AFTERWORD (FOR THE LAYMAN)

0. — This concluding part is written for all the friends, relatives, and the occasional strangers met at parties who have been wondering what, exactly, I have been doing all these years. The most frequent questions were: is your thesis going to be just numbers and formulas, or also words? Did you prove some new theorem or formula? What is your thesis about? Is it useful for something?

The first question needs no answer at this point. The second question has a quick and dirty answer – I proved this formula (that's Theorem B in the Introduction):

$$L'_{p,\mathscr{W}}(f_E,1)=D_F^{-2}\prod_{\wp|p}\left(1-\frac{1}{\alpha_\wp}\right)^2\left(1-\frac{1}{\varepsilon(\wp)\alpha_\wp}\right)^2\langle z_f,z_f\rangle_{\mathscr{W}}.$$

This, of course, means nothing without context, which brings me to the (longer) answer to the third question.[42] Before going there, however, let me offer two words about whether this is *new*: yes, and no. No, that is, because this work is a generalisation of the results of another mathematician. Yes, in the sense that in this situation those results were not known before (were they, no doctorate would follow), and although perhaps it is not surprising (to the experts) that they are true, proving them so has required some "*new*" ideas – which were in turn inspired from other people's work... But this is not an essay in the philosophy of history of scientific ideas.

1. — Esoteric as these pages may look like, the questions they deal with are much older than all the modern-day musings about gravity, or cells, or molecules. In fact, the proper name for this subject, Arithmetic, usually elicits reactions better-known to contemporary artists ("Any child can do that" – arithmetic is identified with primary school mathematics), so that we are stuck with the (uglyish) name of "theory of numbers". The numbers in questions are the integers $0,1,2,3,\ldots$ and their negatives, as well as the fractions (such as $3/4,-7/5,\ldots$) – which go under the name of *rational* numbers. The founding father is usually recognised in Diophantus of Alexandria (3rd century A.D.), who in a treatise named, indeed, *Arithmetica* listed many problems and solutions to equations to be solved in integers or rationals. These questions already had a long history by then, starting with the Pythagorean discovery of the irrationality of the square root of 2 – that is to say, the non-existence of rational solutions to the quadratic equation $x^2=2$.

[42] The more practically-minded reader should jump to the end for an answer to the last question.

The general theory of quadratic equations of one variable is known to most high schoolers, and reduced to the Pythagorean question of the squareness of the discriminant[43] in the system of numbers of interest. Similarly, the study of equations of degree three or four[44] in one variable was reduced to the extraction of third and fourth roots by Italian Renaissance mathematicians. That the same cannot be said in general for equations of degree five or more was the early-Nineteenth-century discovery of Abel and independently Galois, who developed a complete theory of the symmetry of those equations. This was the end of a beautiful story – and the beginning of a new one, but we will not go there.

2. — Many of the equations studied by Diophantus are rather those in *two* variables, like $x^2+y^2=1$ (more generally one could consider systems of several equations in several variables, but two variables are still enough to give us headaches after almost two millennia). The rational solutions to this equation, like $x=3/5$, $y=4/5$, correspond to Pythagorean triples – triples of whole numbers, such as $(3,4,5)$, which can be the sides of a right-angled triangle[45]. Of course $x^2+y^2=1$ is also the equation for a circle: more precisely, this means that the solutions of this equation in the system of the *real* numbers (that is, the infinite decimals like 0.7163538902...) form a circle in the xy-plane. That the same can be said of the real solutions of $x^2+y^2=3$ should convince the reader that solving equations in the system of rational numbers is considerably more complicated than in the system of real numbers, once he or she is told that $x^2+y^2=3$ has no rational solutions at all![46]

We have thus met the idea of attacking the difficult problem of *studying the solutions to an equation in the system of rational numbers* by *first studying them in simpler*

[43] For those struggling to recollect old memories, the discriminant of ax^2+bx+c is the (in)famous $\Delta=b^2-4ac$.

[44] Like $x^3+x+1=0$, or $2x^4-5x^2+7=0$.

[45] Because they satisfy the requirements of Pythagorean Theorem on the sum of the squares of the legs being the square of the hypothenuse: $3^2+4^2=5^2$. There are infinitely many Pythagorean triples, parametrized by $(m^2-n^2, 2mn, m^2+n^2)$ for any integers $m>n$.

