GALOIS REPRESENTATIONS AND MODULAR FORMS

Kenneth A. Ribet

University of California, Berkeley

In this article, I discuss material which is related to the recent proof of Fermat's Last Theorem: elliptic curves, modular forms, Galois representations and their deformations, Frey's construction, and the conjectures of Serre and of Taniyama-Shimura.

1. Introduction

This article is a revised version of the notes which were distributed at my Progress in Mathematics lecture at the August, 1994 Minneapolis Mathfest. When I was first approached in 1993 by the Progress in Mathematics committee, I was asked to discuss "the mathematics behind Andrew Wiles's solution of the Fermat conjecture." As the reader is no doubt aware, Wiles had announced in a series of lectures at the Isaac Newton Institute for Mathematical Sciences that he was able to prove for semistable elliptic curves the conjecture of Shimura and Taniyama to the effect that all elliptic curves over $\mathbf{Q}$ are "modular" (i.e., attached to modular forms in a sense which will be explained below). Fermat's Last Theorem follows from this result, together with a theorem that I proved seven years ago [62].

At the time of the Mathfest, however, a gap which had appeared in Wiles's work had not yet been repaired (see [33])—Fermat's "Theorem" was still a conjecture. Nevertheless, it was readily apparent that the methods introduced by Wiles were significant and deserving of attention. Most notably, these methods had been used to construct for the first time an infinite set $\mathcal{E}$ of modular elliptic curves over $\mathbf{Q}$ with the following property: if E_1 and E_2 are elements of $\mathcal{E}$, then E_1 and E_2 are non-isomorphic even when viewed as elliptic curves over the complex field $\mathbf{C}$ (see [69]). In my lecture at the Mathfest, I stressed this achievement of Wiles and discussed the analogy between the Taniyama-Shimura conjecture and conjectures of Serre [75] about two-dimensional Galois representations.

As I began to revise these notes, I found that the situation had changed dramatically for the better. Wiles announced in October, 1994 that the bound he sought in his original proof could be obtained by a method which circumvented the Euler system construction [34]. This method arose from an observation that Wiles made early in his investigation: the required upper bound would follow from a proof that certain Hecke algebras are complete intersection rings. This statement about Hecke

---

This article was prepared while I was a research professor at MSRI, where research is supported in part by NSF grant DMS-9022140. My work was to some extent supported by NSF Grant DMS 93-06898. I wish to thank N. Boston, H. Darmon, F. Diamond, F. Q. Gouvêa, B. Mazur, V. K. Murty, C. O'Neil, S. Ribet, A. Silverberg, R. Taylor and A. Wilkinson for feedback on drafts of these notes. I am especially grateful to B. Conrad and L. Goldberg for detailed comments.

algebras is the main theorem of an article written jointly by Richard Taylor and Andrew Wiles [88]; these authors announced their result at the same time that Wiles disseminated a revised version of his original manuscript [92].

It is premature to undertake a detailed analysis of the results of Wiles and Taylor-Wiles. The aim of this survey is more modest: to present an introduction to the circle of ideas which form the background for these results. Because of the intense publicity surrounding Fermat's Last Theorem, a good deal of the material I have chosen has been discussed in news and expository articles which were written in connection with Wiles's 1993 announcement. Among these are the author's news item in the Notices [65] and his article with Brian Hayes in American Scientist [30], two pieces in the American Mathematical Monthly [13, 28], the report by K. Rubin and A. Silverberg in this Bulletin [68], a long survey by H. Darmon [14], an elementary formulation of the Langlands program by A. Ash and R. Gross [1], and an article for undergraduates and their teachers by N. Boston [6]. Books containing articles related to Fermat's Last Theorem are soon to appear [10, 55]. Furthermore, two videotapes related to Fermat have been in circulation for some time: a 1993 lecture by the author is available from the AMS [66], and the Mathematical Sciences Research Institute has been distributing a videotape based on the July, 1993 "Fermat Fest" which was organized by the MSRI and the San Francisco Exploratorium. Readers may also consult material available by gopher from `e-math.ams.org`.

In view of the burgeoning literature in this subject, I imagined these notes mostly as a somewhat biased guide to reference works and expository articles. In the end, what I have written might best be characterized as an abbreviated survey with a disproportionately large list of references.

2. Modular forms

We begin by summarizing some background material concerning modular forms. Among reference books in the subject, one might cite [36], [53], and [80]. Also, several books on elliptic curves contain substantial material on modular forms; in particular, Knapp's book [35] has been recommended to the author as a good source for an overview of the Eichler-Shimura theory relating elliptic curves to certain modular forms. For a first introduction to modular forms, a fine starting point is [73, Ch. VII].

The modular forms to be considered are "cusp forms of weight two on $\Gamma_0(N)$," for some integer $N \geq 1$. Here, $\Gamma_0(N)$ is the group of integer matrices with determinant 1 which are upper-triangular mod N. A cusp form on this group is, in particular, a holomorphic function on the upper half-plane $\mathcal{H}$ consisting of complex numbers with positive imaginary part.

These functions are usually presented as converging Fourier series $\sum_{n=1}^{\infty} a_n q^n$, where $q := e^{2\pi i z}$. For the forms which most interest us, the complex numbers a_n are algebraic integers; frequently they are even ordinary integers.

The weight-two cusp forms on $\Gamma_0(N)$ are holomorphic functions $f(z)$ on the complex upper half-plane $\mathcal{H}$. On requires principally that $f(z)\,dz$ be invariant under $\Gamma_0(N)$, i.e., that $f(z)$ satisfy the functional equation

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z)$$

for all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$. (In particular, one has $f(z+b) = f(z)$ for all integers b.) In addition to the holomorphy and the functional equation, one imposes subsidiary conditions at infinity [80, Ch. 2].

