The Letter

Robert Langlands
and the Dictionary of Worlds

William de France

Clementinium Editions

Clem

The Letter

Robert Langlands and the Dictionary of Worlds

William de France

Clementinium Editions

ESSAI Collection – REV260520 – 10pt-print

Clementinium Editions
`https://clementinium.com`

Typeset with LaTeX, in EB Garamond.
Format: 105 mm × 170 mm (ESSAI Collection, 10 pt print).

"Nothing is more fruitful (all mathematicians know it) than those obscure analogies, those disturbing reflections of one theory in another; those furtive caresses, those inexplicable disagreements; nothing also gives more pleasure to the researcher."

André Weil,
letter to his sister Simone, 26 March 1940.

Contents

Foreword

This book tells the story of a letter. Seventeen handwritten pages, sent in January 1967 by a young Canadian mathematician, Robert Langlands, to his senior colleague André Weil. The letter proposed, without proof, a precise correspondence between two very different branches of mathematics: the arithmetic of whole numbers on one side, and harmonic analysis on certain groups on the other. What might have remained a mere speculation became, in the decades that followed, a research program to which several generations of mathematicians have given their careers, and one of whose most ambitious versions was proved in 2024.

The wager of this book is that the story can be told without mathematical prerequisites. Each mathematical object is introduced when it is needed, with enough substance to be understood and no more.

At the end of the volume, appendices gather the usual reference tools: a note on the letter itself and its dating, a glossary, a table of symbols, a chronology, biographical notes on the main figures, an annotated bibliography, and an index of names and concepts.

Anyone wishing a more technical and more detailed exposition of the program will find it in the companion volume Langlands: The Secret Unity of Mathematics *(2024), from the same publisher, where the precise statements, the proofs, and the mathematical apparatus are set out in the discipline's own register.*

WdF

Prologue

Princeton, January 1967

A university office, a winter afternoon. The walls are pale wood, the window looks out onto a park white with snow. On the table, a ruled writing pad, a fountain pen, a cup of tea going cold. A man in his thirties writes slowly, stops, rereads what he has just set down, crosses it out, starts again. He has been working on this letter for several days. When he is finished, it will run to seventeen pages.

The office belongs to Robert Langlands, in the Fine Hall building at Princeton University. Langlands is Canadian, born in British Columbia, into a family of small building-trade merchants. He discovered mathematics almost by accident at a small provincial university and never looked away. He arrived at Princeton seven years earlier and is now an assistant professor. *At thirty, he is unknown outside a small circle of specialists; but he has a curiosity that runs well beyond his thesis topic. Several years of wide, unsystematic reading had left, in his mind, an intuition that gradually took shape.*

The letter is addressed to André Weil, who works a few hundred metres away, in another institute at Princeton. Weil is sixty, with an international career behind him and the unofficial status of patriarch of number theory. The previous summer, at the International Congress of Mathematicians in Moscow, he gave a lecture that people are still talking about. He is not known for patience with vague ideas.

Langlands is not quite sure of his ground. In the first sentence of the letter, he almost apologizes. He writes, in substance, that if his correspondent is willing to read it as pure speculation he would be grateful, and that otherwise he no doubt has a wastepaper basket

close at hand. That sentence has become one of the most celebrated in the history of modern mathematics.

The seventeen pages do not resemble what his contemporaries usually read. It is not an article, nor an isolated conjecture, nor a proof. It is a program. *An overall plan for an entire branch of mathematics yet to come. A prediction about the shape of theorems that do not yet exist and that will, he supposes, take decades to prove.*

The idea is this. Over the twentieth century, mathematics developed two large regions that grew up almost independently. The first is number theory. It studies the integers, the primes, equations with integer coefficients, and the symmetries that permute their solutions. It is a very old domain, with familiar concerns. The second is called harmonic analysis. It studies functions living on spaces with many continuous symmetries, and the way those functions decompose. It is a younger, more abstract region, developed in the twentieth century to study waves, oscillations, spaces of rotation.

At first glance, these two regions have nothing in common. One is discrete, algebraic, turned toward arithmetic. The other is continuous, analytic, closer to physics than to number theory. One can spend a career in either without ever meeting the other. Many mathematicians do.

