either is a prime or is measured by some prime number" (*Elem.* VII.32, equivalent to [6:Theorem 1]) and "can only be resolved into prime factors in one way" (*Elem.* IX.14, equivalent to [6:Theorem 2], the Fundamental Theorem of Arithmetic). Let (a) be expressed in the standard form: a = p(*p*^1,m^1)p(*p*^2,m^2)...p(*p*^k,m^k); m^1 > 0, m^2 > 0, ..., m^k > 0; *p*^1 < *p*^2 < ... < *p*^k. Let further all the prime factors of (a) be divided into two sets so that *p*^i ... *p*^j are not prime factors of (b+c) while the rest *p*^r ... *p*^w are also prime factors of (b+c). Thus b+c = p(*p*^r,n^r) ... p(*p*^w,n^w). Next, let the prime factors of (b+c) be subdivided into three sets so that for *p*^r...*p*^s the exponents n^r > m^r, ..., n^s > m^s; for *p*^v ...*p*^w the exponents m^v > n^v, ..., m^w > n^w; and for the rest p^t ... p^u the exponents are equal m^t = n^t, ..., m^u = n^u. Finally, let gcd(b+c,a) be extracted.

Thus b + c - a = (p(*p*^r,n^r) ... p(*p*^w,n^w)) - (p(*p*^r,m^r) ... p(*p* ^w,m^w))(p(*p* ^i,m^i) ... p(*p* ^j,m^j)) and gcd(b+c,a) = p(*p* ^r,min(n^r,m^r)) ... p(*p* ^w,min(n^w,m^w)) and (b+c-a)/gcd(b+c,a) = A, where A = (p(*p*^r,n^r-m^r) ... p(*p*^s,n^s-m^s)) - (p(*p*^v,m^v-n^v) ... p(*p*^w,m^w-n^w))(p(*p*^i,m^i) ... p(*p*^j,m^j)), and A > 0 because b+c > a (by *Prop. 1*). If for all the exponents n^r ... n^w, m^r ... m^w it holds that min(n^r,m^r) = m^r, ..., min(n^w,m^w) = m^w, then the formula for A is simpler : A = (p(*p*^r,n^r-m^r) ... p(*p*^w,n^w-m^w)) - (p(*p*^i,m^i) ... p(*p*^j,m^j)) > 0. In both cases, let A be expressed in the standard form as A = p(q^1,l^1)p(q^2,l^2)... p(q^f,l^f).

From *Prop. 4*, *Lemma 4* and the argument above we obtain

Lemma 5. If p(a,*p*) = p(b,*p*) + p(c,*p*) has a solution in positive integers and p > 2, then (a) and (b+c) are composite, b+c = p(*p*^r,n^r-m^r) ... p(*p*^w,n^w-m^w) is composed solely of prime factors of a = (p(*p*^r,m^r) ... p(*p*^w,m^w)) (p(*p*^i,m^i) ... p(*p*^j,m^j)) but not conversely, at least for one pair of the exponents n^r...n^w, m^r...m^w, it holds that n^r > m^r,..., or n^w > m^w because a < b+c < 2a-2, and (α) *p* | gcd(b+c,a) and *p* is one of the primes *p*^r,...,*p*^w or (ß) *p* | A, where A = (b+c-a)/gcd(b+c,a) = p(q^1,l^1)p(q^2,l^2)... p(q^f,l^f) > 0 is a function of the prime factors of (a), and *p* is one of the primes q^1, q^2, ..., q^f.

This result transforms the main problem of Fermat's justification into a search for an odd prime (*p*) dependent on the prime

factors of (a) alone. Let us see, how far *Prop. 4* will take us.

It is appropriate to make some comments on the two alternatives: (α) $p \mid \gcd(b+c,a)$ or (β) $p \mid A$ (from *Lemma 5*); their relationship is discussed after *Lemma b*.

If $p \mid \gcd(b+c,a)$, then $p \mid b+c$ and $p \mid a$ but, because a,b,c are pairwise relatively prime (without loss of generality), $p \nmid bc$. This corresponds to the Second Case of FLT ($p \mid xyz$, $\gcd(x,y,z) = 1$) in a sharper form: if the Second Case fails, then p is one of the primes $p\hat{}r$, ...,$p\hat{}w$. Since $p \nmid bc$, $p \mid p(b,p-1) - 1$ and $p \mid p(c,p-1) - 1$ (by FT, 1640), and of course $p \mid p(a,p-1)$. Therefore also $p \mid p(b,p-1) - p(c,p-1)$, $p \mid p(a,p-1) + p(b,p-1) - p(c,p-1)$ and $p \mid p(a,p-1) - p(b,p-1) + p(c,p-1)$.

In the second alternative, if $p \mid A$, then $A \nmid abc$ for otherwise the First Case of FLT cannot fail. Hence $p \mid p(a,p-1) - 1$, $p \mid p(b,p-1) - 1$ and $p \mid p(a,p-1) - 1$ (by FT, 1640), and also $p \mid p(a,p-1) - p(b,p-1)$, $p \mid p(a,p-1) - p(c,p-1)$, $p \mid p(b,p-1) - p(c,p-1)$, $p \mid p(a,p-1) + p(b,p-1) - p(c,p-1) - 1$, $p \mid p(a,p-1) - p(b,p-1) + p(c,p-1) - 1$ and $p \mid p(b,p-1) + p(c,p-1) - p(a,p-1) - 1$ but $p \nmid p(b,p-1) + p(c,p-1) - p(a,p-1)$ (cf. *Fig.2*, where RS = $(p(b,p-1) + p(c,p-1) - p(a,p-1))/p(a,p-2)$).

It is plain that in the search of p, the size of b+c-a is reducible: if $2p \mid b+c-a$, then $2p \mid (b+c- a)/(p(p\hat{}r,\min(n\hat{}r,m\hat{}r)-1) \ldots p(p\hat{}w,\min(n\hat{}w,m\hat{}w)-1))\ (p(q\hat{}1,l\hat{}1-1)\ldots p(q\hat{}f,l\hat{}f-1)) = (p\hat{}r\ldots p\hat{}w)(q\hat{}1\ldots q\hat{}f)$. In case (a) is odd, $(q\hat{}1) = 2$. In case (a) is even, $(p\hat{}1) = 2$ and $(p\hat{}1)$ is the smallest one of the primes $p\hat{}r\ldots p\hat{}w$, for otherwise b+c-a were odd which is impossible because of parity; if further $n\hat{}1 = m\hat{}1$ also $(q\hat{}1) = 2$, otherwise $(q\hat{}1)$ is an odd prime.

Given (b+c) and (a), FLT does *not* fail, therefore, if (b+c) or (a) is a prime, if they are not of the same parity, or if not $a < b+c < 2(a-1)$. The First Case of FLT does *not* fail if $\gcd(b+c,a)$ or $(p\hat{}r\ldots p\hat{}w) = 1$ or any power of 2. The Second Case of FLT does *not* fail if A or $(q\hat{}1\ldots q\hat{}f) = 1$ or any power of 2. Given p, FLT does *not* fail if $p \mid b$ or $p \mid c$. Given a,b,c, one recalls, FLT does *not* fail if (a,b,c) is an *odd-potent* triangle (by *Prop. 4* and *Def. 4*). All these restrictions have an eliminative character, but they do not prove FLT even if combined.

Using the same notation and the parity of (a) as a pilot study, one can further complement the findings of *Lemma 4* and *Lemma 5*.

If $p \mid a$ (and $p \nmid bc$) where $p = 2k+1$ an odd prime, then p is one of the primes $p\hat{}1 \ldots p\hat{}k$, i.e. one of the primes $(p\hat{}i\ldots p\hat{}j)$, $(p\hat{}r\ldots p\hat{}s)$, $(p\hat{}t\ldots p\hat{}u)$ or $(p\hat{}v\ldots p\hat{}w)$. The first subset $(p\hat{}i\ldots p\hat{}j)$ is excluded by *Lemma 5*, however, as none of the primes $(p\hat{}i\ldots p\hat{}j)$ divides $\gcd(b+c,a)$ or A. Hence one can complement

Lemma 4 with

Lemma 6. If $p(a,p) = p(b,p) + p(c,p)$ has a solution in positive integers, $p > 2$ is a prime and $p \mid a$ and $\gcd(b+c,a) = 1$, then $2p \mid a+b+c$.

This states the equivalence $p \mid a$ iff $p \mid b+c$ as $p \mid a$ iff $p \mid a/p(p^i,m^i)...p(p^j,m^j)$.

If p is one of the primes $(p^r...p^s)$, $(p^v...p^w)$, then by *Lemma 5* $p \mid \gcd(b+c,a)$ and $p \nmid A$. If p is one of the primes $(p^t...p^u)$, then $p \mid \gcd(b+c,a)$ and either $p \nmid A$ or $p \mid A$; $p \mid A$ provided $(b+c)/\gcd(b+c,a) = (p(p^r,n^r-m^r)...p(p^s,n^s-m^s))$ and $a/\gcd(b+c,a) = (p(p^v,m^v- n^v)...p(p^w,m^w-n^w))(p(p^i,m^i)... p(p^j,m^j))$ are representatives of the same residue class (mod p). In that case $2p(p,2) \mid b+c-a$.