The group $\Gamma_0(N)$ acts on $\mathcal{H}$ by fractional linear transformations, with $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ acting as

$$z \mapsto \frac{az+b}{cz+d}.$$

The invariance of $f(z)\,dz$ means that $f(z)\,dz$ arises by pullback from a differential on the quotient $Y_0(N) := \Gamma_0(N)\backslash\mathcal{H}$. This quotient is a non-compact Riemann surface with a standard compactification, known as $X_0(N)$. The complement of $Y_0(N)$ in $X_0(N)$ is a finite set, the set of *cusps* of $X_0(N)$. The conditions on infinity satisfied by a cusp form require the differential on $Y_0(N)$ associated to $f(z)$ to extend to $X_0(N)$. This means simply that $f(z)$ is required to vanish at the cusps.

Let us identify the space $S(N)$ of weight-two cusp forms on $\Gamma_0(N)$ with the space of holomorphic differentials on the Riemann surface $X_0(N)$. Then $S(N)$ has finite dimension: $\dim S(N)$ is the genus of the curve $X_0(N)$. This integer may be calculated easily via the Riemann-Hurwitz formula (applied to the covering $X_0(N) \to X_0(1)$ which corresponds to the inclusion of $\Gamma_0(N)$ in $\Gamma_0(1)$). For example, $\dim S(N)$ is zero for $N \leq 10$ and is one for $N = 11$. (See [80, Ch. 2].)

In order to present an example of a non-zero cusp form, we will exhibit a non-zero element of the 1-dimensional space $S(11)$. Namely, consider the formal power series with integral coefficients $\sum a_n X^n$ which is defined by the identity

$$X \prod_{m=1}^{\infty} (1 - X^m)^2 (1 - X^{11m})^2 = \sum_{n=1}^{\infty} a_n X^n.$$

It can be shown that the holomorphic function $\sum a_n q^n$ (with $q = e^{2\pi i z}$) is a cusp form h of weight two on $\Gamma_0(11)$, i.e., a generator of the 1-dimensional space $S(11)$, cf. [80, Example 2.28]. (One has $h(z) = \eta(z)^2\eta(11z)^2$, where $\eta(z)$, the Dedekind η-function, is the standard example of a modular form of weight 1/2.)

For each integer $n \geq 1$, the nth *Hecke operator* on $S(N)$ is an endomorphism T_n of $S(N)$, whose action is generally written on the right: $f \mapsto f|T_n$. The various T_n commute with each other and are interrelated by identities which express a given T_n in terms of the Hecke operators indexed by the prime factors of n. If p is a prime, the operator T_p is defined as a composite $\beta \circ \alpha$, where $\alpha\colon S(N) \to S(Np)$ and $\beta\colon S(Np) \to S(N)$ are standard homomorphisms, known as "degeneracy maps." In terms of Fourier coefficients, T_p is given by the following rules: If p is not a divisor of N and $f = \sum a_n q^n$, then $f|T_p$ has the Fourier expansion

$$\sum_{n=1}^{\infty} a_{np} q^n + p \sum_{n=1}^{\infty} a_n q^{pn}.$$

If p divides N, then $f|T_p$ is given by $\sum_{n=1}^{\infty} a_{np} q^n$. The T_n have a strong tendency to be diagonalizable on $S(N)$: for n prime to N, T_n is self-adjoint with respect to the Petersson pairing on $S(N)$, and is therefore semisimple [80, §3.5].

The elements of $S(N)$ having special arithmetic interest are the *normalized eigenforms* in $S(N)$; these are the non-zero cusp forms $f = \sum a_n q^n$ which are eigenvectors for all the T_n and which satisfy the normalizing condition $a_1 = 1$. (The latter condition is imposed mostly for convenience, since an arbitrary eigenform is a multiple of a normalized one.) If f is such an eigenform, its Fourier coefficients and its eigenvalues coincide: one has $f|T_n = a_n f$ for all $n \geq 1$. The normalized eigenforms are "arithmetic" because the a_n belong to the realm of algebraic number theory. In fact, if f is a normalized eigenform, then the subfield of $\mathbf{C}$ generated by the a_n is a finite (algebraic) extension E of $\mathbf{Q}$ in $\mathbf{C}$, and the elements a_n of E are algebraic integers (see [80, Th. 3.52] or [16, Prop. 2.7]).

It is a mildly complicating fact that the normalized eigenforms in $S(N)$ do not always form a basis of $S(N)$. In other words, the commuting operators T_n are not necessarily all diagonalizable. This problem, which arises from those n which have a common factor with N, can be repaired by the introduction of *newforms* [2]. Briefly, a newform is a normalized eigenform $f = \sum a_n q^n$ for which the space

$$\left\{ g \in S(N) \,\middle|\, g|T_n = a_n g \text{ for all } n \text{ prime to } N \right\}$$

contains only f and its multiples. Atkin and Lehner showed in 1970 that $S(N)$ has a basis built out of suitable transforms of the newforms in the spaces $S(M)$, where M runs over the positive divisors of N. Namely, for each M, there are natural embeddings of the space $S(M)$ into $S(N)$ which are indexed by the positive divisors of N/M. Namely, to each divisor d of N/M is associated the embedding $\delta_d \colon S(M) \hookrightarrow S(N)$ which sends $\sum a_n q^n$ to the series $\sum a_n q^{nd}$. The embeddings δ_d are *degeneracy mappings*; one such mapping was introduced briefly above in connection with the definition of the pth Hecke operator T_p. With the help of these degeneracy mappings, the full space $S(N)$ may be reconstructed purely in terms of the newforms in the spaces $S(M)$ with M dividing N. Specifically, let $S(M)^{\text{new}}$ be the subspace of $S(M)$ spanned by its newforms. Then

$$S(N) = \bigoplus_{M|N} \bigoplus_{d|\frac{N}{M}} \delta_d\big(S(M)^{\text{new}}\big).$$

We conclude this discussion with the remark that the work of Atkin and Lehner was generalized by T. Miyake [52] in the setting of automorphic forms on $\mathbf{GL}(2)$ and then by W. Li [42] in the setting of classical modular forms.