Langlands's intuition is that they speak to each other. More than that: that they are, in a precise sense, two descriptions of one and the same reality. Each arithmetic object should correspond, term for term, to an analytic one. Each question posed in one language should be re-expressible in the other. A dictionary, in short, should exist between these two regions: not a dictionary in the loose sense where two things "resemble" each other, but a precise dictionary, where every word in one language has an exact counterpart in the other.

The idea is not entirely new. Early in the twentieth century, there was already a theory that realized a dictionary of this kind in a very simple case. It is called class field theory. *It took fifty years of work by mathematicians of the first rank (David Hilbert, Teiji Takagi, Emil Artin, Claude Chevalley) to reach its finished form, and it handled*

only the simplest possible case of the dictionary. What Langlands proposes, in his seventeen pages, is to generalize this theory to every case. To every dimension. To every group. Without proof. With only the intuition that such a structure, however vast, cannot fail to exist.

Weil will read the letter. He will not throw it away. He will reply briefly, politely, with measured skepticism. But he will pass it around. In the years that follow, Langlands's intuition will become a program, the Langlands program, *to which several generations of mathematicians will devote their careers. Seminars will meet in Berkeley, Oxford, Paris, Moscow, Bonn. Decades will pass. Pieces of the dictionary will be proved. Fermat's Last Theorem, open since 1637, will fall, almost incidentally, as a consequence. Fields Medals will go to those who advance the program.*

And one day in July 2024, fifty-seven years after the letter, a team of nine mathematicians led by Dennis Gaitsgory and Sam Raskin will announce the complete proof of one of the most ambitious versions of the program. The result will fill five main papers and more than eight hundred pages.

This book tells what was in the letter of 1967, and what it took to prove Langlands right. It assumes no particular mathematical background. All it asks is a little patience, and the desire to watch an idea unfold over time.

Part I

The Two Shores

1

The symmetries of numbers

The equation $x^2 = 2$ has two solutions, which we write $\sqrt{2}$ and $-\sqrt{2}$, roughly equal to 1.414 and -1.414. A slightly strange question suggests itself about them: which of the two, more than the other, deserves to be called $\sqrt{2}$? Which is the "real" one? By convention we give the name to the positive one. But is that convention justified by anything mathematical?

On reflection, no. Nothing whatsoever distinguishes $\sqrt{2}$ from $-\sqrt{2}$ algebraically. Every relation satisfied by one is satisfied by the other. Every polynomial with integer coefficients that vanishes at $\sqrt{2}$ also vanishes at $-\sqrt{2}$. The two values are interchangeable: one can replace either by the other in any algebraic formula without anything changing.

This small observation, apparently banal, contains the seed of an idea that will upend algebra in the nineteenth century. What matters in an equation is not the solutions themselves. It is the *symmetries* that permute them.

The idea was carried to its full maturity by a young French mathematician who did not live to see it bear fruit. Évariste Galois died at the age of twenty, in May 1832, of a gunshot wound received in a duel whose precise cause is still not known. The last years of his life he spent largely in prison on political charges. On the nights before his duel, sensing perhaps what was coming, he hurriedly wrote out his mathematical ideas in a manuscript left to a friend, along with a letter that begins with the famous words "I have no time." It took several decades before mathematicians fully grasped what he had discovered.

The question that occupied Galois had occupied algebraists for

the previous two centuries. Given a polynomial equation, can one write its roots, starting from the coefficients, using only a finite sequence of elementary operations: addition, subtraction, multiplication, division, and the extraction of roots (square, cube, fourth, fifth, and so on)?

For equations of the first degree, of the form $ax + b = 0$, the answer is yes: the solution is $x = -b/a$. For equations of the second degree, of the form $ax^2 + bx + c = 0$, the answer is again yes: the formula has been known since the Babylonians, two thousand years before our era, and is in every school textbook. For the third degree, the formula was discovered by the Italian Renaissance algebraists, Cardano, Tartaglia and Ferrari above all, after the famous public contests whose echoes fill the histories of mathematics. For the fourth degree, a formula exists too, though more complicated. But for the fifth degree, no one could find one, despite two centuries of work by Europe's best mathematicians.

Galois's answer to the question was not a new formula. It was something radically different: a change of viewpoint. Instead of looking for the formula directly, Galois proposed to look at the *permutations* of the roots, that is, the ways one can interchange the solutions while preserving their algebraic relations.