If $p \nmid abc$, then p is none of the primes $p^1...p^k$ and $p \nmid \gcd(b+c,a)$ but, by *Lemma 5*, $p \mid A = p(q^1,l^1)...p(q^f,l^f)$. In this case $(b+c)/\gcd(b+c,a)$ and $a/\gcd(b+c,a)$ must be representatives of the same residue class (mod p) or the First Case of FLT cannot fail. To sum up : if (α), then either (β) or not (β); if not (α), then (β); $p \mid a$ iff (α).

As a corollary of *Lemmas 5-6*, we can give a proposition in the positive tone. To fix the notation, let $p = p^r$ in the first alternative and $p = q^f$ in the second alternative of *Lemma 5*; the corresponding exponents are $m^r, l^f \geq 1$. Then the following proposition holds good

Lemma 7. If $p \neq 2$ and if $a > b > c > 0$ are relatively prime integers such that $p(a,p) = p(b,p) + p(c,p)$ and $p \nmid abc$, then $b+c-a \equiv 0 \pmod{p(p,l^f)}$; and if $p \mid a$, then $b+c \pm a \equiv 0 \pmod{p(p,m^r)}$. Moreover, if $p \mid a$, then $p(p,m^r) \mid a$.

Parts of our argument and partial analogous results are scattered in the literature. See, for instance, Vandiver's application of Furtwängler's theorem : "If $p \neq 2$ and if x,y,z are relatively prime integers such that $p(x,p) + p(y,p) + p(z,p) = 0$, then $p(x,p) \equiv x$, $p(y,p) \equiv y$, $p(z,p) \equiv z \pmod{p(p,3)}$ and $x+y+z \equiv 0 \pmod{p(p,3)}$. Moreover, if $p \mid z$, then $p(p,3) \mid z$", in [19:73-80] where methods of Class Field Theory are used on the Second Case of FLT [cf. 15:170]. We have not seen such arguments and results gathered anywhere as they are in *Lemmas 5-7*, and never deduced from Fermat's premisses by contemporary methods. Yet this is only a watershed: odd exponents await us. The momentum of *Prop. 4* is vaning here.

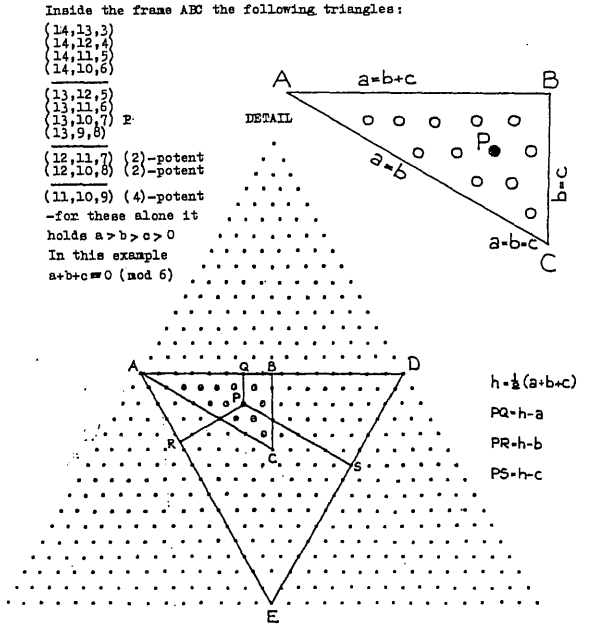
Further Problems

We are now in the position to suggest a strategy of proof

possible for Fermat. In case his premisses consisted of *Props. 1-3* and *Lemmas 1-3*, he could first show (by *Lemma 5*) that, if FLT fails, then the exponent p, an odd prime, is (α) one of the primes p^r, ..., p^w when $p \mid a$ or (β) one of the primes q^1, ..., q^f when $p \nmid abc$. Next, given a,b,c he could determine, independently of *Lemma 5*, the highest potency of the acute-angled scalene (n-1)-potent triangle (a,b,c) starting from *Prop. 2* and *Lemma 3*. Finally, Fermat could try to show that $p = (n-1)+1$ is none of the primes p^r, ..., p^w, q^1, ..., q^f eventually establishing a contradiction. If that succeeds, then FLT is proven. This is an arithmetical strategy, and it is well known that Fermat's predilections were arithmetical.

It was stated earlier (following *Def. 4*) that the potency of an acute-angled scalene triangle (a,b,c), where $a > b > c > 0$, depends both on the perimeter $a+b+c = 2h$ which must be even because of parity, and on how far apart a,b,c are from one another. Consider now all scalene integer-side triangles with the even perimeter 2h. They can be illustrated as a subset of points in a Pythagorean discrete-point equilateral triangle (*Fig. 3*).

Fig.3 Pythagorean triangular numbers generating a discrete-space



Let a = h-(h-a), b = (h-(h-a)) - (a-b) and c = 2(h-a) + (a-b). If (a,b,c) is acute-angled, then by *Def. 1* p(a,2) < p(b,2) + p(c,2). This is the case when (b+c-a)/2 = h-a > (1/4)(-3(a-b) + r(p(a-b,2) + 8h(a-b),2)). The corresponding curve and even an approximate algorism can easily be construed. Using Cardano's Rules, similar operations can be carried out for third and fourth potencies of (a,b,c). That is the end of the road, however, and *Lemma 4* is of no more help.

There is also another possible strategy of proof, more geometric perhaps.

We draw a parallel GC1 to CK and another parallel KG1 to BG (see *Fig. 2*). By *Elementa* VI.2 the sides A*B and CB are divided proportionally at G and C1, and the sides A*C and BC at K and B1. Thus CC1 = a-c and BB1 = a-b. (Conversely, we can cut the segments CC1 = a-c and BB1 = a-b off the line CB = a, obtaining the parallels C1G // CA* and G1K // BA*.) Thus the line segment C1B1 = b+c-a, which is divisible by $2p$ (*Lemma 4*).

If p(a,p) = p(b,p) + p(c,p) (a > b > c > 0), the triangles CKD and BGD are isosceles and CK = CD = p(b,p)/p(a,p-1) and BG = BD = p(c,p)/p(a,p-1). If we then draw parallels l^1 // KD // A*S and l^2 // GD // A*R, also the sides A*B and RB are divided (*Elem.* VI.2) proportionally at G and D, and the sides A*C and SC at K and D.

Thus we can transfer unaltered any line segment from CA* to CS and *vice versa* by means of parallels to l^1, and any line segment from BA* to BR and *vice versa* by means of parallels to l^2. Again, by means of parallels to G1K and to C1G, we can transfer, without altering their divisibility, a line segment from CB to CA* and BA*, and *vice versa* - provided proper care is taken of their scale. For instance, in the triangle C1A1B1 similar to the triangle CA*B, $2p$ | b+c-a = C1B1, but C1A1 = K1K = (b+c-a) p(b,p-1)/p(a,p-1) and B1A1 = G1G = (b+c- a)p(c,p-1)/p(a,p-1) are on a different scale.

This simple geometrical machinery is based on proportionality and on the geometrical illustration of FLT: if FLT fails, then the triangles CKD and BGD are both isosceles. Restricting oneself to line segments that are divisible by p (pointed out in the discussion after *Lemma 5*), it is possible to conceive of geometrical operations both orthodox and unorthodox, toward an estimate of the exponent p. It seems to us that the First and Second Case of FLT must be dealt with separately.

In the Second Case, for instance, one could begin with $2p$ | b+c-a (*Lemma 4*) and apply on (b+c)-(a) methods akin to *Elementa* X.2, attempting to prove a stronger result than *Lemma 5*, viz. that p is not an integer. In the last step, also the propositions

Elementa IX.16-17 can be made use of. Although *Lemma 4* and *Lemma 6* are at one's disposal here, no corresponding results can be obtained for the other two Heronic (actually Archimedean, cf. [7, vol. ii:322]) constituents (a-b+c), (a+b-c). Hence this is one more cul-de-sac; we have not been able to find any promising continuation in this direction.

We have, however, one more possible strategy of proof to offer.

An Ancient Stratagem

We recall that since Fermat himself had proven the case $p(x,4) + p(y,4) \neq p(z,4)$, it is sufficient to prove FLT for all odd prime exponents. By *Prop. 4*, it is sufficient to prove FLT for all odd exponents, say, of the form 2k+1. This gives rise to a proof-idea, perhaps the simplest of them all. If $a > b > c$ are integer sides of a (2k)-potent acute-angled scalene triangle and $p(a,2k+1) = p(b,2k+1) + p(c,2k+1)$, then $(p(a,(2k+1)/2), p(b,(2k+1)/2), p(c,(2k+1)/2))$ is a right-angled triangle, and conversely. The exponents are arithmetical means of two consecutive integers k+1,k ($k \geq 1$), and the sides of the right-angled triangle thus geometric means of sides of an acute-angled triangle $(p(a,k),p(b,k),p(c,k))$ and of an obtuse-angled triangle $(p(a,k+1),p(b,k+1),p(c,k+1))$, which are themselves potencies of the (2k)-potent acute-angled triangle (a,b,c). It suffices, therefore, to prove the following proposition.