3. Elliptic curves

We next introduce some foundational concepts pertaining to elliptic curves. For an in-depth treatment of these concepts, the reader may consult the large number of textbooks and monographs which focus on elliptic curves. (As was indicated above, some of these books discuss modular forms as well.) Rather than list these references here, we refer the reader to the bibliography of a recent book review written by W. R. Hearst III and the author [31]. One book which has appeared since that review was written is the recent "Advanced Topics in the Arithmetic of Elliptic Curves" by J. H. Silverman [85].

Elliptic curves are distinguished by the fact that they are simultaneously *curves*, i.e., varieties of dimension 1, and *abelian varieties*, projective varieties which are endowed with a group law. The definition is very simple: an elliptic curve over a field k is a projective non-singular curve A of genus 1 over k which is furnished with a distinguished rational point O. An elliptic curve is thus a pair (A, O); an isomorphism between two pairs (A, O) and (A', O') is an isomorphism $A \xrightarrow{\sim} A'$ which maps O to O'. By convention, one usually omits explicit mention of the point O and refers simply to an elliptic curve A over k.

One obtains elliptic curves from 5-tuples of elements of k: to $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_6)$ one associates the projective curve whose affine equation is the generalized Weierstraß equation

$$y^2 + \alpha_1 xy + \alpha_3 y = x^3 + \alpha_2 x^2 + \alpha_4 x + \alpha_6.$$

This curve has genus 1 if it is non-singular; the non-singularity is detected by the non-vanishing of the *discriminant* of the equation. (The discriminant is a polynomial in the α_i which will not be reproduced here; it can be found, for instance, on page 36 of [86].) A short calculation shows that the plane curve with the indicated equation has exactly one point which does not lie on the affine plane; we take this point to be the distinguished point O. Using the Riemann-Roch theorem, one checks that every elliptic curve over k is isomorphic to one obtained by this process. (See, e.g., [76, Ch. II] for a discussion of the Riemann-Roch theorem and [84, Ch. III, §3] for a proof that elliptic curves are given by generalized Weierstraß equations.)

As we have suggested, an elliptic curve A over k may be viewed as a commutative *algebraic group* over k. Concretely, suppose that one has chosen a Weierstraß equation for A as above. For each field K containing k, let $A(K)$ be the subset of the projective plane over K which is defined by the Weierstraß equation. This set is endowed with a classical group law, often known as the "chord and tangent operation." In this law, O is the identity element of $A(K)$, and three distinct elements of $A(K)$ sum to O if and only if they are collinear. The composition law on $A(K)$ can be described explicitly in terms of coordinates by a family of polynomial equations with coefficients in k; this family depends on the chosen Weierstraß equation, but is independent of K. (For a recent discussion concerning families of such equations, see [5].)

If A is an elliptic curve over $\mathbf{Q}$, then one can choose the Weierstraß coefficients α_i to be rational *integers*. The coefficients are essentially unique if we demand that the discriminant of the equation have the smallest possible value; the discriminant is then said to be the (minimal) discriminant Δ of A.

Those prime numbers which divide Δ are the primes at which A has *bad reduction*; the others are the primes at which A has *good reduction*. The point is that, for good primes p, the minimal equation, when viewed mod p, yields an elliptic curve $\tilde{A}_p$ over the finite field $\mathbf{Z}/p\mathbf{Z}$. The reduced curve $\tilde{A}_p$ over $\mathbf{Z}/p\mathbf{Z}$ is a projective plane curve over $\mathbf{Z}/p\mathbf{Z}$; the group $\tilde{A}_p(\mathbf{Z}/p\mathbf{Z})$ consisting of the points of $\tilde{A}_p$ with coordinates in $\mathbf{Z}/p\mathbf{Z}$ is then a finite abelian group. A theorem of H. Hasse states that the order of this group is approximately p. More precisely, the integer

$$b_p := p+1 - \#(\tilde{A}_p(\mathbf{Z}/p\mathbf{Z}))$$

is bounded in absolute value by $2\sqrt{p}$ [84, Ch. V, Th. 1.1].

To work with a concrete example, let us introduce the elliptic curve A over $\mathbf{Q}$ defined by the equation $y^2 + y = x^3 - x^2$. The group $A(\mathbf{Q})$ has five evident points: the origin O, and the four affine points gotten by taking $y \in \{0, -1\}$ and $x \in \{0, 1\}$. One may verify that these points form a *subgroup* of $A(\mathbf{Q})$. Since they remain distinct under reduction mod p, $\tilde{A}_p(\mathbf{Z}/p\mathbf{Z})$ has a subgroup of order 5 whenever A has good reduction at p. Accordingly, the number $\#(\tilde{A}_p(\mathbf{Z}/p\mathbf{Z}))$ is divisible by 5, so that $b_p \equiv p{+}1 \bmod 5$ for all primes p which do not divide the discriminant of A.

Now Table 1 of [4] informs us that the minimal discriminant of A is -11. Therefore, integers b_p are defined for all $p \neq 11$. When p is small, it is not hard to compute b_p; one finds $b_2 = -2$, $b_3 = -1$, $b_5 = 1$, and so on. (The mod 5 congruence on the b_p, plus Hasse's bound $|b_p| < 2\sqrt{p}$, determines the first few of these integers.) Anticipating the conjecture of Shimura and Taniyama which will be discussed below, I would like to point out that the arithmetically defined b_p are known to agree with the prime-indexed coefficients a_p of the weight-two modular form h on $\Gamma_0(11)$ which was introduced above. The astounding identity $b_p = a_p$ is a special case of a relation which was discovered by Eicher and Shimura (and which is recalled in [79]). A priori, the Eichler-Shimura relation gives information only about those elliptic curves which are related geometrically to modular curves of the form $X_0(N)$. The Taniyama-Shimura conjecture states that *all* elliptic curves over $\mathbf{Q}$ can be so related.