To each equation, then, corresponds a set of permitted permutations. This set has a particular structure: one can combine two permutations by carrying them out one after the other, and the result is again a permitted permutation. One then speaks of a *group*, and the group attached to an equation is called the *Galois group* of that equation.

Take our equation $x^2 = 2$ again. The roots are $\sqrt{2}$ and $-\sqrt{2}$. The possible permutations are two in number: the identity, which leaves each root in place, and the exchange, which swaps them. The Galois group of this equation therefore has two elements. Its structure fits on a single line.

For a generic equation of the second degree, the Galois group has two elements. For an equation of the third degree, it can have

up to six elements. For one of the fourth, up to twenty-four. For one of the fifth, up to a hundred and twenty. As the degree grows, the maximum size of the group grows very fast. But size is not the decisive criterion. What matters is the *internal structure* of the group.

The precise statement Galois proved, which settled the question that had been open for two centuries, is this. An equation admits a formula for its roots in terms of radicals if, and only if, its Galois group has a certain structure that Galois called *solvable*. Solvable groups are the ones that can be taken apart, layer by layer, into simple pieces (the abelian pieces, that is, the commutative ones, in which the order of combination does not matter). Groups that are not solvable cannot be taken apart in this way, and for them no formula in radicals exists, however clever.

The distinction between solvable and unsolvable equations, therefore, does not depend on the ingenuity of mathematicians. It is inscribed in the equation itself, through the structure of its Galois group. For equations of degree one, two, three, four, the Galois group can always be decomposed into simple pieces. For degree five, this is no longer true in general: the Galois group contains, at bottom, a piece that can no longer be decomposed, and its presence is enough to prevent any formula.

That is for a single equation. But Galois's idea has a reach that extends well beyond this initial case. It says, in substance, that to understand an algebraic problem, one must look not at the solutions themselves, but at the symmetries that permute them. This intuition will drive, throughout the twentieth century, an almost complete reworking of algebra.

Galois had applied his idea to *one* equation at a time. His successors will apply it to *all equations at once*.

Suppose that, instead of a particular polynomial, one takes the whole collection of polynomials with rational coefficients. Suppose one solves every one of them. And suppose one gathers all the solutions so obtained into a single large collection. That

collection is enormous. It contains $\sqrt{2}$, of course, but also $\sqrt[3]{7}$, $\sqrt[5]{11}$, and, trivially, the rationals themselves, along with countless other numbers, provided they are roots of some polynomial with rational coefficients. Such numbers are called *algebraic numbers*. They stand in contrast to the numbers that cannot be obtained in this way, called *transcendental numbers* (the most famous are π and e).

The collection of all algebraic numbers forms what one calls the *algebraic closure* of the rational numbers, written $\overline{\mathbb{Q}}$. It contains the rationals (which are trivially algebraic, since p/q is a root of the polynomial $qx - p$), and all the numbers one can obtain from them by solving polynomial equations. Everything that polynomial algebra can reach, starting from the rationals, is there.

Galois's idea extends, however, to the whole collection $\overline{\mathbb{Q}}$ viewed as an extension of the rationals, rather than to any single polynomial. The symmetries of $\overline{\mathbb{Q}}$ that fix each rational number (that is, that leave $0, 1, 2, 1/3$ and every other rational in place, while permitting permutations among the algebraic non-rational numbers) form a group. This group is called the *absolute Galois group* of the rationals, and is written $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

It is a colossal object. Unlike the Galois group of a particular polynomial, it is not finite. It contains the Galois group of every polynomial equation, all nested within one another in a structure one can in principle describe, but which remains, in practice, impenetrable. It is, in short, the group that gathers all the arithmetic information available about the rational numbers.

The claim is this: all the arithmetic information one can state about the rational numbers is, in one form or another, inscribed in the structure of this group. The way a prime number decomposes in a given extension. The existence or non-existence of integer solutions to a given equation. The reasons certain congruences between seemingly unrelated numbers are inevitable. All these questions have their answers, in principle, somewhere in the structure of the absolute Galois group.

In principle. Because there is one obstacle: this group is too large to be looked at directly. It is infinite, it has no simple explicit description, and one knows its internal structure only through theorems about its finite fragments. To describe this group directly remains out of reach.

So one needs an instrument. An indirect way of extracting from this group fragments of information that, together, reconstitute what it contains. That instrument exists. It was developed from the 1890s on, not for the absolute Galois group, but for finite groups. It is called a *representation*.