> *Prop. 5.* If $a > b > c > 0$ are pairwise relatively prime integers and sides of an acute-angled triangle such that $p(a,k)$, $p(b,k)$, $p(c,k)$ are sides of an acute-angled triangle and $p(a,k+1)$, $p(b,k+1)$, $p(c,k+1)$ are sides of an obtuse-angled triangle and their geometric means are sides of a right-angled triangle, then 2k is not an integer.

The rest of the formal part of this essay pertains to that.

Proof-ideas

Suppose the opposite: k, k+1 are two consecutive integers where $k \geq 1$, $a > b > c > 0$ are pairwise relatively prime integers and sides of an acute-angled (2k)-potent scalene triangle such that $p(a,k)$, $p(b,k)$, $p(c,k)$ are sides of an acute-angled triangle, $p(a,k+1)$, $p(b,k+1)$, $p(c,k+1)$ are sides of an obtuse-angled triangle and their geometric means $p(a,(2k+1)/2)$, $p(b,(2k+1)/2)$, $p(c,(2k+1)/2)$ are sides of a right-angled triangle.

We note first that the sides of the right-angled triangle are not all integers. For if p(a,(2k+1)/2) = p(p ,2)+p(q ,2), p(b,(2k+1)/2) = 2pq or p(p,2)-p(q,2), and p(c,(2k+1)/2) = p(p,2)-p(q,2) or 2pq, then a = r(p(p,2)+p(q,2),(2k+1)/2), b or c = r(2pq,(2k+1)/2) and c or b = r(p(p,2)-p(q,2),(2k+1)/2) where p, q are relatively prime integers not both odd and p > q. By Diophantus's proof, p(p(p,2)+p(q,2),2), p(2pq,2) and p(p(p,2)-p(q,2),2) are squares of integer sides of a right-angled triangle and squares of a primitive triple. But 2k+1 ≥ 3 is odd: a, b, c cannot all be integers; a contradiction. This complements *Prop. 3.*

Second, since the right-angled triangle is, considered side by side, the geometric mean between the obtuse-angled triangle (p(a,k+1),p(b,k+1),p(c,k+1)) and the acute-angled triangle (p(a,k),p(b,k),p(c,k)), it is also the geometric mean between the triangle (p(a,2k),p(b,2k),p(c,2k)) which is obtuse-angled by *Lemma 3*, and the acute-angled triangle (a,b,c). By the same token, the right-angled triangle is the geometric mean between the obtuse-angled triangle (p(a,2k+1),p(b,2k+1),p(c,2k+1)) where, as FLT fails, the obtuse apex angle = 180°, and the acute-angled triangle (1,1,1); and by the counterassumption FLT fails if (p(a,(2k+1)/2),p(b,(2k+1)/2),p(c,(2k+1)/2)) is a right-angled triangle, and conversely. One can accept the obtuse-angled triangle with the apex angle = 180° as an extreme case of obtuse-angled triangles, but the acute-angled triangle (1,1,1) is equilateral; it indicates, in the right-angled triangle assumed to be the geometric mean between two consecutive integer potencies of a *scalene* triangle (a,b,c), one of the potencies being an acute-angled triangle and the other an obtuse-angled triangle, *contradictio in adjecto.*

Third, we now capitalize on the *contradictio in adjecto* discovered, relating the geometric means (p(a,k+1/2), p(b,k+1/2), p(c,k+1/2)) to the harmonic and arithmetical means of (p(a,k),p(b,k),p(c,k)) and (p(a,k+1),p(b,k+1),p(c,k+1)).

By the Pythagorean "most perfect, or musical, proportion" [7:86],(where A = p(a,k+1)+p(a,k), B = p(b,k+1)+p(b,k), C = p(c,k+1)+p(c,k)):

p (a,k) : 2p(a,2k+1)/A = A/2 : p(a,k+1) where p(a,k) < p(a,k).2a/(a+1) < p(a,k).(a+1)/2 < p(a,k+1),

p (b,k) : 2p(b,2k+1)/B = B/2 : p(b,k+1) where p(b,k) < p(b,k).2b/(b+1) < p(b,k).(b+1)/2 < p(b,k+1),

p (c,k) : 2p(c,2k+1)/C = C/2 : p(c,k+1) where p(c,k) < p(c,k).2c/(c+1) < p(c,k).(c+1)/2 < p(c,k+1),

for all acute-angled scalene integer-side triangles (a,b,c), a > b > c > 0.

By the definitions of harmonic, geometric and arithmetical

means [7:87],

$p(a,k).2a/(a+1) : p(a,k+1/2) = p(a,k+1/2) : p(a,k).(a+1)/2$
where $p(a,k).2a/(a+1) < p(a,k+1/2) < p(a,k).(a+1)/2$,
$p(b,k).2b/(b+1) : p(b,k+1/2) = p(b,k+1/2) : p(b,k).(b+1)/2$
where $p(b,k).2b/(b+1) < p(b,k+1/2) < p(b,k).(b+1)/2$,
$p(c,k).2c/(c+1) : p(c,k+1/2) = p(c,k+1/2) : p(c,k).(c+1)/2$
where $p(c,k).2c/(c+1) < p(c,k+1/2) < p(c,k).(c+1)/2$,

for all acute-angled scalene integer-side triangles (a,b,c), $a > b > c > 0$.

Taking now new harmonic and arithmetical means (of second degree) of the previous harmonic and arithmetical means of p(a,k+1), p(a,k); p(b,k+1), p(b,k); p(c,k+1), p(c,k), and so on, *ad infinitum*, one will obtain ever better approximations from below and from above, to the geometric means p(a,k+1/2), p(b,k+1/2), p(c,k+1/2). As the geometric means are, by counter-assumption, sides of a right-angled triangle, all harmonic means of the same degree must be sides of acute-angled triangles and all arithmetical means of the same degree sides of obtuse-angled triangles.

In order to establish a contradiction it is sufficient to prove that any of the triangles with harmonic means as sides is either right-angled or obtuse-angled, or any of the triangles with arithmetical means as sides is either right-angled or acute-angled.

As *Prop. 4* already proves FLT for all even exponents, it suffices in fact to prove that any of the triangles with harmonic means (of the same degree N) as sides, say $(p(a,\sigma\hat{}1), p(b,\sigma\hat{}2), p(c,\sigma\hat{}3))$ where $1 \leq k < \sigma\hat{}1 < \sigma\hat{}2 < \sigma\hat{}3 < k+1/2$, is either right-angled or obtuse-angled. By *Defs. 2–3* that is the case when $p(a,2\sigma\hat{}1) \geq p(b,2\sigma\hat{}2) + p(c,2\sigma\hat{}3)$. We shall claim that in our proof reconstruction of *Prop. 5* (below). Note that the claim implies $p(a,2\sigma\hat{}1) > p(b,2\sigma\hat{}1) + p(c,2\sigma\hat{}1)$, because $p(b,\sigma\hat{}2) > p(b,\sigma\hat{}1)$ and $p(c,\sigma\hat{}3) > p(c,\sigma\hat{}1)$. This additional implication is vital.

Given a,b,c and the degree N, the exponents $\sigma\hat{}1$, $\sigma\hat{}2$, $\sigma\hat{}3$ are computable, of course, but actually the harmonic means of degree N will suffice. For N = 1,2,3 the harmonic means between p(r,k) and p(r,k+1) are (where H = r+1 and G = p(H,2)+4r):

$2p(r,k+1)/H < 4p(r,k+1)H/G < 8p(r,k+1)HG/(p(G,2)+rp(4H,2)) < \ldots < m\hat{}(N,harm)(p(r,k+1),p(r,k)) < m\hat{}geom(p(r,k+1),p(r,k))$; $r > 1$, $k \geq 1$ are integers, $N > 3$ indicates the degree of the harmonic mean, and increases *ad infinitum* (notation is for brevity).

Perhaps we must mention in passing also the Golden Section. Note that if we rewrite $p(a,2k+1) = p(b,2k+1) + p(c,2k+1)$ as $p(a,2k+1)/p(b,2k+1) - p(c,2k+1)/p(b,2k+1) = 1$, three mutually exclusive, necessary conditions for a non-trivial solution are

obtained. Either $p(a,2k+1) : p(b,2k+1) = \mu = (\sqrt{5}+1)/2 = p(b,2k+1) : p(c,2k+1)$ or $p(a,2k+1) : p(b,2k+1) > \mu > p(b,2k+1) : p(c,2k+1)$ or $p(a,2k+1) : p(b,2k+1) < \mu < p(b,2k+1) : p(c,2k+1)$. The first case (*sectio aurea*) will not do: because a, b are relatively prime integers and the (2k+1)th powers in G.P., (c) is not an integer, by *Elementa* VII.27 and IX.16.

In the second case, $a:b > b:c$ and, since $a > b > c > 0$, $(a-b) > (b-c)$. Thus $(a-b) \neq (b-c)$, which is part of Goldziher's result [4]. In the third case, however, all the three alternatives in the relation of (a-b) to (b-c) are possible.