[46] This fact is by no means obvious, although not difficult to show: writing $x=a/c$, $y=b/c$ with a common denominator c, the problem is equivalent to that of solving

$$a^2+b^2=3c^2$$

in integers a,b,c with c not equal to zero. If there is a solution, then the solution with the smallest possible positive c can't have a,b,c all even, since otherwise the halves of a, b and c would give a smaller solution. Now it is easy to see that the square of an even number $(2k)^2=4k^2$ leaves remainder 0 when divided by 4, while the square of an odd number $(2k+1)^2=4(k^2+k)+1$ leaves remainder 1. So looking at the remainders of division by four on either side of our equation, and denoting "$\equiv$" the relation of equality up to the addition of a multiple of 4, we should have $1+0\equiv 3$ or $0+1\equiv 3$ if only one of a or b is odd (so that c is too); or $1+1\equiv 0$ if both a and b are odd (so that c is even). This is clearly not the case, so there is no solution.

systems of numbers, such as the real numbers and the systems of numbers "up to the addition of multiples of a fixed integer", such as the system of numbers up to addition of multiples of 4 used in the last footnote (these are called simply numbers "modulo 4" – note that this is a *finite* number system, containing only the four elements $0, 1, 2, 3$: indeed we have $4 \equiv 0$, and for example $2+3 \equiv 4+1 \equiv 1$). This idea turns out to be very fruitful, and in fact it suffices to give a complete treatment of equations of degree 2 in two variables, such as $x^2+3xy-7y^2+5x-4=0$: there is a solution exactly when there are real solutions[47] *and* solutions in the numbers "modulo N" for any integer N; this can be verified, by a human or a computer, in a finite (and quite short) amount of time, and if there is a solution then there are infinitely many.

3. — Finally, we can approach the topic of this thesis. For general equations of degree three or more, the existence of solutions in the real and finite number systems is not enough to guarantee the existence of a solution in the system of rational numbers. Even if we know that there is a solution, in general we still don't know whether there are finitely many or infinitely many others. Actually, for equations of degree at least five, one of the most important recent results in the subject, a theorem of the German mathematician Faltings [12], says that there is always at most a finite number of solutions. (On the other hand, the study of equations of degree four can be reduced to the case of degree three.) So in a sense the most important equations to be studied are those of degree three, called cubic equations, like for example $x^3+y^3=1729$.[48] Here are the main questions: is there an efficient way (an algorithm) to detect whether a given cubic equation has a rational solution or not? If there is a solution, what can be said about how many solutions there are?

The answer to the first question is not known, and this thesis adds (almost) nothing to that. For the second question, it was known to Diophantus that given two solutions one could construct a third one by a geometric method: picturing the two given solutions as points P and Q on the cubic curve $\mathscr{C}$ in the xy-plane corresponding to the equation, the third one is constructed by intersecting the line through P and Q with $\mathscr{C}$ – since the degree is three, there will be three intersections, that is P, Q and the new solution. One can perform the same construction starting from just one point P and using the tangent line to that point (this corresponds to the 'degenerate' case $P=Q$); and then iterate the procedure. Then two things can happen: either one returns to P after a certain number of iterations (in this case P is called

[47] This may not be the case: consider for example $x^2+y^2=-1$. When there are real solutions, they form a conic in the xy-plane: an ellipse, a parabola or a hyperbola.

[48] I am choosing this equation because of its curious history: two famous mathematicians, Hardy and Ramanujan, were once meeting at Ramanujan's house. Hardy said upon arriving that his taxicab had a rather unremarkable number, 1729. Ramanujan immediately replied that the number was remarkable indeed, for being the smallest number expressible in two ways as the sum of two cubes: $1729 = 1000+729 = 10^3+9^3$, and $1729 = 1+1728 = 1^3+12^3$.

a *torsion* point); or one keeps getting new points (i.e. solutions) indefinitely. A fundamental 1922 theorem of Mordell says the following: all the rational solutions to the cubic equation of interest can be *generated from a finite number of points* by performing the geometric construction just described. The smallest possible such number of generating points (excluding the torsion points) is called the *rank* of the cubic: it is 0 or a positive integer, and it is 0 precisely when the number of solutions is finite; otherwise, the rank gives as a basic measure of "how infinite" the set of solutions is.

