THE WORK OF PIERRE DELIGNE

W.T. GOWERS

1. Introduction

Pierre Deligne is indisputably one of the world's greatest mathematicians. He has received many major awards, including the Fields Medal in 1978, the Crafoord Prize in 1988, the Balzan Prize in 2004, and the Wolf Prize in 2008. While one never knows who will win the Abel Prize in any given year, it was virtually inevitable that Deligne would win it in due course, so today's announcement is about as small a surprise as such announcements can be.

This is the third time I have been asked to present to a general audience the work of the winner of the Abel Prize, and my assignment this year is by some way the most difficult of the three. Two years ago, I talked about John Milnor, whose work in geometry could be illustrated by several pictures. Last year, the winner was Endre Szemerédi, who has solved several problems with statements that are relatively simple to explain (even if the proofs were very hard). But Deligne's work, though it certainly has geometrical aspects, is not geometrical in a sense that lends itself to pictorial explanations, and the statements of his results are far from elementary. So I am forced to be somewhat impressionistic in my description of his work and to spend most of my time discussing the background to it rather than the work itself.

2. The Ramanujan conjecture

One of the results that Deligne is famous for is solving a conjecture of Ramanujan. This conjecture is not the obvious logical starting point for an account of Deligne's work, but its solution is the most concrete of his major results and therefore the easiest to explain. What is less easy to explain is why the conjecture is interesting – I will come to that later.

2.1. **Statement of the conjecture.** Consider the expression

$$q\prod_{n=1}^{\infty}(1-q^n)^{24}.$$

The notation $\prod_{n=1}^{\infty}$ stands for "the product as n runs through all the positive integers", so the expression is shorthand for

$$q(1-q)^{24}(1-q^2)^{24}(1-q^3)^{24}(1-q^4)^{24}(1-q^5)^{24}\dots$$

Now one thing one can do with a product of expressions in brackets is multiply them out so that there aren't any brackets any more. It is not quite as obvious that this can be done when there are infinitely many brackets involved, but under suitable circumstances it can. To see this, note first that the above expression is itself shorthand for

$$q(1-q)(1-q)\dots(1-q)(1-q^2)(1-q^2)\dots(1-q^2)(1-q^3)(1-q^3)\dots(1-q^3)\times$$
$$\times(1-q^4)(1-q^4)\dots(1-q^4)(1-q^5)(1-q^5)\dots(1-q^5)(1-q^6)(1-q^6)\dots(1-q^6)\dots$$

where there are twenty-four $(1-q^n)$s for each n. Each term in the product of all these brackets is obtained by choosing either a 1 or a $-q^n$ from each bracket. We cannot pick the $-q^n$ term more than finitely many times or we would have an infinite power of q, which is not allowed. To be sure that the product makes sense, we need to know that no finite power of q can be obtained in more than finitely many ways, since that would give us an infinite coefficient. But if we want to obtain a power such as q^{53}, we can't pick the $-q^n$ term for any n greater than 53, or we would end up with a larger power of q. So there are indeed only finitely many ways of ending up with q^{53}, so the coefficient of q^{53} must be finite, and the same goes for any other power.

We can say slightly more about this. Imagine that we have the sequence

$$1,1,\dots,1,2,2,\dots,2,3,3,\dots,3,4,4,\dots 4,5,5,\dots 5,\dots$$

laid out in front of us, with 24 of each number and our task is to select some terms of the sequence that add up to 53. If we ignore the q at the beginning of the product, then the coefficient of q^{53} is the number of ways of selecting an even number of terms of the sequence that add up to 53 minus the number of ways of selecting an odd number of terms of the sequence that add up to 53. Because of the q at the beginning, however, this is in fact the coefficient of q^{54} in the product.

The coefficient of q^n is traditionally called $\tau(n)$, and τ is called *Ramanujan's tau function*. Thus, we have (by definition) the identity

$$q\prod_{n=1}^{\infty}(1-q^n)^{24}=\sum_{n=1}^{\infty}\tau(n)q^n.$$

Ramanujan's conjecture states that $\tau(p)$ lies between $-2p^{11/2}$ and $2p^{11/2}$ for every prime number p. More generally, it states that $\tau(n)$ lies between $-d(n)n^{11/2}$ and $d(n)n^{11/2}$, where $d(n)$ is the number of divisors of n (which is 2 when n is prime, since then the only divisors are 1 and n).