2

Representations

One of the most fruitful ideas in modern algebra fits in a single phrase: to study a group, one turns it into matrices.

A matrix is an array of numbers arranged in rows and columns. A matrix of size 2×2, for example, is a square of four numbers. One can add two matrices, one can multiply them, and multiplication obeys rules close to those of ordinary numbers, with one difference: when one multiplies two matrices A and B, the order matters. The product AB is not, in general, equal to BA.

It is precisely this last property that makes matrices interesting for our subject. For the groups one wants to study, beginning with the Galois groups, are not commutative either. Combining two symmetries in one order or the other can give different results. Matrices, in their multiplication, reproduce that property. They therefore offer a natural setting in which to translate groups into objects one can calculate with.

The idea was formulated in the 1890s, at the University of Berlin, by a German mathematician named Ferdinand Georg Frobenius. He was answering a question raised by his colleague Richard Dedekind, and in thinking it over he realized that there was a systematic way of turning any finite group into a family of matrices. The rule is simple. To each element g of the group one assigns a square invertible matrix $\rho(g)$ of size $n \times n$. One asks only one thing of the assignment: if one combines two elements g and h in the group, the product gh must correspond to the product $\rho(g)\rho(h)$ of the associated matrices. In other words, the group's law of combination must be respected by the matrices.

Such an assignment is called a *representation* of the group.

The integer n is called the *dimension* of the representation. The larger n, the larger the matrices, and the more informative the representation can be.

Why does this change anything? Because once the group has been turned into matrices, one can calculate. Matrices have eigenvalues, a trace (the sum of the diagonal entries), a determinant, a characteristic polynomial. The whole machinery of linear algebra, developed over two centuries to study systems of linear equations and geometry, becomes available. A group that was only a combination table, often impossible to write out completely, now sits on a vector space where its action becomes visible. One can examine it, compare it, decompose it.

The most informative representations are those that cannot be decomposed further. What is meant by that is simple. Suppose a representation sends the elements of the group into $n \times n$ matrices. Suppose also that there is, inside the n-dimensional space, a smaller subspace that every one of those matrices leaves invariant (sends to itself). Then one can restrict to that subspace and obtain a smaller representation, of lower dimension. A representation that has no nontrivial invariant subspace is called *irreducible*. It is, in a sense, an atom of the theory: it cannot be broken into simpler pieces.

Frobenius, and after him Issai Schur and Richard Brauer, established for these irreducible representations a series of identities of a cleanness rare in algebra. The number of irreducible representations of a finite group is exactly equal to the number of its conjugacy classes (a partition of the group's elements determined by its internal structure alone). The sum of the squares of the dimensions of the irreducible representations equals the order of the group (that is, its total number of elements). The full information of the group fits into a small table called the *character table*, which encodes everything there is to know. For finite groups this theory achieves an uncommon economy: it reduces the study of a group to that of its character table, and all calculations proceed from that table.

But the group Langlands cared about, the absolute Galois group of the rationals, is not finite. It is infinite. It even has a particular structure: it is what is called *profinite*, which means, without going into detail, that it is the limit of an infinite sequence of finite groups nested within one another, like Russian dolls. To study such a group by means of representations, one has to adapt Frobenius's framework.

Two modifications suffice. First, one no longer takes matrices with coefficients in just any numbers. One takes them with coefficients in a particular type of number, called *p-adic fields*. This is a mathematical construction fascinating in its own right, and deserving of a book of its own, but for which it is enough here to retain the general idea. The p-adic fields are numbers in which the notion of closeness works differently from what ordinary arithmetic teaches. In ordinary numbers, two values are close if their difference is small. In the p-adic fields, two values are close if their difference is divisible by a large power of a prime p. This strange way of measuring closeness happens to be exactly the one that Galois representations need in order to capture the arithmetic structure of the rational numbers.

Second, one requires the representation to be *continuous*, which means, intuitively, that it has no sudden jumps. If two elements of the Galois group are very close (in the particular sense of the profinite topology described above), their images under the representation must be close as well (in the p-adic sense). This continuity condition forces the representation to respect the internal structure of the group, rather than simply gliding over its elements.