Next, suppose that A and A' are elliptic curves over k. An *isogeny* $A \to A'$ over k is a non-constant map of curves over k which takes the distinguished point O of A to the corresponding point O' of A. If such a map exists, then it is a map of algebraic groups, and A and A' are said to be isogenous over k. (As the terminology suggests, the relation of being isogenous is symmetric; it is in fact an equivalence relation.) It is not difficult to show that two isogenous elliptic curves over $\mathbf{Q}$ have the same primes of bad reduction. (This follows from the criterion of Néron-Ogg-Shafarevich—see [77] or [84, Ch. VII, §7].) Moreover, if A and A' are isogenous over $\mathbf{Q}$, then for each good reduction prime p, one has

$$\#(\tilde{A}_p(\mathbf{Z}/p\mathbf{Z})) = \#(\tilde{A}'_p(\mathbf{Z}/p\mathbf{Z})).$$

In other words, the numbers b_p are the same for A and A', so that one might make the informal statement that A and A' are "equivalent arithmetically." A striking theorem of G. Faltings [19,§5] states that, conversely, two elliptic curves A and A' over $\mathbf{Q}$ are isogenous if $\#(\tilde{A}_p(\mathbf{Z}/p\mathbf{Z})) = \#(\tilde{A}'_p(\mathbf{Z}/p\mathbf{Z}))$ for all primes p at which the two curves have good reduction. (In [19], Faltings proves a more general statement about homomorphisms between abelian varieties over number fields, thus confirming conjectures of J. Tate.)

Each elliptic curve A over $\mathbf{Q}$ has a *conductor*, which is a positive integer divisible precisely by the primes at which A has bad reduction. For example, the curve with equation $y^2{+}y = x^3{-}x^2$ has conductor 11. Although the definition of the conductor is somewhat unenlightening (see for example [71, §2]), a well known algorithm of J. Tate [86] makes it possible to compute conductors by hand in specific cases. Alternatively, the computer algebra program `gp` [3], the Mathematica package [44] and the Maple package `apecs` all implement Tate's algorithm—and more generally make it easy to perform elliptic curve calculations on a workstation or personal computer. (The latter package is recommended by F. Gouvêa.)

Because the conductor is divisible exactly by the set of "bad primes," the conductor and the minimal discriminant of A are divisible by the same set of prime numbers. Nevertheless, these integers have a completely different feel. For one thing, the conductor of an elliptic curve depends only on its $\mathbf{Q}$-isogeny class, while the discriminant may change under isogeny. (The curve with equation $y^2+y = x^3-x^2-10x-20$ is isogenous to the conductor-11 curve that we have been discussing; its conductor is 11, but its minimal discriminant is -11^5.) For another, the conductor is always positive; the minimal discriminant may be positive or negative. Finally, the latter number may be divisible by a large power of a given prime, whereas the conductor of an elliptic curve tends to be divisible only by low powers of its prime divisors. (In particular, a prime $p \geq 5$ can divide the conductor at most to the second power.) A beautiful formula of A. Ogg [57] expresses the conductor of a given curve A in terms of the minimal discriminant and the Néron model of A. (The latter is a curve over $\mathbf{Z}$ which may be regarded as the "best possible" model for A. See also [84, App. C, §16] for a statement of Ogg's formula.)

Suppose that the conductor of A is N. Then A is said to be *semistable* at a prime number p if p^2 does not divide N. This means either that p is prime to N, in which case A has good reduction at p, or else that p "exactly" divides N, in which case the reduction of A at p is bad—but not too bad (it is said to be multiplicative). If A is semistable at all primes p, i.e., if N is square free, then the elliptic curve A is said to be semistable. Semistable elliptic curves occur in connection with Fermat's Last Theorem, and in other applications.

This is a good point to insert a short digression about abelian varieties, whose arithmetic theory is exposed in various chapters of [12] and the recent graduate-level text [54]; see also the books by such authors as A. Weil, S. Lang [37], D. Mumford, H. P. F. Swinnerton-Dyer, G. Kempf and H. Lange. As was mentioned very briefly above, an abelian variety over a field k is a projective algebraic variety over k which is equipped with the structure of an algebraic group over k. The first examples of abelian varieties are obtained by taking Jacobians of curves; the Jacobian of a curve of genus g over k is an abelian variety over k of dimension g. (For the construction of Jacobians, the reader may consult, e.g., Chapter V of [76].) If A is an elliptic curve over k, then the chosen origin O of A enables one to identify A (endowed with its group structure) with the Jacobian of A. In fact, an elliptic curve over k is nothing other than an abelian variety over k of dimension one. Since an abelian variety of dimension one is an elliptic curve (and since an abelian variety of dimension zero is just a single point), one might describe abelian varieties as "higher-dimensional analogues" of elliptic curves.

Abelian varieties play an inevitable role in our story because the conjecture which we are about to discuss (a priori one involving elliptic curves over $\mathbf{Q}$ and weight-two newforms with integer coefficients) extends naturally to a conjectural dictionary between arbitrary newforms of weight two and a certain class of abelian varieties over $\mathbf{Q}$ (see Conjecture 2 below).

4. The Taniyama-Shimura conjecture

The conjecture of Shimura and Taniyama relates elliptic curves over $\mathbf{Q}$ and certain modular forms. Its history is the subject of an engrossing "file" compiled by S. Lang [38]. If I understand correctly, the conjecture was first posed as a tentative question by Taniyama at the Tokyo-Nikko conference of 1955. Shimura stated the conjecture in its present form in the early 1960s. In 1968, A. Weil proved a theorem to the effect that certain necessary conditions for an elliptic curve to be "modular" are, in fact, sufficient [91]. Because of this theorem, Weil's name has sometimes been attached to the conjecture.

Before giving a formal description of the conjecture, we refer once again to the elliptic curve $y^2 + y = x^3 - x^2$, which has conductor 11. The integers b_p which control the numbers of mod p points on this curve might be viewed initially as mysterious quantities for which we seek a "formula." The relation $b_p = a_p$, where a_p is the pth coefficient of the normalized eigenform in $S(11)$ may be regarded as such a formula.