2.2. **Why is Ramanujan's tau function interesting?** If you are not familiar with the conjecture just described, then it will probably seem rather bizarre and arbitrary. For it to be a good conjecture, one needs to have some reason to be interested in the product $q\prod_{n=1}^{\infty}(1-q^n)^{24}$ in the first place, as well as some reason to expect the conjecture to be true. In short, one wants an *interpretation* of the conjecture, rather than a piece of uninterpreted algebra.

I cannot give a fully adequate interpretation in this talk, but I can at least try to convince you that such an interpretation exists. To start with, I need to mention that complex numbers are involved in an important way, which is not apparent from the way I have formulated the conjecture. Their involvement comes because q is actually standing for $e^{2\pi iz}$, where z is a complex number.

At this point, I shall temporarily have to assume a little familiarity with complex numbers. If we split z into its real and imaginary parts, writing it as $z=a+ib$, then

$$e^{2\pi iz}=e^{2\pi i(a+ib)}=e^{2\pi ia-2\pi b}=e^{2\pi ia}e^{-2\pi b}.$$

The modulus of this complex number is $e^{-2\pi b}$, since $|e^{2\pi ia}|=1$ for every real number a. The real number $e^{-2\pi b}$ is less than 1 if and only if $b>0$. This tells us that $|q|<1$ whenever the imaginary part of z is positive. We need $|q|$ to be less than 1 for the infinite product $q\prod_{n=1}^{\infty}(1-q^n)^{24}$ to converge, so as a function of z it is defined on the upper half plane of the complex numbers.

This function is called $\Delta(z)$, and it turns up in several places in mathematics. Usually when that is the case, it is not because the *formula* for the function is particularly interesting, but because the function has various distinctive *properties*. That is certainly the case for $\Delta(z)$. A basic property is that it is *holomorphic*, which means that it can be differentiated in the sense appropriate for complex functions. There is a sense in which holomorphic functions are very "rigid": for example, the values of a holomorphic function near one point determine the values of the function everywhere else. (This is a little bit like an infinitary analogue of the statement that the values of a polynomial of degree d at $d+1$ points determine the values of the polynomial everywhere else.) This rigidity makes it hard

to combine differentiability with certain other kinds of properties, so when a holomorphic function *does* have those other properties, it tends to be a rare and interesting function.

The function $\Delta(z)$ is easily seen to be periodic, in the sense that $\Delta(z) = \Delta(z+1)$ for every z. This is simply because $q = e^{2\pi iz}$ doesn't change when you add 1 to z and has nothing to do with the strange formula involving q itself. However, it also has a less obvious property: it turns out that $\Delta(-1/z) = z^{12}\Delta(z)$ for every z. A final restriction is that as the imaginary part of z goes off to infinity, $\Delta(z)$ goes to zero.

These properties of Δ make it a very special function. In fact, it is an example of a *modular form*, a class of functions that played an essential role in Andrew Wiles's proof of Fermat's Last Theorem and that is central to much of modern number theory.

2.3. **Sums of 24 squares.** When a function is interesting in this sense of having important properties that are hard to combine, it usually turns up in many different contexts. The function $\Delta(z)$ is no exception. The reason Ramanujan was interested in it is that it was closely connected, via the τ function, to the number of ways of writing a number as a sum of 24 squares.

When I say "closely connected", I don't mean that it actually *was* the number of ways of writing a number as a sum of 24 squares, or even close to that number. To explain the connection, I need to say a little about what formulae for complicated quantities such as this tend to look like. Very often, there is no hope of a useful exact formula for a given mathematical quantity, but what one can do instead is obtain a useful *approximate* formula. Given such a formula, one would then like to know how accurate it is: this accuracy is measured by an "error term". The formula itself we then call the "main term".

It turns out that the number of ways of writing n as a sum of 24 squares has an approximate formula in terms of some well-known functions from number theory: for each k, the function $\sigma_k(n)$ is defined as the sum of the kth powers of the factors of n, and a few of these functions are used in the formula.