Such a representation is called a *p-adic Galois representation*. It is one of the two types of objects Langlands will put into correspondence in his letter. It is both concrete (its values are matrices, one can compute with them) and intimately tied to arithmetic (it encodes information about the rational numbers that would otherwise be inaccessible).

But how exactly? To see why these representations are so useful, one has to describe a particular type of element of the Galois group, called the *Frobenius*.

For each prime p, there exists within the absolute Galois group a particular element (or more precisely, a class of elements), called the Frobenius at p and written Frob_p. It is defined by its action on numbers considered modulo p. Working modulo p means not distinguishing two integers that differ by a multiple of p: modulo 7, the numbers 3, 10, 17, 24 are all considered the same, since they differ from one another by multiples of 7. One obtains in this way a system of numbers smaller and simpler than the ordinary integers. The Frobenius at p is the symmetry which, in that system, sends each element x to x^p. This operation preserves all the algebraic relations that hold modulo p: it is indeed a symmetry.

The Frobenius matters because it condenses, in a single element, all the arithmetic information tied to the prime p. And when one applies a Galois representation to that element, one obtains a matrix. That matrix, like any matrix, has a characteristic polynomial, a trace, a determinant, eigenvalues. These quantities are finite, concrete, and computable in principle. They are the *local data* attached to the prime p by the representation.

Here is the observation. If one gathers these local data for every prime p (for $p = 2, 3, 5, 7, 11$, and so on), one obtains a sequence of matrices, or more simply a sequence of characteristic polynomials, or more simply still a sequence of numbers (the traces of the Frobenius matrices). This sequence is the *imprint* of the representation: a collection of information, prime by prime, that characterizes it completely.

In practice, two distinct representations give distinct imprints. The local information at each prime, taken together, determines the representation. And it is this imprint, this collection of information prime by prime, that the L-function will condense into a single analytic object.

The Galois representations are only one of the two shores that

Langlands wants to connect. The other shore is that of *automorphic forms*. They arose in mathematics along a completely different road, discovered by mathematicians who were not thinking of number theory at all, and their beauty, on first acquaintance, has something almost unaccountable about it.

3

Automorphic forms

Here is a sequence of integers.

$$1,\ -24,\ 252,\ -1472,\ 4830,$$
$$-6048,\ -16744,\ 84480,\ \ldots$$

No pattern jumps to the eye. The signs do not alternate regularly, the values do not grow by any visible law, and one could go on listing them for several pages without discerning any form. It looks like a random sequence. Yet it is not.

It satisfies arithmetic identities of startling precision. For any two coprime integers m and n (integers with no common divisor other than 1), the term at position m multiplied by the term at position n gives exactly the term at position mn. And for every prime p, the terms at positions p, p^2, p^3, and so on, are bound together by a simple recursion. These regularities are not coincidences. They are the trace of a hidden structure the sequence obeys, without anything in its appearance hinting at it.

The sequence is called $\tau(n)$, the Ramanujan tau function. It was discovered in 1916 at Trinity College, Cambridge, by the Indian mathematician Srinivasa Ramanujan, who had extracted it, by hand, from an innocent-looking infinite product:

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

If one expands this product as a power series (an infinite polynomial in powers of q), the coefficients one obtains are $\tau(1) = 1$,

$\tau(2) = -24$, $\tau(3) = 252$, and so on. Ramanujan computed the first ones by hand. He observed the multiplicative properties over hundreds of cases. But he could not prove it. The proof came in 1917 from a British colleague, Louis Mordell.

The recursion satisfied by τ, for its part, carries an unexpected coefficient: for every prime p,

$$\tau(p^{k+1}) = \tau(p)\,\tau(p^k) - p^{11}\,\tau(p^{k-1}).$$

The exponent 11 is not there by accident. It is linked in a precise way to the construction of Δ, a link Ramanujan sensed without being able to prove. And it is this arithmetic precision, this tie between an apparently random sequence and a number as specific as 11, that will turn out to be the signature of an entire theory.

That theory is the theory of *modular forms*.

The terrain shifts here. Ramanujan's sequence looks like a purely arithmetic fact, but it comes from an object that lives in an entirely different world, the world of geometry. To understand where its regularities come from, one has to visit first the place where that object lives. The tie with arithmetic will come back into view at the end of the road.