The Taniyama-Shimura conjecture affirms that there is an analogous relation for *all* elliptic curves over $\mathbf{Q}$. Namely, if A is an elliptic curve over $\mathbf{Q}$ of conductor N, then one conjectures that there is a newform $f = \sum a_n q^n$ in $S(N)$ such that $b_p = a_p$ for all primes $p \nmid N$. All coefficients a_n of f are then necessarily rational integers. (For another account of the conjecture, see the article by K. Rubin and A. Silverberg [68, §1].)

A geometric formulation of the Taniyama-Shimura conjecture may be given in terms of the construction presented in Chapter 7 of [80] and, from a different perspective, in [83]. Suppose that $f \in S(N)$ is a normalized eigenform and let E be the field generated by the coefficients of f. Shimura associates to f an abelian variety A_f over $\mathbf{Q}$ whose dimension is the degree $[E\colon \mathbf{Q}]$ and whose arithmetic incorporates the eigenvalues a_p for p prime to N. Although the construction of A_f is perfectly precise, it is fruitful in this context to regard the association $f \rightsquigarrow A_f$ as a flabby one linking f to a "clump" of isogenous abelian varieties, rather than a specific abelian variety which is singled out up to isomorphism. (The relevant notion of isogeny is an appropriate generalization of the notion of isogeny for elliptic curves.)

If the Fourier coefficients of f are rational integers, then $E = \mathbf{Q}$ and A_f has dimension 1. This means that A_f is an elliptic curve over $\mathbf{Q}$, which we shall regard as being defined only up to isogeny. According to a theorem of Eichler and Shimura, the eigenvalues a_p are reflected in the arithmetic of the elliptic curve $A = A_f$ in the following way. If p does not divide N, then A has good reduction at p. Further, for such p, the pth Fourier coefficient of f coincides with the quantity $b_p := p+1-\#(\tilde{A}_p(\mathbf{Z}/p\mathbf{Z}))$. In other words, one has $b_p = a_p$ for all $p \nmid N$. (If A_f has dimension greater than 1, the relation between f and A_f is a bit more complicated to formulate, but involves no new ideas.)

In its geometric form, the Taniyama-Shimura conjecture states that every elliptic curve A over $\mathbf{Q}$ is *modular* in the sense that it is isogenous to some curve A_f. In other words, the conjecture asserts the surjectivity of the construction $f \rightsquigarrow A_f$, viewed as a map from eigenforms with integral coefficients (in some $S(N)$) to isogeny classes of elliptic curves over $\mathbf{Q}$. In analogy with the arithmetic formulation, when A has conductor N, A is conjectured to be isogenous to an A_f with $f \in S(N)$.

The connection between the two formulations is as follows. Suppose that A is an elliptic curve over $\mathbf{Q}$ of conductor N, and assume that A is related arithmetically to an eigenform in $S(N)$. Specifically, assume that $f = \sum a_n q^n$ is a normalized eigenform in $S(N)$ with integral coefficients and that $a_p = b_p$ for all p prime to N. (The b_p have their usual meaning.) Let A' be the elliptic curve A_f, and write b'_p for the analogues of the b_p for A'. Then one has $b'_p = a_p$ for all $p \not| N$ by the formula of Eichler-Shimura, and hence $b_p = B'_p$ for all $p \not| N$. A theorem of Faltings which was quoted above then ensures that A and $A' = A_f$ are isogenous over $\mathbf{Q}$, so that the geometric form of the Taniyama-Shimura conjecture is true for A. Conversely, if A is isogenous to an A_f, then numbers b_p computed for A coincide with their analogues for A_f. By the Eicher-Shimura formula, these latter numbers are the prime-indexed coefficients of f.

In connection with the construction $f \rightsquigarrow A_f$, let us consider the situation where f is a normalized eigenform in $S(N)$ with integral Fourier coefficients but where f is not necessarily a newform. What is the conductor of the elliptic curve $A = A_f$? The answer to this question begins with the fact that f is necessarily built from a newform in $S(M)$, for some unique divisor M of N. It is then true that the conductor of A is precisely this divisor M. This theorem was proved by H. Carayol in [9], following work of Shimura, Igusa, Deligne and Langlands.

A formulation of the Taniyama-Shimura conjecture with a completely different flavor is provided by Mazur's article [48]. In this article, Mazur rephrases the conjecture as a statement about the Riemann surface associated with an elliptic curve over $\mathbf{Q}$. If A is such an elliptic curve, we write $A(\mathbf{C})$ for this Riemann surface, which may be realized as the subset of the complex projective plane which is defined by a Weierstraß equation for A. This surface is holomorphically a complex torus. For each integer N, consider the subgroup $\Gamma_1(N)$ of $\Gamma_0(N)$ consisting of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$ for which $a \equiv d \equiv 1 \bmod N$. By considering $\Gamma_1(N)$ in place of $\Gamma_0(N)$, one obtains an analogue of $X_0(N)$ which is called $X_1(N)$. Mazur shows that an elliptic curve A over $\mathbf{Q}$ is modular if and only if there exists a non-constant holomorphic map from $X_1(N)$ to $A(\mathbf{C})$, for some positive integer N. As Mazur explains in his article, one can paraphrase this condition as the statement that the arithmetic object A is *hyperbolically uniformized*, since $\mathcal{H}$ is a model for the hyperbolic plane. This circumstance has led to the charge that Wiles's proof of Fermat's Last Theorem could not possibly be correct since its logical structure involves a statement that may be interpreted in terms of hyperbolic geometry [90]. For a refutation, see [7].

Mazur's observation led Serre to ask for a description of the set of elliptic curves over the complex field which satisfy Mazur's condition: Which elliptic curves A over $\mathbf{C}$ are modular in the sense that one can find a non-constant holomorphic map $X_1(N) \to A(\mathbf{C})$ for some positive integer N? (According to Mazur's theorem, an elliptic curve over $\mathbf{Q}$ is modular in the usual sense if and only if it is modular in this new sense when viewed over $\mathbf{C}$.) In [64], I provide a conjectural answer to Serre's question. Namely, I show that the conjectures made by Serre in [75] imply the following statement, which is similar in spirit to the analysis in §10 of [82].