Often the error term is so to speak a piece of junk that one just hopes is as small as possible. But here it is far from junk: it is Ramanujan's tau function, with all its connections to other parts of mathematics. Nevertheless, it is also an error term, so we are interested in its size.

2.4. **Why might one expect Ramanujan's conjecture to be true?** I have had to prepare this talk in complete isolation, so am not certain that the explanation I am about to give is correct in every detail. However, it gives the right answer, so I am prepared to take the risk.

Let us think first about the number of ways we would expect to be able to write a randomly chosen positive integer n as $x_1^2 + x_2^2 + \cdots + x_{24}^2$, with $x_1, \ldots, x_{24}$ all integers. This is the same as asking how many points $(x_1, \ldots, x_{24})$ there are in 24-dimensional space with integer coordinates and distance $\sqrt{n}$ from the origin. (The distance estimate is the natural generalization to 24 dimensions of Pythagoras's theorem.) The points with integer coordinates form a grid-like structure called a *lattice*, so another way of phrasing the question is this: how many lattice points are there on a sphere of radius $\sqrt{n}$?

Now let us ask a slightly different question: how many lattice points are there on *or inside* a sphere of radius $\sqrt{n}$? We can make a reasonable guess at this by noting that the volume of a ball of radius r is proportional to r^{24}, and thus to n^{12} if $r = \sqrt{n}$. Writing C for the constant of proportionality, we might therefore guess that when n is large, the number of lattice points in the ball of radius $\sqrt{n}$ is roughly Cn^{12}. (The reason this is a sensible guess is that we can fill up 24-dimensional space with 24-dimensional unit cubes centred on the lattice points, so in a certain sense the density of the lattice points is 1.) The number of lattice points on the sphere of radius $\sqrt{n}$ is the number of lattice points in the ball of radius $\sqrt{n}$ minus the number of lattice points in the ball of radius $\sqrt{n-1}$, since the distance from every lattice point to the origin is the square root of an integer. So we might expect a reasonable estimate to be $Cn^{12} - C(n-1)^{12}$, which works out at roughly $12Cn^{11}$ if one does the calculations.

Now the above argument is very crude, and ignores the fact that divisibility properties of n make a significant difference to the estimate. Nevertheless, it gives the right order of magnitude, so it gives us some clue about how the more refined estimates using the functions $\sigma_k(n)$ will behave.

But what about the error term? Well, at this point let us make a rather bold assumption: that the error term behaves rather as it might if the lattice points near a given sphere were choosing randomly whether to belong to it or not. If you toss a coin m times, then you'll expect to get heads roughly half the time and tails roughly half the time. But you would be amazed if as you kept tossing the coin, the number of heads never differed from the number of tails by more than 1: a certain error is very much expected.

We can say more than this. The *standard deviation*, which is a measure of how far you can expect to be from the average, has order of magnitude $\sqrt{m}$. That is, you shouldn't be surprised if the number of heads differs from the number of tails by around $\sqrt{m}$ but you should be very surprised if they differ by a lot more than that.

Going back to the lattice points deciding whether to jump into spheres, if the average number is proportional to n^{11} and the decisions behave as if they were random, then the standard deviation should be proportional to $\sqrt{n^{11}} = n^{11/2}$. So this is the kind of error one might expect, and the strange and unnatural-seeming number 11/2 actually arises very naturally.

So now we have an interpretation of Ramanujan's conjecture: it says that the error term in a formula for the number of ways of writing n as a sum of 24 squares should be of about the size one would expect based on a probabilistic model of what is going on.

You may still wonder what is so interesting about the number 24. Why didn't we go for sums of 32 squares, or 50, or 96? One answer is that the error term we get, the tau function, comes from the independently interesting function $\Delta(z)$ – this kind of connection is not the norm at all. Another answer is that 24-dimensional space is particularly interesting because of the existence of an extraordinarily symmetric structure called the *Leech lattice*, which is also a grid of points but with far more symmetries than the grid of points with integer coordinates. (A much milder version of this phenomenon occurs in two dimensions, where a lattice based on equilateral triangles has six-fold rotational symmetry, whereas a lattice based on squares has only four-fold rotational symmetry.)