The first place to introduce is the *upper half-plane*, written $\mathbb{H}$. It is simply the upper half of the complex plane: all complex numbers $z = x + iy$ with strictly positive imaginary part y. On a sheet of paper, it is the region above the horizontal axis. An infinite half-sheet, bounded below by a horizontal line one never touches.

This half-plane has a geometric property that makes it interesting. If one equips it with a certain way of measuring distances (its *natural metric*, whose details do not matter here), it becomes a space of constant negative curvature, called the *hyperbolic plane*. Negative curvature means, intuitively, that at every point the space spreads away from itself like a saddle, unlike a sphere, which has positive curvature. In such a space, two apparently parallel lines

always diverge, and what looks close to the boundary of the half-plane is in fact at infinite distance.

Every geometric space has a group of symmetries: the transformations that leave distances unchanged. For the Euclidean plane, these are the translations and rotations. For the hyperbolic plane, the group of symmetries goes by a technical name, $\mathrm{SL}(2,\mathbb{R})$. It consists of 2×2 matrices with real entries and determinant 1. Each such matrix sends a point z of the half-plane to another point, by the rule:

$$z \longmapsto \frac{az+b}{cz+d},$$

where a, b, c, d are the entries of the matrix. This rule is called a *Möbius transformation*. It sends $\mathbb{H}$ to itself and preserves the hyperbolic distances between points.

Inside $\mathrm{SL}(2,\mathbb{R})$ a particular subgroup stands out. It is $\mathrm{SL}(2,\mathbb{Z})$, the set of 2×2 matrices with integer entries and determinant 1. It is merely a discrete part of the total group, but it inherits the ability to act on the half-plane. This subgroup brings in the arithmetic ingredient that was missing: its entries are integers, so its actions on $\mathbb{H}$ are constrained in a very particular way.

When one lets $\mathrm{SL}(2,\mathbb{Z})$ act on $\mathbb{H}$ and identifies all the points it connects, one obtains a new geometric object, called the *modular curve*. By tradition, it is called a curve, even though to our eyes it is a two-dimensional surface. It is a surface of finite area that tapers off to a single point at infinity, called its *cusp*. Each point of this curve corresponds to an elliptic curve (up to equivalence), which is why the arithmetic of elliptic curves can be read off the geometry of the modular curve.

Now to the modular forms. One is looking for functions on the half-plane that are compatible with the action of $\mathrm{SL}(2,\mathbb{Z})$: functions that, when a point is moved by a matrix of the group, change value according to one fixed law. More precisely, a *modular form of weight k* is a holomorphic function (that is, a function

differentiable in the complex sense) on the half-plane $\mathbb{H}$ that satisfies, for every matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\mathrm{SL}(2,\mathbb{Z})$, the relation:

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k\, f(z).$$

The exponent k is the weight of the form. The constraint is strong: the group $\mathrm{SL}(2,\mathbb{Z})$ contains infinitely many matrices, and each imposes its own equation on f. A holomorphic function chosen at random will violate every one of them. And yet functions that satisfy them all exist. For each weight k they form a finite-dimensional vector space whose dimension can be computed exactly. That, in itself, is striking.

Two families of classical modular forms deserve to be named. The *Eisenstein series* are modular forms constructed explicitly as sums over pairs of integers. There is one for each even weight $k \geq 4$. And Ramanujan's Δ is a modular form of weight 12 with an extra property: it tends to zero as the imaginary part of z grows large. Modular forms that vanish in this way at the boundary are called *cusp forms*, and they form a particularly rich subfamily.

Back to the sequence τ. The exponent 11 in the recursion it satisfies is nothing other than $k-1$ for $k=12$. The number that Ramanujan had ended up guessing through hand computations was already written into the weight of the form. The arithmetic of Ramanujan's sequence is entirely determined, up to that exponent, by the modular form it comes from.

Why do modular forms produce sequences with such precise arithmetic properties? Because the symmetries they satisfy are rigid enough to fix their internal structure almost completely. A cusp form of weight 12 is unique up to a multiplicative factor. Its coefficients are therefore determined, up to that factor, by the modularity condition and the boundary condition alone. The arithmetic identities between coefficients follow directly from the symmetry.

In the 1930s, the German mathematician Erich Hecke introduced a family of operators, now called *Hecke operators*, which act on the space of modular forms of a given weight. A form that is an eigenvector for all these operators at once is called a *Hecke eigenform*. Its coefficients automatically satisfy the multiplicative and recursive identities that Ramanujan had observed by hand. Ramanujan's Δ is a Hecke eigenform, and so are the cusp forms at low weights, where the space has dimension at most 1.