Conjecture 1. *Let A be an elliptic curve over $\mathbf{C}$. Then A is modular if and only if A is isogenous to all elliptic curves ${}^{\sigma}A$ obtained by conjugating A by algebraic automorphisms σ of the field $\mathbf{C}$.*

Conjecture 1 is related to Serre's conjectures via a second conjecture, which we will state after some motivating remarks. Consider a normalized eigenform $f = \sum a_n q^n$ in the space of weight-two cusp forms on $\Gamma_1(N)$. Let $E = \mathbf{Q}(\ldots, a_n, \ldots)$ be the field generated by the coefficients of f, and let $d = [E\colon \mathbf{Q}]$ be the degree of E, i.e., the dimension of E as a $\mathbf{Q}$-vector space. The abelian variety A_f is an abelian variety over $\mathbf{Q}$ of dimension d which comes equipped with an action of E. To give sense to the last statement, one introduces the ring $\mathrm{End}(A_f)$ whose elements are maps $A_f \to A_f$ in the category of algebraic varieties over $\mathbf{Q}$ which respect the group structure on A_f. (Such maps are the endomorphisms of A_f.) The ring $\mathrm{End}(A_f)$ turns out to be a free rank-d module over $\mathbf{Z}$; the associated $\mathbf{Q}$-algebra $\mathrm{End}(A_f)\otimes\mathbf{Q}$ is isomorphic to E. Thus A_f has many endomorphisms; moreover, it is "modular" in the sense that there is a non-constant map $X_1(N) \to A$ which is defined over $\mathbf{Q}$.

Conjecture 2. *Let A be an abelian variety over $\mathbf{Q}$ for which $\mathrm{End}(A) \otimes \mathbf{Q}$ is a number field of degree equal to $\dim A$. Then for some $N \geq 1$, there is a non-constant map $X_1(N) \to A$ which is defined over $\mathbf{Q}$.*

In [64], I prove that Serre's conjectures imply Conjecture 2 and that Conjecture 2 implies Conjecture 1.

It is natural to regard Conjecture 1 and Conjecture 2 as generalizations of the Taniyama-Shimura conjecture. The first conjecture pertains to elliptic curves which are not necessarily defined over $\mathbf{Q}$, while the second pertains to abelian varieties over $\mathbf{Q}$ which are not necessarily elliptic curves. Neither of these conjectures is proved in [92]. The Taniyama-Shimura conjecture can be generalized still further. Indeed, a common generalization of Conjectures 1 and 2 will be presented in a forthcoming work of E. Pyle. See also [22].

5. Galois representations attached to elliptic curves

Let A be an elliptic curve over $\mathbf{Q}$. A model for $A(K)$ when $K = \mathbf{C}$ is given by the Weierstraß theory of complex analysis: the group $A(\mathbf{C})$ is the complex torus $\mathbf{C}/L$, where L is the lattice of periods associated to the given cubic equation. (Explicitly, L is obtained by integrating the differential $\dfrac{dx}{y}$ on A over the free abelian group $H_1(A(\mathbf{C}), \mathbf{Z})$ of rank two.) Let n be a positive integer, and let $A[n]$ be the group of elements of $A(\mathbf{C})$ whose order divides n. This group of *n-division points* on A may be modeled as $\frac{1}{n}L/L$; it is therefore a free module of rank two over $\mathbf{Z}/n\mathbf{Z}$, since L is free of rank two over $\mathbf{Z}$.

On reflection, one sees that $A[n]$ in fact lies in $A(\overline{\mathbf{Q}})$, where $\overline{\mathbf{Q}}$ is the subfield of $\mathbf{C}$ consisting of all algebraic numbers. Indeed, the group $A[n]$ is a finite subgroup of $A(\mathbf{C})$ which consists of those points satisfying a certain set of polynomial equations with rational coefficients; it follows that the coordinates are algebraic numbers. Moreover, let $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ be the group of automorphisms of $\overline{\mathbf{Q}}$. Then the same reasoning shows that $A[n]$ is stable under the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $A(\overline{\mathbf{Q}})$ which results from the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $\overline{\mathbf{Q}}$. Thus $A[n]$ comes equipped with a canonical action of the Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

It is important to observe, for each $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, that the automorphism $P \mapsto {}^{\sigma}P$ is a *group* automorphism of $A[n]$; in symbols, ${}^{\sigma}(P + Q) = {}^{\sigma}P + {}^{\sigma}Q$ for $P, Q \in A[n]$. This equation is a consequence of the fact that the composition law which expresses $P+Q$ in terms of P and Q involves only polynomials with rational coefficients. Our observation means that the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $A[n]$ may be viewed as a (continuous) homomorphism

$$\rho_{A,n}\colon\ \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(A[n])$$

in which $\mathrm{Aut}(A[n])$ stands for the group of automorphisms of $A[n]$ as an *abelian group*. Since $A[n]$ is isomorphic to the group $(\mathbf{Z}/n\mathbf{Z})^2$, one has

$$\mathrm{Aut}(A[n]) \approx \mathbf{GL}(2, \mathbf{Z}/n\mathbf{Z}),$$

where the group on the right consists of two-by-two invertible matrices with coefficients in $\mathbf{Z}/n\mathbf{Z}$. While there is no canonical isomorphism between these groups, each choice of basis $A[n] \approx (\mathbf{Z}/n\mathbf{Z})^2$ determines such an isomorphism; moreover, the various isomorphisms obtained in this way differ by inner automorphisms of $\mathbf{GL}(2, \mathbf{Z}/n\mathbf{Z})$. Therefore, each element of $\mathrm{Aut}(A[n])$ has a well-defined trace and determinant in $\mathbf{Z}/n\mathbf{Z}$.