2.5. **Another famous error term.** If you know a little mathematics, then the discussion in the previous section may have reminded you of the way that the prime number theorem is often discussed. As with the number of lattice points on a sphere of radius $\sqrt{n}$ in 24 dimensions, it does not seem to be possible to give a useful exact formula for the number of primes less than n, but it is certainly possible to give a useful approximate formula. Furthermore, the approximate formula can be given a probabilistic interpretation: the number of primes up to n is roughly what one would expect if each number $n \geq 2$ had a one in $\log n$ chance of being prime. In other words, it is roughly

$$\frac{1}{\log 2} + \frac{1}{\log 3} + \frac{1}{\log 4} + \cdots + \frac{1}{\log n} .$$

This sum works out as roughly $n/\log n$, with most of the terms in the sum of roughly the same size, so we would expect the error term to be in the region of the square root of this. However, whether that is really the case (or to be more accurate, whether for every $c > 0$ the error will be smaller than $n^{1/2+c}$ if n is large enough) is the most famous unsolved problem in mathematics, since it is one way of formulating the Riemann hypothesis.

3. A link between algebraic geometry and number theory

The comparison with the Riemann hypothesis may look a little fanciful, but in fact it is not. To explain why not, I need to say a little about a branch of mathematics called arithmetic algebraic geometry and then tell another mathematical story, which starts with the work of the great French mathematician André Weil (or rather, that is where I shall start it – Weil himself built on the work of earlier mathematicians).

Algebraic geometry is the study of curves and higher-dimensional sets that are defined by polynomial equations. A simple example of such a curve is the unit circle in the plane, which is defined to be the set of all points (x, y) such that $x^2 + y^2 = 1$. Some of the questions one can ask about sets defined by polynomial equations are what their topological properties are, what kinds of singularities they have (if any), how they intersect each other, and so on.

A polynomial equation makes sense in a number of different contexts, according to what kind of numbers the variables are allowed to stand for. For example, the equation $x^2 + y^2 = 1$ describes the unit circle in the plane if x and y are real numbers, but what if they are complex numbers? Then for every x we can solve the equation $y^2 = 1 - x^2$ – indeed, there are two solutions unless $x^2 = -1$ – so the set has become unbounded. It is called a "curve" because locally it can be described with one parameter, but when that parameter is a complex number the curve is geometrically two-dimensional because a complex number needs two real parameters to describe it.

3.1. **Finite fields.** A *field* is, roughly speaking, an algebraic system that resembles the three richest of the great number systems: the rational numbers, the real numbers and the complex numbers. In each of these systems, we can add, subtract, multiply and divide numbers (except that we mustn't divide by 0), and we have various useful rules such as that $x + y = y + x$ or $x(y + z) = xy + xz$.

A finite field is what it sounds like: a finite set of "numbers" where one can add, subtract, multiply and divide and the same kinds of rules hold that hold in the rationals, reals and complex numbers. The most basic examples are the fields of *integers mod p*. The idea here is that we pick a prime p and stipulate that the numbers go from 0 up to $p - 1$ and then back to 0 again. (It is helpful to imagine them going round in a circle, like the numbers on a clock.) To add or multiply two numbers, one pretends they are ordinary numbers but then replaces the result by its remainder on division by p. For example, if $p = 13$, then $7 + 8 = 2$ (because the remainder when you divide 15 by 13 is 2) and $5 \times 8 = 1$ (because the remainder when you divide 40 by 13 is 1).

Subtraction is also easy to define; the big surprise is that we can even define division. In a sense the definition is easy: a/b is defined to be the number c such that $bc = a$. But c is supposed to be one of the numbers $0, 1, \dots, p-1$. Why should we be able to find one with the property that $bc = a$? The answer to that question is not obvious – the explanation belongs in a typical first-year undergraduate mathematics course – but one always can, provided that b is not 0. For example, if $p = 13$ and we want c such that $4c = 7$, then we can take c to be 5, since $4 \times 5 = 20$ and 20 leaves a remainder of 7 when you divide by 13.

The field I have just described is called $\mathbb{F}_p$. Other finite fields can be built out of $\mathbb{F}_p$ by a process rather similar to the process by which the complex numbers are built out of the real numbers. In that process, we observe that the equation $x^2 = -1$ has no solution in the real numbers, and we then "adjoin" a number, traditionally called i, stipulating that it squares to -1 and that it obeys all the usual laws of arithmetic. Then the set of all numbers of the form $a + bi$ forms the complex number system.