The first case (*sectio aurea*) being excluded, it is clear that in the second and third case $p(a,2k+1)$, $p(b,2k+1)$, $p(c,2k+1)$ must be three consecutive terms of the Fibonacci series (which Fermat knew), Lucas series, or of some other similar series with other initial numbers. Hence FLT now reads: there are no three consecutive terms all (2k+1)th powers, in any such series. What complicates the matter is the start. For in the third case even the first three terms of an appropriate series will do.

Now, the series similar to the Fibonacci series have been studied in particular by D.H. Lehmer [cf. 6:148], and there are studies on the Fibonacci numbers that are powers, but it is not fair to assume that Fermat would have anticipated them. [Yet he must have studied *sectio aurea* like almost all of his contemporaries and predecessors.]

In search for the hidden lemmas

What could a musician like Lully do without an enchanting theme? What about a master navigator like Pytheas on unknown seas without his teacher's device for finding out his bearings [cf. 11]? For a mathematician, lemmas have the same vital importance. Blessed are those whose proofs succeed on the wings of "hidden lemmas", for they do not know what they are doing. But there lies a curse on the others: the "hidden lemmas" must be found and, with luck, refined.

As to the refinenemt of "hidden lemmas", Fermat may have learned it from Diophantus' use of *regula falsi* (e.g. at *Arithmetica* IV.15,37). It cannot be doubted, either, that Fermat was in constant search for "hidden lemmas". For he had tried to restore Apollonius' lost *Plane loci* on the basis of Pappus' account [7,II:185], and commented on statements probably from Diophantus' lost *Porisms* (e.g. at *Arithmetica* III.19;IV.29,30;V.9,11,14).

As for *Prop. 5*, what is needed is a lemma that will set

bounds to the ratio $(p(a,k+1/2)-p(a,\sigma\hat{}1)) : [(p(b,k+1/2)-p(b,\sigma\hat{}2)) + (p(c,k+1/2)-p(c,\sigma\hat{}3))]$, for computations show that it increases when N increases (see *Porism 2* below).

It turns out, however, that a very modest lemma will suffice here.

Lemma 8. If $p(a,2k+1) = p(b,2k+1) + p(c,2k+1)$ has an integer solution $a > b > c > 0$, $k \geq 1$, then $1 > p(a,(2k+1)/2) : p(b,(2k+1)/2) + p(c,(2k+1)/2) > 1 : \sqrt{2}$.

As for proof, if $p(a,2k+1) = p(b,2k+1) + p(c,2k+1)$, then $p(a,(2k+1)/2) = r(p(b,2k+1) + p(c,2k+1),2)$. The upper and lower bounds are proven indirectly. If $r(p(b,2k+1) + p(c,2k+1),2) \geq p(b,(2k+1)/2) + p(c,(2k+1)/2)$, then $p(b,(2k+1)/2).p(c,(2k+1)/2)$ or the geometric mean of $p(b,2k+1)$, $c(p,2k+1) \leq 0$; a contradiction because $a > b > c > 0$. If $p(b,(2k+1)/2) + p(c,(2k+1)/2) \geq \sqrt{2}.r(p(b,2k+1) + p(c,2k+1),2)$, then $p(b,(2k+1)/2).p(c,(2k+1)/2) : (1/2).(p(b,2k+1)+p(c,2k+1))$ or the ratio of the geometric mean of $p(b,2k+1)$, $p(c,2k+1)$ to their arithmetical mean ≥ 1; a contradiction because the geometric mean is lesser than the arithmetical mean. QED

Thus the claim is proven. We can restrict ourselves to triangles where it is.

Now, *Lemma 8* actually reduces the number of such acute-angled (2k)-potent scalene triangles with integer sides, on which FLT could fail. In contradistinction to the previous *Lemmas 1-6*, however, *Lemma 8* focuses directly on the (2k+1)th powers of a,b,c. Obviously *Lemma 8* is an implication of a more general proposition. Its full potential can best be exploited by conducting the proof of *Prop. 5* in such a manner that *Lemma 8* serves as a necessary and sufficient condition.

Where did the great mathematicians find their lemmas, then? There is one general source, the collections of propositions called porisms by the ancients. In Fermat's case, we already noted his likely contact with Diophantus' *Porisms*, but it is even more intimate with Euclid's. To quote Heath [7,I:435], "The great Fermat (1601-65) gave his idea of a 'porism', illustrating it by five examples which are very interesting in themselves; but he did not succeed in connecting them with the description of Euclid's *Porisms* by Pappus, and, though he expressed a hope of being able to produce a complete restoration of the latter, his hope was not realized."

As Euclid's *Porisms* remain lost even today, we must turn to the commentators, Pappus and Proclus. From Pappus *Collection* we gather that the enunciations of porisms are contracted and in fact comprehend many propositions in one enunciation [VII:648-60]. This is corroborated by his examples [7,I:432-433]. Pappus

and Proclus (*Commentary on Euclid* I: 212,14;301,22) also agree on the intermediate status of porisms between theorems and problems. Porisms, therefore, are useful in two directions.

In accordance with these views we conjecture that Fermat made use of three porisms which turn out to help prove *Prop. 5.*

Porism 1. $1 > r(b+c,2) : (r(b,2)+r(c,2)) > 1 : \sqrt{2}$; this implies *Lemma 8.*

For the other two porisms, we must agree on the common notation.

Let E^1 = r(p(a,2k+1),2) - m^(N,harm)(p(a,k+1),p(a,k)) and m^(N,harm)(p(a,k+1),p(a,k)) = p(a,σ^1)

Let E^2 = r(p(b,2k+1),2) - m^(N,harm)(p(b,k+1),p(b,k)) and m^(N,harm)(p(b,k+1),p(b,k)) = p(b,σ^2)

Let E^3 = r(p(c,2k+1),2) - m^(N,harm)(p(c,k+1),p(c,k)) and m^(N,harm)(p(c,k+1),p(c,k)) = p(c,σ^3)

Porism 2. If $a > b > c > 0$, then E^1 > E^2 > E^3 > 0. The order of the exponents of a, b, c depends on the coefficients of p(a,k), p(b,k), p(c,k) in the harmonic means of a finite degree N, for which it holds

$2a/(a+1) : a < 2b/(b+1) : b < 2c/(c+1) : c$, $4(a+1)a/(p(a+1,2)+4a) : a < 4(b+1)b/(p(b+1,2)+4b) : b < 4(c+1)c/(p(c+1,2)+4c) : c$, etc. Thus $k < σ^1 < σ^2 < σ^3 < k + 1/2$. Because in the geometric means the exponent is the same $(k + 1/2)$ and $a > b > c > 0$, the differences E^1 > E^2 > E^3. Finally, because even the harmonic mean of a finite degree N, of p(c,k+1), p(c,k), is smaller than the geometric mean of p(c,k+1), p(c,k), E^3 > 0 when c > 0. Therefore E^1 > E^2 > E^3 > 0. QED

It may be noted that *Porism 2* and *Lemma 8* (or *Porism 1*) together set bounds to the ratio E^1 : (E^2+E^3) which increases when the degree N does.

Porism 3. p(p(b,(2k+1)/2) - (E^2+E^3)/2,2) + p(p(c,(2k+1)/2) - (E^2+E^3)/2,2) > p(p(b,(2k+1)/2) - E^2,2) + p(p(c,(2k+1)/2) - E^3,2).

Using the same notation as in *Porism 2*, the proof of *Porism 3* becomes straightforward computation. Raising the bracketed terms to square and subtracting, we obtain 2(E^2 - E^3)(p(b,(2k+1)/2) - p(c,(2k+1)/2)) > p(E^2 - E^3,2). Dividing both sides by (E^2 - E^3) and recalling how E^2, E^3 were characterized, we finally obtain p(b,(2k+1)/2) - p(c,(2k+1)/2) > m^(N,harm)(p(c,k+1),p(c,k)) - m^(N,harm)(p(b,k+1),p(b,k)), which holds good because the left hand side is greater than zero and the right hand side less than zero when b > c > 0. QED. This result illustrates the usefulness of porisms in problems.

Note that *Porism 3* could have been enunciated in a different way also, e.g. as *Porism 3** : the sum of the left-hand side of the

two squares is the greatest one because the arithmetical mean of E^2 and E^3 has been subtracted. In that case the problem becomes one of determining the maximum sum, where Fermat could have used his method of maxima and minima with the characteristic auxiliary variable E.