It is often fruitful to fix a choice of basis $A[n] \approx (\mathbf{Z}/n\mathbf{Z})^2$ and to view $\rho_{A,n}$ as taking values in the matrix group $\mathbf{GL}(2, \mathbf{Z}/n\mathbf{Z})$. Once this choice is made, $\rho_{A,n}$ becomes matrix-valued *representation* of the Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$; it is the representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ defined by the group of n-division points of A. The kernel of this representation corresponds, via Galois theory, to a finite Galois extension K_n of $\mathbf{Q}$ in $\overline{\mathbf{Q}}$. Concretely, this extension is obtained by adjoining to $\mathbf{Q}$ the coordinates of the various points in $A[n]$. The Galois group $G_n := \mathrm{Gal}(K_n/\mathbf{Q})$ may thus be identified with the image of $\rho_{A,n}$, which is a subgroup of the target group $\mathbf{GL}(2, \mathbf{Z}/n\mathbf{Z})$. The elliptic curve A and the positive integer n have given rise to a finite Galois extension $K_n/\mathbf{Q}$ whose Galois group is a subgroup of the group of two-by-two invertible matrices with coefficients in $\mathbf{Z}/n\mathbf{Z}$.

It is natural to ask for a description of G_n as a subgroup of $\mathbf{GL}(2, \mathbf{Z}/n\mathbf{Z})$, cf. [79]. There is a (relatively rare) special case to consider: that where A has *complex multiplication* (over $\mathbf{C}$). When A is viewed as $\mathbf{C}/L$, the complex multiplication case is that for which there is a complex number $\alpha \notin \mathbf{Z}$ such that $\alpha L \subseteq L$. The group G_n then has an abelian subgroup of index ≤ 2, so it is much smaller than the ambient group $\mathbf{GL}(2, \mathbf{Z}/n\mathbf{Z})$. In the more common case where A has no complex multiplication, Serre showed in [72] that the index of G_n in $\mathbf{GL}(2, \mathbf{Z}/n\mathbf{Z})$ is bounded as a function of n. In particular, $G_p = \mathbf{GL}(2, \mathbf{Z}/p\mathbf{Z})$ for all but finitely many primes p.

As background, we point out that the Taniyama-Shimura conjecture was proved for complex multiplication elliptic curves over $\mathbf{Q}$ by Shimura in 1971 [81]. The result of [81] is suggestive and may be regarded as evidence for the general case of the Taniyama-Shimura conjecture. As we recall below, however, the elliptic curves which appear in connection with Fermat's Last Theorem are semistable. And it is a fact that semistable elliptic curves over $\mathbf{Q}$ *never* have complex multiplication. (One possible proof can be summarized as follows: An elliptic curve with complex multiplication has an integral j-invariant, i.e., potentially good reduction. Hence if it is semistable, it has good reduction everywhere. However, a theorem of Tate states that there is no elliptic curve over $\mathbf{Q}$ with everywhere good reduction, cf. [21].) Accordingly, the theorem of [81] cannot be used to prove Fermat's Last Theorem.

A key piece of information about the extension $K_n/\mathbf{Q}$ (which depends on A as well as on n) is that its discriminant is divisible only by those prime numbers which divide either n or the conductor of A. In other words, if $p \nmid n$ is a prime number at which A has good reduction, then $K_n/\mathbf{Q}$ is unramified at p; one says frequently that the representation $\rho_{A,n}$ is unramified at p. Whenever this occurs, a familiar construction in algebraic number theory produces a Frobenius element σ_p in G_n which is well defined up to conjugation.

We shall now summarize this construction with K_n replaced by an arbitrary finite Galois extension K of $\mathbf{Q}$. Let $\mathcal{O}_K$ be the ring of algebraic integers in K. The Galois group $\mathrm{Gal}(K/\mathbf{Q})$ leaves $\mathcal{O}_K$ invariant, so that one obtains an induced action of $\mathrm{Gal}(K/\mathbf{Q})$ on the ideals of $\mathcal{O}_K$. The set of prime ideals $\mathfrak{p}$ of $\mathcal{O}_K$ which contain the prime number p is permuted under this action. For each $\mathfrak{p}$, the subgroup $D_\mathfrak{p}$ of $\mathrm{Gal}(K/\mathbf{Q})$ consisting of those elements in $\mathrm{Gal}(K/\mathbf{Q})$ which fix $\mathfrak{p}$ is called the *decomposition group* of $\mathfrak{p}$ in $\mathrm{Gal}(K/\mathbf{Q})$. Meanwhile, the finite field $\mathbf{F}_\mathfrak{p} := \mathcal{O}_K/\mathfrak{p}$ is a finite extension of the prime field $\mathbf{F}_p$. The extension $\mathbf{F}_\mathfrak{p}/\mathbf{F}_p$ is necessarily Galois; its Galois group is the cyclic group generated by the Frobenius automorphism

$$\varphi_p \colon x \mapsto x^p$$

of $\mathbf{F}_\mathfrak{p}$. There is a natural map $D_\mathfrak{p} \to \mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$, gotten by associating a given $\delta \in D_\mathfrak{p}$ to the automorphism of $\mathcal{O}_K/\mathfrak{p}$ induced by δ. This map is surjective; its injectivity is equivalent to the assertion that p is unramified in the extension $K/\mathbf{Q}$. Therefore, whenever this assertion is true, there is a unique $\sigma_\mathfrak{p} \in D_\mathfrak{p}$ whose image in $\mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$ is φ_p. The automorphism $\sigma_\mathfrak{p}$ is then a well defined element of $\mathrm{Gal}(K/\mathbf{Q})$, known as the Frobenius automorphism for $\mathfrak{p}$. It is easy to show that the various $\mathfrak{p}$ are all conjugate under $\mathrm{Gal}(K/\mathbf{Q})$ and that the Frobenius automorphism for the conjugate of $\mathfrak{p}$ by g is $g\sigma_\mathfrak{p}g^{-1}$. In particular, the various $\sigma_\mathfrak{p}$ are all conjugate; this justifies the practice of writing σ_p for any one of them and stating that σ_p is well defined up to conjugation.