Similarly, to build a larger field out of $\mathbb{F}_{13}$, say, we can take a simple equation with no solution and "adjoin" a solution. The equation $x^2 = -1$ will not do, because $5^2 = -1$ in $\mathbb{F}_{13}$ (since $25 = 2 \times 13 - 1$), so there is already a solution. However, the equation $x^2 = 2$ turns out not to have a solution in $\mathbb{F}_{13}$, so we can adjoin a solution, giving it a name such as w, and define a new field to be the set of all "numbers" of the form $a + bw$ where a and b belong to $\mathbb{F}_{13}$.

It is not quite obvious that the mathematical structure we have created is a field, but this can be shown without too much difficulty. One key step is to show that for every $a + bw$ we can find some $c + dw$ such that $(a + bw)(c + dw) = 1$. To do this, we start by noting that $(a + bw)(a - bw) = a^2 - b^2w^2 = a^2 - 2b^2$. Now $a^2 - 2b^2$ cannot equal 0 unless $a = b = 0$, or else, dividing through by b^2, we would find that $(a/b)^2 = 2$, but we know that 2 does not have a square root in $\mathbb{F}_{13}$. If we write u for $a^2 - 2b^2$, then we can find v such that $uv = 1$, and then $(a + bw)(a - bw)v = 1$. So for our $c + dw$ we can take $v(a - bw) = va - vbw$.

This gives us a field of size $13^2 = 169$. In general, if p is any prime number and m is any positive integer, we can create a field of size p^m by starting with $\mathbb{F}_p$ and adjoining the root of an *irreducible polynomial* of degree m – that is, a polynomial of degree m that cannot be written as a product of two polynomials of smaller degree. It turns out that any two fields of the same size are "isomorphic" – that is, they have exactly the same algebraic structure, even if they have been defined in different ways – so we talk about "the" field of size p^m and denote it by $\mathbb{F}_{p^m}$.

3.2. **Algebraic geometry over finite fields.** Algebraic geometry over the complex numbers is very convenient because of the *fundamental theorem of algebra*, which guarantees that every polynomial equation has a root. We saw in the example above how this made a difference: for every complex number x we can find y such that $y^2 = 1 - x^2$, but if x and y are required to be real numbers and $|x| > 1$, then we can no longer solve the equation.

It is slightly surprising that algebraic geometry should be interesting and important when the variables take values in finite fields, since here the fundamental theorem of algebra is very far from true. (Indeed, we relied on the existence of polynomials without roots to build larger fields out of smaller ones.) It is all the more surprising when one thinks about the topics that interest algebraic geometers, such as the topological properties of sets defined by polynomial equations. In a finite field, a set defined by polynomial equations is just a finite set, so it completely lacks any kind of continuity, which is a basic requirement for a set to have interesting topological properties.

Nevertheless, we can at least make sense of the idea of the basic object of study: a set defined by polynomial equations. For example, we can easily talk about a "circle" in $\mathbb{F}_{13}^2$: we define it to be the set of points of the form (x, y) such that x and y belong to $\mathbb{F}_{13}$ and $x^2 + y^2 = 1$. (An example of a point that belongs to this "circle" is $(2, 6)$ since $2^2 + 6^2 = 40$ and 40 leaves a remainder of 1 when divided by 13.)

We can also partially deal with the first criticism of finite algebraic geometry by using a trick that is used throughout mathematics, which is to consider *families* rather than *individuals*. Here, each prime number p gives us an obvious family of finite fields: the fields $\mathbb{F}_p$, $\mathbb{F}_{p^2}$, $\mathbb{F}_{p^3}$, etc. Although for each individual field there will be polynomials that don't have roots, those same polynomials will have roots in later fields in the sequence.

Once we have decided to look at families, there is a very basic question we can ask. Given a system of polynomial equations, how many solutions are there in $\mathbb{F}_{p^m}$, and how does this number increase as m goes to infinity? For instance, if we take the system consisting of the single equation $x^2 + y^2 = 1$ and we take $p = 13$, we could ask how big a "circle" is in $\mathbb{F}_{13}$, $\mathbb{F}_{169}$, $\mathbb{F}_{13^3}$, etc. This will give us an increasing sequence of numbers $a_1, a_2, a_3, \ldots$ which turns out to encapsulate a great deal of fascinating information.