Classical modular forms are the simplest instance of a much broader theory. When one replaces $\mathrm{SL}(2,\mathbb{Z})$ with other groups of symmetries built from integers, and the half-plane with spaces of higher dimension, one obtains other families of functions satisfying symmetry relations of the same kind. They are called *automorphic forms*. The classical modular forms are their prototype.

That is the second shore. On one side, the Galois world, of algebraic symmetries and whole numbers. On the other, the automorphic world, of functions on curved spaces. In 1967, both theories were mature enough to be compared with precision. What neither, on its own, possessed, was the bridge that was to connect them. The only instrument visible on both shores was the L-function.

4

L-functions

There is a mathematical equality that, at first sight, should not have existed. It was discovered in 1737 by Leonhard Euler, then working at Saint Petersburg, and it goes as follows.

On one side, an infinite sum over all positive integers:

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots$$

On the other, an infinite product over the prime numbers only:

$$\prod_{p \text{ prime}} \frac{1}{1-p^{-s}} = \frac{1}{1-2^{-s}} \cdot \frac{1}{1-3^{-s}} \cdot \frac{1}{1-5^{-s}} \cdot \frac{1}{1-7^{-s}} \cdots$$

The two expressions are rigorously equal for every complex number s whose real part exceeds 1.

Why should they be equal? One adds up reciprocals of powers of integers; the other multiplies together fractions involving only the primes. At first sight, there is no obvious reason for the two to coincide.

The reason lies in a simple property of the integers : the decomposition into prime factors is unique. Every positive integer

can be written, in exactly one way up to the order of its factors, as a product of prime numbers. The number 12 is $2 \times 2 \times 3$, the number 30 is $2 \times 3 \times 5$, and so on.

To move from the product to the sum, one expands each factor of the product as a geometric series:

$$\frac{1}{1 - p^{-s}} = 1 + p^{-s} + p^{-2s} + p^{-3s} + \cdots$$

and then multiply all these series together. Every term in the expanded product corresponds to a choice of a power of p for each prime p, and the product of those choices is an integer written in its prime factorization. Since each integer has a unique decomposition, each integer appears exactly once in the expanded product. The product thus recovers the full sum of the reciprocals of the integers.

This identity, known as the *Euler product*, is the first equation in the history of the *L-function*. It contains, in its purest form, the idea that runs through the whole theory: an analytic object, defined by a sum over the integers, equals an arithmetic object, defined by a product indexed by the primes. Every statement about one must translate, one way or another, into a statement about the other.

In 1859, Bernhard Riemann published a paper of eight pages that would turn this observation into theory. He showed that the function defined by the Euler sum and product, now written $\zeta(s)$, can be analytically continued to the whole complex plane, except at the single point $s = 1$. This continued function satisfies a functional equation relating its value at s to its value at $1 - s$. And above all, it has a family of zeros (values of s at which $\zeta(s) = 0$) whose location is conjectured to lie on a certain vertical line in the complex plane. That is the *Riemann Hypothesis*, which remains one of the open problems in mathematics with the longest list of

failed attempts.

For our purposes, the Riemann Hypothesis itself is not central. What is, is the pattern Riemann laid bare: an arithmetic object (the sequence of primes) generates an analytic object (the function ζ) whose analytic properties encode arithmetic information. This pattern will repeat.

Peter Gustav Lejeune Dirichlet had anticipated it before Riemann. In 1837, to prove that there are infinitely many primes in every arithmetic progression $a, a+N, a+2N, \ldots$ (when a and N share no common divisor), he attached to each *Dirichlet character* χ modulo N a function analogous to ζ:

$$L(s,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}.$$

A Dirichlet character is, roughly, a way of colouring the integers according to their remainder modulo N, subject to certain multiplicative rules. The functions so obtained are called the *Dirichlet L-functions*. They too have an Euler product, an analytic continuation, and a functional equation. In the classification that emerged later, they are the simplest L-functions after ζ.