4. — What does the number of solutions in other number systems tell us about the rank of a cubic equation? After extensive computer simulations (a new thing at the time), the mathematicians Birch and Swinnerton-Dyer found in the 1960s a conjectural answer. Consider for example the equation $x^3+y^3=1729$; we can look at its solutions in the systems of numbers modulo p for various prime numbers p (we don't need to restrict to prime numbers, but they are sufficient, and easier to deal with since one can define the operation of division on the associated number systems); for example, in the system of numbers modulo 5 we have $2^3+1^3\equiv 9\equiv 5+4\equiv 4$, and $1729\equiv 1725+4\equiv 4$ so that $x=2$, y= 1 is a solution in this system. How many solutions can we expect, in general, in the system of numbers modulo p? We are looking at one equation in two variables, so roughly speaking we expect to have one free variable (for example, if every number modulo p had a unique cube root, then we could take x as the free variable and then $y=\sqrt[3]{x^3-1729})$ – since the free variables can assume the p values $0,1,\dots,p-1$ we then expect about p solutions.

The actual number of solution will vary around p, and here is the idea: if this number is often larger than p, then this should be because of the existence of *rational solutions* which "reduce" to solutions in the system of numbers modulo p.

How does one make this idea precise? If N_p is the number of solutions modulo p to a fixed cubic, a good measure of the discrepancy with the expected number of points is its ratio to p, that is N_p/p: if there are no rational solution (or only finitely many of them) this should always be around 1, while if there are infinitely many then it should often be larger. This information can be packaged into a function of a variable x (a real number),

$$L(x)=\frac{N_2}{2}\cdot\frac{N_3}{3}\cdot\frac{N_5}{5}\cdots\frac{N_{p_x}}{p_x}$$

where p_x is the largest prime number smaller than x. Again, if there are only finitely many rational solutions, each factor should be about 1 so that we expect that $L(x)$ does not grow when x grows. On the other hand if there are infinitely many solutions, Birch and Swinnerton-Dyer observed in their examples that $L(x)$

grew like

$$L(x) \sim C \log(x)^r,$$

for some real constant C and some non-negative integer r; and that r was equal to the *rank* of the curve.[49] They conjectured that the above relation between the growth rate of $L(x)$ and the rank should hold for all cubic equations, and moreover they predicted what the value of C should be (in terms of various quantities associated to this equation).

Fifty years later, this conjecture is still unproven and likely to remain so for a long time – notwithstanding the prize of one million dollars offered by a foundation who ranked it among the seven most important unsolved[50] mathematical problems. This thesis, as part of a wider circle of ideas developed in the past three decades, makes some progress towards it.

5. — The readers who have followed me to this point now know what we are talking about – yet they may (and should) complain that after several pages they still have received no explanation as to how any of the words in the title of this work relates to the problems I described. In fact, at least two words make sense: I have tried to explain above how the problem of finding rational solutions to an equation can be viewed as that of finding points with rational coordinates on a curve. I will try to explain the other words.

Shimura curves. — Shimura is a Japanese mathematician who studied a certain class of curves, called indeed Shimura curves, which *parametrise* certain other geometric objects (let us call those "Shimura objects"). An example should illustrate what this means: a circle in the plane can be identified by three numbers, the two coordinates of its centre (two real numbers) and the length of its radius (a positive real number); if we consider circles to be equivalent when they can be rigidly moved to coincide, than the radius is the only parameter: that is, the positive real numbers *parametrise* the circles up to equivalence.

Shimura curves are less concrete than other curves given by explicit equations, yet they have a crucial advantage: one can find rational points on them in a natural way, since they correspond to the "rational Shimura objects" – in the above example, circles with rational radius can be thought of as "rational circles" (of course, in the example it is trivial to find rational points on the – rather straight – "curve" constituted by the positive real numbers, but by now my reader will be convinced

[49] To get a sense at why this conjecture could not have been made before the advent of computers, notice that $\log(x)$, the inverse function of the exponential function, grows very slowlly. So if in the study of the growth of $L(x)$ one wants to see when it surpasses, say, the value of 9, and if $r = 1$, then one needs to find the number of solutions to his equation modulo primes up to $10^9 = 1,000,000,000$.

[50] One of the seven has been solved since being included in the list – but the million dollars offered for it was refused by the winner.

that for other curves this is not so). Shimura and others made the striking conjecture that *every cubic curve is related to some Shimua curve*, in the precise sense that one can find a transformation (a function) from a Shimura curve to the cubic curve, which sends rational points to rational points. (Here is an example to illustrate the concept: the parabola $y = x^2$ is related to the real line with coordinate t by the function which sends t to the point $(x = t, y = t^2)$ on the parabola – and if t is rational then so are x and y.) The proof of this conjecture by Andrew Wiles in 1993 was the essential ingredient towards his proof of the famous Fermat's Last Theorem, which had been waiting for one for four centuries.