One of the things one can do with sequences of numbers is form functions out of them. For example, if we take Ramanujan's tau function, or equivalently the sequence $\tau(1), \tau(2), \tau(3), \ldots$, we can form the function $\sum_{n=1}^{\infty} \tau(n) q^n$ with $q = e^{2\pi i z}$ and obtain the function $\Delta(z)$ discussed earlier. We can do something similar with the sequence just described. For reasons that are beyond the scope of this talk, the "right" function to take

is

$$Z(x) = \exp\Big(\sum_{n=1}^{\infty} \frac{a_n x^n}{n}\Big) \ ,$$

and it is called the *zeta function* associated with the system of polynomial equations and the prime p. As we shall see later, there are close analogies between these functions and the Riemann zeta function, the function at the heart of the Riemann hypothesis.

4. The Weil conjectures

One might think that asking mathematical questions is far easier than answering them, and most of the time one would be right. However, sometimes coming up with questions is a major achievement in itself. For example, it took extraordinary insight for Ramanujan to ask his conjecture about the τ function: the probable truth of that statement was a brilliant observation somewhat akin to the brilliant theorizing of a scientist. Sometimes, both the asking and the answering are major achievements: the asking of the question leads to an unexpected research programme that turns out to be extremely difficult and also extremely fruitful. One of the best examples of this is a set of conjectures due to Weil.

I mentioned earlier an obvious objection to the idea of algebraic geometry over finite fields: that the sets one is studying are finite and therefore of no topological interest. André Weil, building on the work of others before him, had the remarkable insight that this objection is *completely* wrong. He came to believe that a topological concept known as *cohomology*, which is a central tool for topologists, should be applicable to algebraic geometry over finite fields. Although he did not develop an appropriate cohomology theory himself, his guess that such a theory ought to exist led him to expect that several results in topology had analogies in the finite-field context, which in turn led him to formulate a series of conjectures about the zeta functions defined in the previous section. These conjectures have had a huge influence on the shape of mathematics ever since. Here is what they say.

Let $Z(x)$ be the zeta function arising from a system of polynomial equations of degree n and a prime number q. Then the following statements hold.

(1) $Z(x)$ can be written in the form $P(x)/Q(x)$ for two polynomials P and Q with integer coefficients.
(2) More precisely, there is a formula of the form

$$Z(x) = \frac{P_1(x)P_3(x)\dots P_{2n-1}(x)}{P_0(x)P_2(x)\dots P_{2n}(x)}$$

where each P_i is a polynomial with integer coefficients. The reciprocals of the roots of P_i are algebraic integers (that is, solutions of polynomials with integer coefficients and leading coefficient 1), and the roots themselves have modulus $q^{-i/2}$.

(3) The function $z \mapsto 1/q^n z$ interchanges the roots of P_i with the roots of P_{2n-i}.

(4) Under appropriate conditions (which I won't try to state here) the degree of P_i is equal to the ith *Betti number*, an important topological invariant, of the set determined by the system of polynomial equations when the coefficients are complex numbers.

4.1. **The relation between the Weil conjectures and topology.** I have repeatedly mentioned that finite sets have no interesting topological structure. However, there is a simple notion that links topology to the study of sets determined by polynomial equations over finite fields, which is that of a *fixed point*. A fixed point of a function f is simply a point x such that $f(x) = x$. Some of the best known results in topology concern fixed points: for example, Brouwer's fixed point theorem says that every continuous function from an n-dimensional ball to itself must have at least one fixed point. More generally, the Lefschetz fixed point theorem allows one to determine, for a very wide range of spaces X, how many fixed points a continuous function f from X to X must have, provided that f is not too wild and the fixed points are counted with an appropriate multiplicity that ensures that the number of fixed points does not change if f is modified in a continuous way. The number of fixed points is given by a formula that involves the homology groups of the space X.