It is worthwhile to conclude this section on porisms with a philosophical note. Every mathematician interested in FLT must have faced the fact that FLT is an extremely isolated proposition. Now, a proof is ordinarily so conducted - no matter which method is employed - that either an absurdity, contradiction or compatibility is established with respect to previously proved propositions or accepted axioms, although even the accepted rules of inference might do. As a rule, the absurdity, contradiction or compatibility can ultimately be referred to an axiom. The porisms, despite their great generality, as in *Porism 1*, are not axioms, however. When a contradiction (or absurdity, or compatibility) is established with respect to a porism, the proof might be considered less than perfect by our standards. Fermat must have thought differently, following Pappus who had stated about Euclid's *Porisms* that "these porisms embody a theory subtle, natural, necessary, and of considerable generality, which is fascinating to those who can see and produce results" [7,I:431, in a sentence bracketed by Hultsch who edited the *Collection*, but in full agreement with Pappus' words and tenor in the same context]. At any event, Fermat's use of *Lemma 8* (which rests on *Porism 1*) and *Porisms 2-3*, should be considered as an attempt to break the isolation of FLT. In this respect it is to be compared with Yoichi Miyaoka's reliance on the principles of avantgarde theoretical physics, in his attempt at a proof of FLT (1988).

As to *Porism 2*, it is far less precisely worded than the modern taste demands. That is due to the historical restriction: Fermat did not use the limit concept, contrary to what earlier interpretations often claim (see below). Finally, in *Porism 3*, Fermat may well have tried the geometric mean instead of the arithmetical mean. But that leads to a blind alley.

Proof Reconstruction

In our proof reconstruction of *Prop. 5* by Fermat's methods we employ *reductio ad absurdum* and attempt to prove that if (a,b,c) is a (2k)-potent acute-angled scalene triangle with integer sides such that $a > b > c > 0$ are relatively prime, p(a,k), p(b,k), p(c,k) are sides of an acute-angled triangle, p(a,k+1), p(b,k+1), p(c,k+1) are sides of an obtuse-angled triangle, their geometric means

p(a,k+1/2), p(b,k+1/2), p(c,k+1/2) are sides of a right-angled triangle and 2k+1 ≥ 3 is an integer, then also, for some (high) finite N, their harmonic means of degree N are sides of a right-angled or of an obtuse-angled triangle.

This is equivalent, by *Defs. 2-3*, to proving that

(1) p(m^(N,harm)(p(a,k+1),p(a,k)),2) ≥ p(m^(N,harm)(p(b,k+1),p(b,k)),2) + p(m^(N,harm)(p(c,k+1),p(c,k)),2),

which is absurd because the geometric and the harmonic means of degree N of the same numbers, cannot both be sides of a right-angled triangle, nor the geometric means sides of a right-angled triangle and the harmonic means of degree N of the same numbers sides of an obtuse-angled triangle. For if the geometric means are sides of a right-angled triangle, as they are in case FLT fails for odd exponents, then the harmonic means of any degree of the same numbers must be sides of an acute-angled triangle; and if *not*, an absurdity follows.

Proof.

(2) If the geometric means are sides of a right-angled triangle, then p(a,2k+1) = p(b,2k+1) + p(c,2k+1) and conversely, and p(a,(2k+1)/2) = r(p(b,2k+1) + p(c,2k+1),2).

(3) In the notation of *Porisms 2-3*, by (2), the proof of (1) is equivalent to proving p(r(p(b,2k+1) + p(c,2k+1),2) - E^1,2) ≥ p(p(b,(2k+1)/2) - E^2,2) + p(p(c,(2k+1)/2) - E^3,2).

(4) By *Porism 3*, (3) holds if the following does p(r(p(b,2k+1) + p(c,2k+1),2) - E^1,2) ≥ p(p(b,(2k+1)/2) - (E^2+E^3)/2,2) + p(p(c,(2k+1)/2) - (E^2+E^3)/2,2).

(5) We introduce an auxiliary variable 1 > E > 0 (which may be imagined very small, in accordance with a high value for N) such that for two positive numbers r,s it holds that rE = E^1 and sE = (E^2+E^3). We next determine the bounds to the ratio (r:s) by a method of Fermat.

(6) Substituting rE for E^1 and sE for E^2+E^3, by (5), we can rewrite (4): p(r(p(b,2k+1) + p(c,2k+1),2) - r E,2) ≥ p(p(b,(2k+1)/2) - sE/2,2) + p(p(c,(2k+1)/2) - sE/2,2).

(7) Raising to square and computing we obtain from (6) the following: s(p(b,(2k+1)/2) + p(c,(2k+1)/2)) + r^2E ≥ s^2E/2 + 2r.r(p(b,2k+1) + p(c,2k+1),2).

(8) Let E = 0. This operation is characteristic to Fermat; E is *not* let "approach to zero" as he did not operate with the concept of the limit. The same result is obtained if, instead of (8), the following steps are taken. Both sides of (7) are divided by the coefficient of E. The proof then bifurcates as either $2r^2 < s^2$ or $2r^2 > s^2$; if $2r^2 = s^2$, a short-cut leads to (9). Finally, in both branches, let E = 0, and (9) is obtained.

(9) By (8) from (7), s : 2r ≥ r(p(b,2k+1) + p(c,2k+1),2) :

p(b,(2k+1)/2) + p(c,(2k+1)/2).

(10) Now (4) and (3), and hence (1), will be proven by *Lemma 8* iff $1 > s : 2r \geq r(p(b,2k+1) + p(c,2k+1),2) : (p(b,(2k+1)/2) + p(c,(2k+1)/2)) > 1 : \sqrt{2}$, that is iff $1 : \sqrt{2} > (p(b,(2k+1)/2) + p(c,(2k+1)/2)) : 2r(p(b,2k+1) + p(c,2k+1),2) \geq r : s > 1 : 2$.

We now study E^1, E^2, E^3 within the bounds 1, 1:2 in preparation to proof.

(11) Retranslating (10) into the notation of *Porisms 2-3* by (5), we obtain the bounds $1 : \sqrt{2} > E^1 : (E^2 + E^3) > 1 : 2$ from (10). Thus $2E^1 > E^2 + E^3 > \sqrt{2}E^1$ and *a fortiori* $E^2 + E^3 > E^1$. By *Porism 2*, $E^1 > E^2 > E^3$; and $E^3 > 0$ because the harmonic mean of any finite degree N, is smaller than the geometric mean of p(c,k+1), p(c,k) when $c > 0$. Therefore E^1, E^2, E^3 are sides of a scalene triangle by *Elem. I*.20; (cf. the proof of *Prop. 1*).

(12) When $1 : \sqrt{2} > E^1 : (E^2 + E^3) > 1 : 2$, the upper bound indicates that the triangle (E^1, E^2, E^3) is acute-angled, for $p(E^2 + E^3,2) > 2p(E^1,2)$. Most (but not all) scalene, rational-side, acute-angled triangles satisfy a similar condition.

(13) Now, $(s:2r)$ and $(r:s)$ are interchangeable in steps (5-10) of the proof-scheme. If, *mutatis mutandis*, $1 > r:s > 1:\sqrt{2} > s:2r > 1:2$ in (10), then, retranslating into the notation of *Porisms 2-3* by (5), $1 > E^1:(E^2 + E^3) > 1:\sqrt{2} > (E^2 + E^3):2E^1 > 1:2$. Hence now $2p(E^1,2) > p(E^2 + E^3,2)$. The rest of scalene, rational-side, acute-angled triangles, which did not satisfy the condition of (12), satisfy this type of condition. For if $2p(E^1,2) = p(E^2 + E^3,2)$, all sides cannot be rational. (Note that Euclid takes a wider view of 'rational' [7,I:403]).

(14) It is worthwhile to express the findings of (12) and (13) in an analogical way. For all scalene, rational-side, acute-angled triangles either $1 > (E^2+E^3):2E^1 > 1:\sqrt{2} > E^1:(E^2+E^3) > 1:2$ or $1 > (E^2+E^3):\sqrt{2}E^1 > 1:\sqrt{2} > E^1:\sqrt{2}(E^2+E^3) > 1:2$. *A fortiori*, these two sets of bounds cover all scalene, integer-side, acute-angled triangles with the same ratios of their sides considered in the same manner as above.

(15) On the bounds, *contradictiones in adjecto* spring up. An example for Fermat could have been Erycinus' paradoxes in Pappus' *Collection* III, Third section. For if $E^1:(E^2+E^3) = 1:\sqrt{2}$, then E^2, E^3 could be the sides and E^1 the diagonal of a square; if it is = 1:2, then E^1, E^2, E^3 could be sides of an equilateral triangle; and if it is = 1, then E^1, E^2, E^3 could be sides of an obtuse-angled triangle in the extreme case that the apex angle is 180°. It is reasonable to think that Fermat wanted to avoid such paradoxes, just as modern mathematicians do.

(16) We set first $aE = rE$, $(b+c)E = sE$ in (5), and using the

proof scheme (5)-(10) thus prove $1 > (b+c):2a \geq r(p(b,2k+1) + p(c,2k+1),2):(p(b,(2k+1)/2) + p(c,(2k+1)/2)) > 1:\sqrt{2}$ at step (10) by *Lemma 8*. Here $a:(b+c) \leq E\hat{}1:(E\hat{}2+E\hat{}3)$ and $\sqrt{2}a < b+c$.