Heegner points. — Heegner was a German high school teacher and amateur mathematician, who in 1952 had the brilliant idea of producing rational points on cubic curves from the natural rational points on the related Shimura curve. This is to date the only systematic way of finding solution to cubic equations.[(51)] One can raise the question of whether the so-called Heegner points thus obtained are torsion points or not, and here is the answer:[(52)] suppose that $L(x)$ grows like $C\log(x)^r$; then the point is torsion exactly when $r = 1$ (as an exercise, the reader can try to extract what this implies for the conjecture of Birch and Swinnerton-Dyer). Notice that, since r is supposed to equal the rank, it is no surprise that the Heegner point is torsion when $r = 0$, as in this case *every* solution is torsion. On the other hand, when r is at least two, we should expect many (non-torsion) rational points, yet the only known method for finding them fails, and we are at a loss.

p-adic heights. — A measure of the complexity of a rational number written as a reduced fraction m/n is the maximum of the number of digits of m and n. One can refine this notion to study the complexity of a point on a cubic curve – and the resulting measure of complexity of the point is called the *height* of the point. It is a non-negative real number, which is zero precisely when the point is torsion. A study of the height of Heegner points is what has led to the result mentioned just above on the "growth versus rank" part of the Birch and Swinnerton-Dyer conjecture.

[(51)] Here is an example of its power – relegated to a footnote in order to avoid scaring any readers with the size of the numbers involved. The equation $1063y^2 = x^3 - x$ has rank one, and the simplest non-torsion solution has the x-coordinate $q^2/1063$, where

$$q = \frac{11091863741829769675047021635712281767382339667434645}{317342657544772180735207977320900012522807936777887}.$$

(This was found as a Heegner point by Noam Elkies. The reader may make a guess on how long it would take to find it with a naive search.)

[(52)] Due to Gross and Zagier [13]: my more attentive followers may recognize in the latter the name of my *laurea* thesis advisor.

The *p-adic height* is another measure of the complexity of a point, which is not a real number but rather a *p-adic number*, that is an element of a certain infinite system of numbers obtained by combining together the systems of numbers modulo $p, p^2, p^3, \ldots$ for a fixed prime number p. Now reading the title should give some ideas as to what this thesis studies – which is useful for understanding the constant C in the growth of $L(x)$. (This constant is important: it is related to the problem of deciding on the very existence of any points on some *other* related cubic curves.)

6. — Finally, a word of explanation on the formula I started this section with (and which is indeed my main result): the Heegner point being studied is z_f (here f is the name of the corresponding cubic curve on which it lies), and $\langle z_f, z_f \rangle_{\mathscr{W}}$ is its p-adic height; on the left-hand side, the quantity $L'_{p,\mathscr{W}}(f_E, \mathbf{1})$ is related to the function $L(x)$ introduced above.

As mentioned before, the results of this work are not new in the study of rational solutions to cubic equations.[53] The novelty here is that they are proved for a certain class of systems of equations in any number of variables which share some of the features of cubic equations of one variable (some of these systems are related to the study of one-variable equations of degree higher than three), and over systems of numbers which are more general than the system of rational numbers.

7. — I have thus far escaped the question of what all this is useful for. Jacobi's noble answer *"pour l'honneur de l'esprit humain"* may not satisfy every palate in an era of tight budgets. It is certainly unclear that the world is a better place after these pages. Yet the same could be said of the vast majority of basic research.[54] *Si parva licet*, nothing could have sounded more abstruse than antimatter when Dirac suggested its existence in 1928; yet today it is used daily in PET[55] scans in every modern hospital. More to the point, no one from Diophantus to Heegner could foresee the advent of the internet; yet the cryptographic methods that protect our electronic transactions and communications are using all the arithmetic knowledge developed in the past two millennia – and cubic curves have a significant part in it. Whether the ideas of this work will prove useful for improving on those methods, for nothing at all, or for something else which has not been invented yet, is as unclear as it is unpredictable: we will have to wait and see.

Daniel Disegni • *E-mail :* disegni@math.columbia.edu

[53] They are due to the French mathematician Bernadette Perrin-Riou [30].

[54] Not to mention many other more marketable cultural artifacts: opinion pieces, TV ads, mortgage-backed securities…

[55] Positron Emission Tomography (positrons are the antiparticles of electrons).