What has this to do with the finite sets we are interested in? The answer starts with the observation that in the field with q^m elements, the function $x \mapsto x^{q^m}$, which raises every number to the power q^m, takes every x to itself. In other words, every x is a fixed point of this map, which is known as Φ_{q^m}. Having a complicated way of defining the identity function may not look very useful, but it is, because the function Φ_{q^m} makes sense in larger fields. Therefore, if we take any larger field F that contains $\mathbb{F}_{q^m}$, we find that we have a very useful characterization of the points that belong to $\mathbb{F}_{q^m}$: they are precisely the fixed points of Φ_{q^m}! (The reason there can't be any more fixed points when we enlarge the field is that the equation $x^{q^m} = x$ cannot have more than q^m solutions, since it is a polynomial equation of degree q^m.) The function Φ_{q^m} is called the *Frobenius endomorphism*.

Another useful fact about the Frobenius endomorphism is that $\Phi_{q^m}(x + y) = \Phi_{q^m}(x) + \Phi_{q^m}(y)$ for every x and y. From this, it is not hard to show that if S is a subset of F^d determined by some polynomial equations, and we take a point $(x_1, \dots, x_d)$ in S, then the

point $(\Phi_{q^m}x_1,\dots,\Phi_{q^m}x_d)$ also belongs to S. In other words, if we apply the Frobenius endomorphism to each coordinate, then we obtain a function from S to S. Moreover, the fixed points of this function are precisely the points that belong to the part of S that lives in $\mathbb{F}_{q^m}^d$. So if we want to count the latter (as we do), then we can instead count the fixed points of a certain function from S to S.

The reason this is potentially useful is that if the number of fixed points turns out to be rather stable under changes to the function, as the Lefschetz fixed point theorem tells us it would be in a more topological set-up, then we are no longer forced to think about the specific function we started with, but can instead extract the information from homology groups, which we may be able to compute in a completely different way. But to make all this work, we have to say what we mean by "homology groups" in this context.

This was done by Grothendieck, who was Deligne's supervisor, and Artin. Grothendieck suggested a theory called *étale cohomology*, and he and Artin worked out the details a couple of years later. This was a major advance, with repercussions throughout algebraic geometry, and it led to the solution of all of the Weil conjectures apart from the second one.

4.2. **Analogies between the Weil conjectures and the Riemann hypothesis.** Riemann revolutionized the study of prime numbers by showing that they were intimately connected with what we now call the *Riemann zeta function*, the function ζ that takes complex numbers to complex numbers and is defined by the formula

$$\zeta(s)=1^{-s}+2^{-s}+3^{-s}+\dots$$

This formula in fact only makes sense if the imaginary part of s is greater than 1, but Riemann realized that the function itself made sense for all complex numbers. This is connected with the "rigidity" I mentioned earlier: once you know the values of $\zeta(s)$ for all s with imaginary part greater than 1, there is precisely one complex-differentiable function defined on the rest of the complex plane (except that it tends to infinity at 1) that agrees with ζ for those s. Not only does $\zeta(s)$ make sense for all s, but it is an extremely important and interesting function.

Riemann proved that the zeta function has a kind of symmetry that relates $\zeta(s)$ to $\zeta(1-s)$. It is not quite true that the two are equal, but one can be expressed very neatly in terms of the other. This result is known as the *functional equation*.

Using the functional equation, one can show that $\zeta(s) = 0$ whenever $s = -2n$ for some positive integer n. These zeros of the zeta function are called *trivial zeros*. The *Riemann hypothesis* is the following statement.

Conjecture. *Let s be a non-trivial zero of the Riemann zeta function. Then the real part of s is $1/2$.*

Earlier I said that the Riemann hypothesis was a statement about the error term in the prime number theorem. It was another of Riemann's great results that the hypothesis above and the statement about the error term in the prime number theorem are equivalent.

The second part of the Weil conjectures, and in particular the statement that the roots of P_i have modulus $q^{-i/2}$, is known as the Riemann hypothesis for varieties over a finite field. Why is this a reasonable name?

One obvious connection is that the number $1/2$ plays an important role in both statements. One can make this observation look a little less superficial by means of a change of variables. The Riemann hypothesis says that the zeros of the ζ lie on the line $\Re(s) = 1/2$, while Weil's conjecture says that the roots of P_i lie on the circle of radius $q^{-i/2}$ about the origin. There is a simple function, based on the complex logarithm function, that takes the circle to the line, and using this one can change variables so that the zeros of P_i lie on the line $\Re(s) = 1/2$ as well. (However, we can't do this simultaneously for all i.)