(17) Second, we set $aE = rE$, $\sqrt{2}(b+c)E = sE$ in (5), and using the proof scheme (5)-(10) thus prove $1 > (b+c):\sqrt{2}a \geq r(p(b,2k+1) + p(c,2k+1),2):(p(b,(2k+1)/2) + p(c,(2k+1)/2)) > 1:\sqrt{2}$ at step (10) by *Lemma 8*. Here $a:\sqrt{2}(b+c) \leq E\hat{}1:\sqrt{2}(E\hat{}2+E\hat{}3)$ and $\sqrt{2}a > b+c$.

(18) Thus the proofs (16), (17) cover all scalene, integer-side, acute-angled (2k)-potent triangles (a,b,c) divisible in two subsets as in (14).

(19) Hence (1) is proven, an absurdity established, and *Prop. 5*, that is to say, FLT for odd exponents $2k+1 \geq 3$, proven 'truly remarkably'. QED

(20) Finally, we explicate the core of (16), (17). Consider first $1 > (b+c):2a \geq p(a,(2k+1)/2):(p(b,(2k+1)/2) + p(c,(2k+1)/2)) > 1:\sqrt{2}$. Recalling our heuristic analysis and the application of *Elem.* I.21 we put the right-angled triangle with the geometric means as its sides on the same scale with the triangle (a,b,c), with (a) as the common base (cf. *Fig.1*). Thus $1 > (b+c):2a \geq a:(\beta+\tau) > 1:\sqrt{2}$, where $\beta = r(p(b,2k+1)/p(a,2k-1),2)$ and $\tau = r(p(c,2k+1)/p(a,2k-1),2)$. Here β and τ are (if FLT fails for odd exponents) sides of a right-angled triangle with the base (a); the right apex angle is contained by the semicircle with the radius (a/2). The maximum value for the sides of such a right-angled triangle is $\sqrt{2}a = 2r(p(a/2,2)+p(a/2,2),2) > \beta+\tau$.

This maximum value can be obtained in many ways (e.g. the apex of the right-angled triangle, say A*, is the only common point of the semicircle BA*C and an ellipse of which B and C are the foci). It was known to the ancient geometers, too. But, as it happens, it is also an immediate consequence of a result of Fermat, proved by his method of maxima and minima using the auxiliary variable E (*Oeuvres* I:133-4,147-51;III:121-2; cf. Supplement:120-125 and Boyer 1949:155-6).

The core of (16) can be expressed as $(b+c)(\beta+\tau) \geq 2a^2$ or $r((b+c)(\beta+\tau),2) \geq \sqrt{2}a$. On the other hand, the relationship $2a > b+c > \sqrt{2}a > \beta+\tau$ holds good for the triangle (a,b,c) in (16). Thus the core of (16) can be expressed in words in this way : $\sqrt{2}a$, the maximum value for $\beta+\tau$, is equal or smaller than the geometric mean of (b+c) and $(\beta+\tau)$. Because the geometric mean is greater than the harmonic mean and smaller than the arithmetical mean, it can be expanded into this: the maximum value for $\beta+\tau$ is smaller than the arithmetical mean, smaller than or equal with the geometric mean, and greater than or equal with the harmonic mean, of (b+c) and $(\beta+\tau)$.

The core of (17) can be explicated in a similar manner. The

core of (17) can be expressed as $(b+c)(\beta+\tau) \geq \sqrt{2}a^2$ or $r((b+c)(\beta+\tau),2) \geq r((\sqrt{2}a)a,2)$. On the other hand, the relationship $2a > \sqrt{2}a > b+c > \beta+\tau$ or $\sqrt{2}a > r((\sqrt{2}a)a,2) > r(a(b+c),2) > r(a(\beta+\tau),2)$ holds good for the triangle (a,b,c) in (17). Again expansion and verbal summary can be given. We conjecture that such expansions and geometric considerations were the genesis of FLT.

Thus triangles, and in particular acute-angled triangles with scalene integer sides, pervade our reconstruction of Fermat's proof from *Prop. 1* to *Prop. 5. Defs. 1-4* and *Lemmas 1-8* adumbrate, in our opinion, his borders to this newly discovered kingdom in a valley between barren peaks.

Fermat's auxiliary variable E first appears in his method for determining maximum and minimum values (1638), but in a letter to Roberval (1636) Fermat mentions that already in about 1629 he was in possession of the method. It is quite possible, therefore, that Fermat made use of the same technique of an auxiliary variable E in 1637 in the proof of FLT. Fermat probably learned it from the ancient geometers [cf.12:120-122] or from Pappus (see the analysis of Prop. 12 at *Collection*, Book IV). Pappus' influence on the method of maxima and minima is clear, however.

Pappus had spoken of a "minima et singularis proportio" which led Fermat to consider the fact that in a problem that in general has two solutions, the minimum and maximum value gives only one solution (as Fermat explained in a letter 1643; cf. P.G. Giovannozzi, "Pierre Fermat. Una lettera inedita", *Archivo di Storia della Scienzia*, I:137-140 and Boyer 1949:155-6). The argument in Fermat's first application of this method (1638) runs like this: Given a line segment (a), mark off the distance (x) from one end. The area on the segments is $A = x(a-x)$. If one marks off the distance (x+E), however, the area is $A = (x+E)(a-x-E)$. For the maximum area the two values will be the same, as Pappus had noted, and the points marking off the distances (x) and (x+E) will coincide. Now, setting the two values of A equal and letting $E = 0$, the result is $x = a/2$. This is the result mentioned above at step (20); we consider it a piece of circumstantial historical evidence for our proof reconstruction.

On the other hand, also Pappus' account of the ancient heuristic method of analysis and synthesis [cf.12] lends some support to our reconstruction. Note, in particular, his general description: "Now analysis is the way from the *zetoumenon* (what is sought) - as if it were admitted - through its *akoloutha* (concomitants) in order to something admitted in synthesis. For in analysis we suppose the *zetoumenon* to be already done, and we inquire from what it results, and again what is the anteced-

ent of the latter, *until we on our backward way light upon something already known and being first in order*. And we call such a method analysis, as being a solution backwards" (*Collection* VII, The method of analysis, 8; tr. Jaakko Hintikka & Unto Remes, based on Ivor Thomas and Thomas Heath; *The method of analysis. Its geometrical origin and its general significance*, 1974, Dordrecht, Boston; cf. [12:116]). "Something already known and being first in order" is a very appropriate notion when the triangles enter the proof again in steps (11) to (13). Assuming that the proof was cast in the mould of a foregoing heuristic analysis, (11) to (13) may well remind us of the turning point of analysis, after which the synthesis will follow. And when the initial triangles reappear into the proof in steps (16) and (17), this may well account for Fermat's own characterization of the surprise element.

But is (1)-(20) really a proof and not only heuristics ? One reason for Fermat to drop his note on a "truly remarkable proof", was probably Descartes' unduly bitter criticism against the method of maxima and minima. Moreover, Fermat also believed that his method of tangents was an application of his method for maxima, but facing the criticism, was unable to specify which quantity he was maximizing (cf. Boyer 1949:158). That type of criticism can be directed against the reconstructed proof also. After all, the maxima of $(\beta+\tau)$ needed, play only a minor role. Perhaps it would be better to speak about an independent "E-method" or of a proof scheme in steps (5)-(10), as we have done.

Descartes' criticism and Fermat's hallmark: solution of geometrical problems rather than generalization of methods, must be weighed against each other. We are inclined to answer our main question in the affirmative. With respect to the standards of his time, Fermat was justified in claiming that he had invented "a truly remarkable proof" to FLT.

The steps (11)-(13) and (16), (17) mark a pregnant moment in the history of mathematics. It may strike one that the auxiliary variable E was brought in by a postulate, but that was (and still is) the style of the Old Masters. It may intrigue one to note that the explication (20) sounds very prosaic in comparison with the formulation of FLT, but even if the problem was a fruit of a skilful device, it advanced the art. Sometimes the man finds his problem, sometimes the problem its man. But in the last analysis, what matters is the problem. And Fermat did discover a most unexpected symmetry between the microcosm and the macrocosm of triangles, between the triangles (E^1,E^2,E^3) and (a,b,c). His proof scheme was his microscope and the theory of means his

telescope.

But it is not Fermat's fault that the course in the history of mathematics differed from Fermat's pivotal idea, which was founded on discrete entities at all levels of mathematical existence. Nor can he be blamed for the superficial applications of his Principle of the Least Time, the physical analogue of FLT.

Ek Panton Hen Kai Ex Henos Panta

Our reconstruction of Fermat's reasoning follows a *via negativa*, befitting a Requiem to his Greek predecessors. First, by Fermat's extant proof using his Method of Infinite Descent (the principle of well-ordering of natural numbers), if FLT fails, then the exponent $n \neq 4$. Second, by *Prop. 4*, n is not an even integer. Third, by *Prop. 5* using *reductio ad absurdum*, n is not an odd integer. A truly constructivist proof might be required to yield also the (non-integer) value for n, depending on the perimeter $2h = a+b+c$ and on how far apart a,b,c are from one another in each case. Fermat, if anybody, must have required that type of perfection in a proof, "an early Bourbakian" as he is.

The fact that a sentence (e.g. FLT put into a universal garb [15:216]) is a theorem depends essentially on how rich is the given collection of axioms. Similarly, one's method of establishing a contradiction depends on it. Truth and Contradiction are Twin Daughters of Time.