The pattern generalizes further. What began with Euler as a coincidence between a sum and a product, what Riemann turned into an analytic theory, what Dirichlet extended to characters modulo N, turned out to be a general principle. To every continuous Galois representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(n, K)$ (subject to a technical ramification condition), one attaches an L-function:

$$L(s,\rho) = \prod_p L_p(s,\rho),$$

$$L_p(s,\rho) = \det\begin{pmatrix} I - \rho(\mathrm{Frob}_p) & \\ & \cdot\, p^{-s} \end{pmatrix}^{-1}.$$

the product running over the primes at which ρ is *unramified*, with modified local factors at the exceptional primes. The Frobe-

nius eigenvalues reappear in this product, gathered into a single analytic object. For $n = 1$ with ρ trivial, one recovers ζ. For $n = 1$ with ρ a Dirichlet character, one recovers Dirichlet's $L(s, \chi)$. Higher dimensions give richer L-functions, but the structure is the same.

On the other side of Langlands's letter there is a parallel construction. To every automorphic representation π on an adelic group $\mathrm{GL}(n, \mathbb{A}_{\mathbb{Q}})$, one attaches an L-function, again as a product over the primes. The Japanese mathematician Ichiro Satake showed, in the 1960s, how to extract from the local component π_v a diagonal $n \times n$ matrix, called the *Satake parameter*, which condenses the information π carries at the place v. The local factor of the automorphic L-function at v is obtained from this matrix by the same recipe as on the Galois side. At ramified places the construction is more technical, but it is perfectly well defined.

On both sides, one obtains the same type of object. An L-function on the Galois side, generated by a representation ρ. An L-function on the automorphic side, generated by a representation π. Each one is a Dirichlet series converging in a half-plane, admitting an analytic continuation to the whole complex plane, and satisfying a functional equation.

When a Galois representation ρ and an automorphic representation π produce the same L-function, a whole web of identities follows. The Frobenius eigenvalues of ρ at each prime coincide with the Satake parameters of π at the same prime. The functional equations agree. The conductors (another numerical invariant attached to both sides) are the same. The local data, translated through the L-function, match on both sides. Such an agreement, when it holds, runs all the way through the structure of both objects. It means that ρ and π are two presentations of the same underlying object, seen through two different vocabularies.

The L-function is the bridge between the two shores: the instrument by which two objects living in worlds that share no immediate vocabulary can be set side by side. If the L-functions

agree, the correspondence is at work.

In 1967, the whole apparatus needed to state this idea was in place. Riemann had given ζ its analytic continuation and functional equation a century before. Dirichlet had extended the theory to characters. Hecke had shown, in the 1930s, that modular forms too generate L-functions satisfying analogous functional equations. The conjecture that an elliptic curve over $\mathbb{Q}$ should have an L-function agreeing with that of a modular form was already circulating, through the work of Taniyama, Shimura, and Weil. John Tate's thesis, defended at Princeton in 1950, had reformulated Dirichlet's theory in the adelic language. Satake had given the recipe for extracting parameters from the local unramified representations. What was missing was only the statement of the correspondence itself. Langlands was about to supply it.

Part II

The Letter

5
Princeton, January 1967

André Weil received the letter a few days after it was sent. He opened it in his office at the Institute for Advanced Study, a few hundred metres from the Fine Hall office where it had been written. He unfolded the seventeen handwritten pages, read the first one, stopped, went back to the start, reread. What he held in his hands was not an ordinary article, not an isolated conjecture. It was an overall plan for an entire branch of number theory, laid out with the modesty of a man who expected to be dismissed, and the precision of a man who would be proved right.

The first concrete claim of the letter extends a theory that was already proved, called *class field theory*. That theory, completed in the first half of the twentieth century by Hilbert, Takagi, Artin, Chevalley, and a few others, describes what are called the *abelian extensions* of a number field. An extension is abelian when its Galois group is abelian, that is, when its symmetries commute with each other. For the rational numbers $\mathbb{Q}$, class field theory establishes that every abelian extension corresponds to a *Hecke character*, the adelic counterpart of Dirichlet characters. And the two L-functions, the one from the Galois extension and the one from the Hecke character, coincide exactly.

In the vocabulary Langlands will introduce, this situation is called the correspondence for $n = 1$. On the Galois side, a representation of dimension one, a 1×1 matrix, which is to say a single number. On the automorphic side, a Hecke character. L-functions as the seal of the match. Everything is proved. Each pairing can be written out explicitly.

Langlands proposes to extend the claim to $n = 2$, then