Similarly, the third Weil conjecture, about the interchanging of roots, tells us that Z has a certain symmetry that is reminiscent of the functional equation, and after a suitable change of variables one can make this symmetry relate values at s with values at $1 - s$, just as with the functional equation for the Riemann zeta function.

More important than these cosmetic similarities, however, is the fact that the *uses* to which the Riemann hypothesis for varieties over a finite field can be put are similar to the uses to which the Riemann hypothesis for ζ can be put. In particular, both tell us about the sizes of error estimates. The original Riemann hypothesis tells us that the error term in the prime number theorem is about as small as we could reasonably hope if we believe in a probabilistic model for the primes. Similarly, the Weil conjecture tells us that many other error estimates are about as small as they could reasonably be. One example of such an application is Ramanujan's conjecture: as we have seen, the Ramanujan τ function is an error term resulting from the number of ways of writing a number as a sum of 24 squares, and the conjecture states that it is as small as could be hoped; the combined work of a number of authors showed (in not at all a simple way) that the conjecture was a consequence of the Weil conjectures.

5. Deligne's proof of the last remaining Weil conjecture

Deligne's most famous result is his proof of the Riemann hypothesis for varieties over a finite field, the conjecture that did not follow from the work of Grothendieck. I wish I could say something enlightening about this proof, but I cannot. However, I can attempt to convey how amazing the proof was, and why explaining it to a non-mathematical audience is out of the question, by repeating what others have said about it.

Grothendieck thought he knew how to go about proving the final conjecture. There were certain properties that a suitable cohomology theory ought to have, and from those, all the Weil conjectures would follow. However, proving those properties was hard: Grothendieck called them the *standard conjectures*.

One of the surprises about Deligne's result was that he did not prove the final Weil conjecture by proving the standard conjectures: instead, he found a different route. He did, however, make significant use of étale cohomology. (Indeed, I have heard it said that one of the reasons for Deligne's success is that he was the only person in the world apart from Grothendieck who truly understood Grothendieck's work.)

Here are a few of the ingredients that went into Deligne's proof.

- A theorem of Kazhdan and Margulis about monodromy groups of Lefschetz pencils.
- A method of Rankin for estimating Ramanujan's tau function.
- A cohomology theory of Grothendieck for certain L-functions.
- The classical invariant theory of the symplectic group.
- A Leray spectral sequence argument.
- The "tensor-power trick".

The only one of those that I know anything about is the last. Deligne manages to show that the zeros of P_i have modulus at least $q^{-(i+1)/2}$, when the aim was to show an estimate of $q^{-i/2}$. However, if one takes Cartesian powers of the set from which the zeta function is derived, then one ends up showing a bound of $q^{-(ri+1)/2}$ when the target is $q^{-ri/2}$. Because the ratio of $ri+1$ to ri tends to 1, this can be used to show that the correct bound for the original set was in fact $q^{-i/2}$, as desired.

The remaining ingredients are just words to me, and pretty terrifying ones at that. But I am somewhat reassured by an article by Nick Katz, who wrote about Deligne's work when he was awarded his Fields Medal. Two sentences stand out in his description of the proof. The first is this: "With characteristic daring, Deligne simply ignores the preliminary problem of establishing independence of l." I do not know exactly what is going on, but it is clear that Deligne faced an apparently insurmountable obstacle, and just blithely

continued where any normal mathematician would have given up. The second sentence gives an idea of what Katz, who knows far more about this area than I ever will, thought of the heart of Deligne's argument: "Deligne's proof of this is simply spectacular; no other word will do." And as he proceeds to describe what Deligne did (it is from that description that I extracted the list of ingredients above), he cannot stop himself adding exclamation marks to indicate just how extraordinary the steps of the argument are.

6. Conclusion

At the end of his article, Nick Katz says, "We have passed over in silence a considerable body of Deligne's work, which alone would be sufficient to mark him as a truly exceptional mathematician." That is even more true of what I have just written, partly because Katz mentioned several results that I have not mentioned, and partly because Deligne has proved many more results since 1978. However, I hope that the small and unsatisfactory glimpse I have been able to offer gives some idea of Deligne's greatness and of why he is such a natural choice for the Abel Prize.