It seems that Fermat had not much of a choice. We think he considered first (c. 1637–1640) that his *reductio ad absurdum* is enough, and did not distinguish the Janus faces of an actual construction from one another. But around 1640 he did it: in a letter to Mersenne (where the problem was proposed to Frénicle de Bessy), and in another in 1657 to Digby (where it was proposed to Wallis and Brouncker), Fermat no more mentions his "remarkable proof" [15:2]. We conjecture that by then he fully realized the utter isolation of the theorem $p(a,2k+1) \neq p(b,2k+1 + p(c,2k+1)$. In fact he had proven a useful theorem [6:Theorem 366]. But he probably had not determined the (non-integer) value for n in its various cases in $p(a,n) = p(b,n) + p(c,n)$, as a true constructivist might want to.

Nevertheless, Fermat had discovered an interesting line of attack that made use of the ancient theory of means (Nicomachus and Pappus have preserved altogether eleven different means, cf. [7:87]). And he had realized the possibilities embedded in Erycinus's Paradoxes about Euclid's *Elem.* 1.21, as well as the effectivity of Heron's approach in his *Metrics*, and Pappus's

account of heuristics in his *Collection*. All this was possible for him, because he had a direct access to Greek sources thanks to his interest in Greek philology. Fermat's Principle of the Least Time in optics, again, suggests another main trait in his intellect: Gallic sagacity. And it connects FLT with physics.

As for the constructivist proof of *Prop. 5*, we presume that Fermat tried to determine the value for a non-integer n and pursued perhaps the two other strategies of proof outlined above, until the end.

But even if he never succeeded in these self-imposed tasks (for he left no vestiges of his eventual labours), it is worthwhile to study his heuristics today. As Arpad Szabó wisely says about the general problem: "Ich bin überhaupt der Ansicht, dass nicht nur die mathematische Heuristik in den Dienst der historischen Forschung gestellt werden kann, sondern auch umgekehrt die historische Forschung die keimende Wissenschaft der Heuristik fördern soll" [16:482, N.19].

As for wider historical perspectives, Fermat's problem is intimately connected with two of Hilbert's Problems [8], viz. IV and X. In spite of Hamel's positive solution to the former already in 1901 and Ju.V. Matijasevic's and G.V. Cudnovskij's negative solution to the latter in 1970 [cf. 13], it is interesting to ponder about some historical contrafactuals, or rather "counter-eventuals". What would be missing in modern mathematics without Fermat's problem, for instance? What would have ensued from an early solution to it? What could follow now, and in the future?

These are, of course, very big questions. But at least some answers are possible. Thus it is worthwhile to repeat the early optical experiments within the context of a discrete point geometry (a fragment of the Euclidean geometry) where the Principle of the Least Time applies. It is generated by the Pythagorean triangular numbers [7:76-7] illustrated (in the two-dimensional case) by *Fig. 3*. The results may concern modern science, too.

Another case of interest is Hamilton's quaternions *v.* Fermat's FLT. Hamilton's quaternions were invented after frustrating efforts to find three-dimensional numbers that would possess the normal properties of real and complex numbers. It turned out that "the principle of permanence" could not be satisfied which requires that a number field should fulfill the following properties: (1) sequence of natural numbers can be identified, (2) criteria of rank can be established and (3) a scheme of addition and multiplication can be devised that will have the commutative, associative, and distributive properties of the natural operations (cf. Tobias Dantzig 1954(4):92-93).

Hamilton had to sacrifice the commutative law of multiplica-

tion and four dimensions were needed for the new numbers. As Kline puts it, geometric hindsight can show us that the rotation and the stretching or contraction of a given vector in physical space requires four parameters; three angles and a stretching factor. More formally, Frobenius has proved that "the only linear associative algebras with real coefficients (of the primary units), with a finite number of primary units, a unit element for multiplication, and obeying the product law are those of the real numbers, the complex numbers and real quaternions" (cf. Morris Kline 1972:779,793). Nevertheless, triples of integers are flexible in other respects. Although three-dimensional *vectors* are used in the vector analysis of modern physics or in mathematics as a special case of a n-dimensional linear algebra, it seems that the study of triangular numbers as separate entities has been sidestepped because of the historical process that led to the physics-oriented quaternions. Second, interpreted as integer-side *triangles*, they constituted Fermat's ontology of mathematics. Their metamorphosis into *furca crosses* we have barely alluded to (cf. *Fig.3*).

There is no need to dwell on the ancestry of mathematical atomism (at times distinguished from, at times merging together with, physical atomism). It may be recalled, however, that a very eloquent prologue was given in Zeno's "Stadion" and "Flying Arrow". In the subsequent dialogue, *atoms* or *indivisibles* were made use of in many ways by men like Democritus, Plato, Aristotle, Archimedes, Heron, Oresme, Nicholas of Cusa, Galileo, Kepler, Pascal, Huygens, Cavalieri, Torricelli and Roberval besides Fermat. With some notable exceptions, like Kronecker, atomism is rejected in modern analysis; its kingdom is number theory.

It is worthwhile, however, to get rid of a misunderstanding as to Fermat's use of the auxiliary variable E. In his method of the maxima and minima, Fermat's procedure is almost precisely the same now employed in the differential calculus, except that dx is substituted for E. No wonder, therefore, that Fermat's argument for his method is at times interpreted in terms of the limit concept (so that E becomes a variable quantity approaching zero; cf. for instance Duhamel's (1864) "Mémoire sur la méthode des maxima et minima de Fermat, et sur les méthodes des tangentes de Fermat et Descartes"). A much more reasonable interpretation is, however, that Fermat let E vanish in the sense of actually being zero (cf. Tannery 1902:344; Wallner 1904:122-123; Boyer 1949:154 ff.). We have adopted this interpretation throughout. For firstly, in so doing we need not ascribe to Fermat advanced notions not corroborated by his own words.

And secondly, this interpretation is in full agreement with Fermat's predecessors. To put it shortly: Fermat remained faithful to Greek Antiquity and refused (Boyer seems to contend that he was unable) to use the limit concept of continuous variables. Fermat also tried to avoid the concept of actual infinity, and used equalities, inequalities and pseudo-equalities (his own term is *adaequalitas*; see, e.g., *Oeuvres* I:133-79) instead.

Aware as we are of the endless paradoxes connected with infinite sets, it is good to consider the wisdom of Fermat's approach. But things were in the move: already 1656 in Wallis' *Arithmetica Infinitorum* actual infinity enters heuristics.

As to the present state of the art, we may recall that Paul Cohen's solutions (1963 and 1966) are not the end of Cantor's "Problem von der Mächtigkeit des Kontinuums (1878)" which Hilbert proposed as his first and foremost problem in 1900 (for the references, see [13:143-4]). The problem arises when the Continuum Hypothesis is to be proven consistent with, or independent of, the rest of the axiom systems with or without the Axiom of Choice in particular (that was introduced by Zermelo).

In 1940 Gödel proved the consistency of an axiom system (Σ) with the generalized Continuum Hypothesis and the Axiom of Choice. Other important results that paved the way for Cohen's work were Sierpinski's proof (1947) that the Axiom of Choice implies the well-orderedness of a set and conversely, and Specker's proof (1952) that the generalized Continuum Hypothesis implies the Axiom of Choice. Paul Cohen's solutions are relative to an axiom system of Zermelo and Fraenkel (ZF; in 1951 I.L. Novak showed that Σ and ZF are mutually consistent), whereas Cantor had presented the problem within the context of a non-axiomatic ("naive") set theory. Especially Cohen's proof (1966) that, with respect to Σ without the Axiom of Choice, it is consistent to say that the Continuum includes Dedekind's set, leads into the new problems of unreachable cardinal numbers and transfinite induction - to mention just a few of the new problems. Hilbert's final aim, the proof of consistency for entire mathematics, is not reached. On the contrary, one can easily visualize new Continuum Problems for systems other than Σ. In set theory, the real problem is consistency.

As for the analysis, then, the "rigorization" of analysis in the nineteenth century and the entire underlying theory of "the flowing world", calculus, have given birth to a horrendous collection of conceptual paradoxes. The effort to algebrize geometry and to remove motion and geometric intuition, was made at a great price. Were it not called mathematics, philosophers

would (and should) have attacked these developments.

Analysis is based on infinite sets and on the idea of comparing the number of elements in them, which presupposes the existence of actually infinite sets. But set theory, which is needed in the foundation of analysis, is as full of paradoxes as Zeno. What is the rigour gained, then?

The continuum of real numbers assumes that numbers are not really separate entities. The least thing one can say is that the concept of real number rests on hazy grounds when compared to natural number. Infinitesimals, infinitely small or large magnitudes have perplexed mathematicians, and rightly so, for centuries. There is every reason to believe that the present day definitions are considered satisfactory merely because of their expediency - calculus works well in practice. But that is exactly what Dedekind and Weierstrass could have said of their predecessors, like Eudoxus (assuming he was the author of Euclid's *Elements*, Book V, Def. 5) who relied on the concept of motion in geometry and on geometrical intuition.

What is more, continuous analysis requires or implies an old-fashioned and paradoxical cosmology, where both time and space are continuous and have no smallest or indivisible or even separate constituents. It is historically tied to the study of heavenly bodies flowing smoothly in space. Perhaps the strongest argument for that view comes from theology. An omnipresent, omnipotent and omniscient divinity or demon guarantees the continuous rotation of the celestial bodies.

Continuous analysis also produces pseudoproblems both in physics and in social sciences: problems that have nothing to do with the subject matter, but reflect the conceptual difficulties in the methodology. Most real world entities are best described with discrete or finite concepts, and best measured with finite yardsticks (for concrete examples, see Kasanen, E. "Dilemmas with infinitesimal magnitudes", *Journal of Economic Dynamics and Control* 1982; 4:295-301).

The geometric counterpart of natural numbers is a discrete-point geometry. Indeed, supposing there is a smallest distance in space and time, we can in principle number all points and moments using natural numbers alone. There are neither philosophical nor mathematical reasons for using numbers other than natural. Thus all the pseudo-problems of analysis can be removed from the theoretical level. But calculus may remain, of course, as a useful method for engineers. Had not Hamel been so fast (1901) in solving Hilbert's fourth problem (*Elem.* I.20 applies in Euclidean and elliptic geometries), it is possible that already then Fermat's discrete-point geometry could have been given

more serious reconsideration in mathematics, and its physical counterpart, the Principle of the Least Time, in physics.

We would like to conclude this essay by a prediction. As we see it, there are two choices available for our successors (and contemporaries). Either to strive for simplicity, natural numbers, harmonic world-view and for a discrete-point geometry both in the ontology of mathematics and in philosophical cosmology, or to continue to accept doubtful primary concepts, paradoxes, real numbers, chaotic world-view, and continuous analysis.

Our forecast on the progress of science is this: Great harmony in nature will be found, analysis based on natural numbers will take off (the pivotal metaphor of science being the digital computer instead of the old orrery), and a discrete cosmology will unite the basic forces, the basic constituent of nature, the basic forms of matter (triangles and *furca* crosses), and this will be based on atomistic, separate concepts. For a harbinger, look at the beauty and hierarchic depth of the fractals evolving by the Sea-Horse Bay.

Once the Greeks played with pebbles on sand and discovered the metamorphosis of triangles into *furca* crosses in the microcosm as well as in the macrocosm. Two millennia later, Fermat saw the same vision anew. Perhaps two millennia from now, after the Bomb, when a new Polar Star is seen heralding a new science, Dolphins will be playing with pebbles again?

A marginal note

What was Fermat's dessert in 1637, then? Why, indeed, could he conclude the proof at Step 19 in our proof reconstruction of *Prop. 5*? He had started from the counter-assumption that FLT fails for odd exponents, then transformed (1) by legitime means which include *Porisms 1-3*, his own 'E-method' and the ancient theory of means, and finally reached the transformed formulae (16) and (17). But (16) $1/2(b+c)(\beta+\tau) = a^2 = \beta^2+\tau^2$ and (17) $1/\sqrt{2}(b+c)(\beta+\tau) = a^2 = \beta^2+\tau^2$ with the equality sign do not say more than the counter-assumption! Thus the *reductio ad absurdum* succeeds if FLT fails for odd exponents, and it does not succeed if FLT does not fail for them. This is another version of Eubulides' (originally Epimenides') Liar Paradox, later known to Adam of Balsham and Schoolmen as one of the *insolubilia*, and also one of the main targets in Russell's, Gödel's and Tarski's philosophies of mathematics. In other words, Fermat's final strategy appears as a proof that it is impossible to solve the Diophantine equation $p(a,2k+1) = p(b,2k+1) + p(c,2k+1)$ where $a > b > c > 0$, $k \geq 1$. No wonder he called the proof 'truly remar-

kable'.

Moreover, if FLT fails for odd exponents, (16) and (17) with the equality signs give rise to second degree equations whose roots determine when the equilibrium holds. Let the roots be called (β^*,τ^*). For (16) they are $(a^2 \pm a.r(1/2(b+c)^2-a^2,2))$: (b+c) and for (17) $(1/\sqrt{2})(a^2 \pm r((b+c)^2-a^2,2))$:(b+c).

University of Tampere

NOTES

This essay is dedicated to the memory of our friends Prof. Hiromichi Takeda - the last Samurai in the fields of philosophy, and Prof. Tadao Yamada - humanist, scientist, musician. They were wonderful interpreters of the harmony of Kyoto, of the song of nature, and of the best aspects of Japan's culture.

* *Editor's note*: In this article, an option has been taken for a linear format. Thus the n-th power of x, x^n, is written p(x,n) and the n-th root of x is written r(x,n). x with subscript y is written x^y.

REFERENCES

The main biographical source is Mahoney [10] and the best overview of literature Ribenboim [15], with bibliographies and primary sources. Fermat's *Oeuvres*, ed. Tannery & Henry, Paris (1891) 1922.

[1] Bashmakova, I.G. "Diophante et Fermat", *Rev. Hist. Sci.* XIX:289-306.

[2] Bashmakova, I.G. & Slavutin, E.I. *Istoriya diofantova analiza ot Diofanta do Ferma*, Moscow, 1984 ("Nauka").

[3] Dickson, L.E. *History of the Theory of Numbers* II, Washington, 1920.

[4] Goldziher, K. "Hátvanyszamok Telbontása hátvanyszamok összegere", *Középiskolai Math. Lapok* 21:177-184 (1913); his result was rediscovered in 1952 by Mihaljinec and in 1969 by Rameswar Rao [15:69].

[5] Grünert, J.A. "Wenn n > 1, so gibt es unter den ganzen Zahlen von 1 bis n nicht zwei Werte von x und y, für welche, wenn z einen ganzen Wert bezeichnet, $x^n+y^n = z^n$ ist", *Archiv*

Math. Phys. 27:119-120 (1856).
[6] Hardy, G.H. & Wright, E.M. *An Introduction to the Theory of Numbers*, London, 1960 (reprint with corrections 1968).
[7] Heath, Th. *A History of Greek Mathematics* I-II, Oxford, 1921 (rep. 1965).
[8] Hilbert, D. "Mathematische Probleme", *Nachr. Ges. Wiss. Göttingen* 1900:253-297 = *Gesamm. Abh.* Bd. III, Berlin 1935:290-329.
[9] Inkeri, K. & Hyyrö, S. "Ueber die Anzahl der Lösungen einiger Diophantischer Gleichungen", *Ann. Univ. Turku.* A, I, 1964, No. 78:3-10.
[10] Mahoney, M.S. *The Mathematical Career of Pierre de Fermat*, Princeton, 1973 (Princeton UP).
[11] Maula, E., Kasanen, E. & Mattila, J. "The Spider in the Sphere. Eudoxus Arachne", *Philosophia* V/VI:225-257, Athens, 1976; Maula, E. "From time to place: the paradigm case", *Organon* 15:93-120, Warsaw, 1979.
[12] Maula, E. "An End of Invention", *Annals of Science* 38:109-122.
[13] Maula, E. "Proof, history and heuristics", *Historia Scientiarum* 29:125-155, Tokyo, 1985.
[14] Perisastri, M. "On Fermat's last theorem", *Amer. Math. Monthly* 76:671-5 (1969).
[15] Ribenboim, P. *13 Lectures on Fermat's Last Theorem*, New York, 1979 (Springer).
[16] Szabó, A. *Anfänge der griechischen Mathematik*, Budapest, 1969.
[17] Szabó, A. & Maula, E. *Les débuts de l'astronomie, de la géographie et de la trigonométrie chez les grecs*, Paris, 1986 (C.N.R.S.).
[18] Terjanian, G. "Sur l'équation p(x,2p)+p(y,2p) = p(z,2p)", *C.R. Acad. Sci.* Paris 285 (1977).
[19] Vandiver, H.S. "A property of cyclotomic integers and its relation to Fermat's last theorem", *Annals of Math.* 21:73-80 (1919).
[20] van der Waerden, B.L. "Die Postulate und Konstruktionen in der frühgriechischen Geometrie", *Arch. Hist. Exact Sci.* 18:343-357; cf. the review by E. Maula at *Zentralblatt für Mathematik* 432:9-11.

References to ancient authors and their commentators in full in [7] and [17].

LITERATURE OF MORE GENERAL INTEREST

Boyer, C.B. *The History of the Calculus and Its Conceptual Development*, New York, 1949 (Dover Publications).

Dantzig, Tobias *Number, The Language of Science*, New York, 1954 (4th ed.) (The Free Press).

Kline, Morris *Mathematical Thought from Ancient to Modern Times*, New York, 1972.

Tannery, Paul "Notions historiques" in J. Tannery *Notions de Mathématiques*, Paris, 1902 (Paul Tannery and Charles Henry edited Fermat's works 1891-1922).

Wallner, C.R. "Entwickelungsgeschichtliche Momente bei Entstehung der Infinitesimalrechnung", *Bibliotheca Mathematica* (3), V, (1904), pp. 113-124.