For later use, we prolong this digression and introduce the concept of Frobenius elements in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Let p again be a prime number, and let $\mathfrak{p}$ now be a prime of $\overline{\mathbf{Q}}$ lying over p. (One can think of $\mathfrak{p}$ as a coherent set of choices of primes lying over p in the rings of integers of all finite extensions of $\mathbf{Q}$ in $\overline{\mathbf{Q}}$.) To $\mathfrak{p}$ we associate: (1) its residue field $\mathbf{F}_\mathfrak{p}$, which is an algebraic closure of the finite field $\mathbf{F}_p$, and (2) a decomposition subgroup $D_\mathfrak{p}$ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. There is again a surjective map $D_\mathfrak{p} \to \mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$. The Frobenius automorphism $\varphi_p \colon x \mapsto x^p$ generates the target group in the topological sense: the subgroup of $\mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$ consisting of powers of φ_p is dense in $\mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$. We shall use the symbol Frob_p to denote any preimage of φ_p in $D_\mathfrak{p}$ and refer to Frob_p as a Frobenius element for p in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

One thinks of Frob_p as a specific element of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, albeit one which is doubly ill-defined. The ambiguities in Frob_p result from the circumstance that $\mathfrak{p}$ needs to be chosen and from the fact that $D_\mathfrak{p} \to \mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$ has a large kernel, the inertia subgroup $I_\mathfrak{p}$ of $D_\mathfrak{p}$. The usefulness of Frob_p stems from the fact that the various $\mathfrak{p}$ are conjugate by elements of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, so that all the subgroups $D_\mathfrak{p}$ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ are conjugate, and similarly all $I_\mathfrak{p}$ are conjugate. Thus if ρ is a homomorphism mapping $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ to some other group, the kernel of ρ contains one $I_\mathfrak{p}$ if and only if it contains all $I_\mathfrak{p}$. In this case, one says that ρ is unramified at p; the image of Frob_p is then an element of the target group of ρ which is well defined up to conjugation. Note that one may write $\rho_{A,n}(\mathrm{Frob}_p) = \sigma_p$ for all primes p at which $\rho_{A,n}$ is unramified. As has been stated, these include all primes which divide neither n nor the conductor of A.

Let us return now to the matrix group G_n. We pointed out above that each element of G_n has a trace and determinant in $\mathbf{Z}/n\mathbf{Z}$ which are independent of any choice of basis. The Frobenius element σ_p is an element of G_n which is well defined only up to conjugation. Nevertheless, the trace and determinant of σ_p are well-defined, since conjugate matrices have the same traces and determinant. The number $\det \sigma_p$ is the residue class of p mod n. On the other hand, one has the striking congruence

$$\mathrm{tr}(\sigma_p) \equiv b_p \bmod n,$$

where b_p is the number $p+1 - \#(A(\mathbf{Z}/p\mathbf{Z}))$ introduced above. This means that the representation $\rho_{A,n}$ encapsulates information about the numbers b_p (for prime numbers p which are primes of good reduction and which are prime to n); more precisely, it determines the numbers b_p mod n.

6. Galois representations attached to modular forms

Suppose that $f \in S(N)$ is a normalized eigenform. The coefficients a_n of f are always algebraic integers, but not necessarily ordinary integers. If it happens that the a_n all lie in $\mathbf{Z}$, then the abelian variety $A = A_f$ is an elliptic curve. By considering the family $\rho_{A,n}$, one obtains a series of representations of the Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. These representations are related to f by the congruence $\mathrm{tr}(\sigma_p) \equiv a_p \bmod n$, valid for the nth representation and all primes $p \nmid nN$. We are especially interested in the case where n is a prime number ℓ; the ring $\mathbf{Z}/n\mathbf{Z}$ is then the *finite field* $\mathbf{F}_\ell$.

The representations ρ are associated to A, which in turn arises from f. Hence it is tempting to write $\rho_{f,\ell}$ for the representations $\rho_{A,\ell}$. The obstacle to doing this arises from the circumstance that A is determined up to isogeny, but not always up to isomorphism. If one replaces A by an isogenous elliptic curve, the representations $\rho_{A,\ell}$ may change! To circumvent this difficulty, we introduce the "semisimplifications" of the $\rho_{A,\ell}$.

These representations are defined as follows. If ρ is a two-dimensional representation of a group over a field, ρ is either irreducible, or else "upper-triangular," i.e., an extension of a one-dimensional representation α by another, β. In the case where ρ is irreducible, we declare its semisimplification to be ρ itself. In the reducible case, the semisimplification of ρ is the *direct sum* of the two one-dimensional representations α and β. Clearly, the trace and determinant are the same for ρ and for its semisimplification.

When A is fixed, the results of [72] show that $\rho_{A,\ell}$ can be reducible only for a finite number of ℓ. In fact, a theorem of Mazur [46] shows that $\rho_{A,\ell}$ is irreducible for all ℓ not in the set $\{2, 3, 5, 7, 13, 11, 17, 19, 37, 43, 67, 163\}$. Hence the replacement of $\rho_{A,\ell}$ by its semisimplification can be thought of as "fine tuning" which affects only a small number of the representations. One shows easily for all ℓ that the semisimplification of $\rho_{A,\ell}$ depends only on f and on ℓ (but not on the choice of A).

Introducing

$$\rho_{f,\ell} := \text{semisimplification of } \rho_{A,\ell},$$

one obtains a sequence of semisimple representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ which are well defined up to isomorphism. The characteristic property of $\rho_{f,\ell}$ may be summarized in terms of Frobenius elements Frob_p in the Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, elements which were introduced above. Namely, if p is a prime number not dividing ℓN, then $\rho_{f,\ell}(\mathrm{Frob}_p)$ has trace a_p mod ℓ and determinant p mod ℓ.

It is natural to generalize this process by considering the situation where $f = \sum a_n q^n$ is a normalized eigenform whose coefficients a_n are algebraic integers, but not necessarily rational integers. As we indicated above, the field $E = E_f$ generated by the a_n is a number field, i.e., a finite extension of $\mathbf{Q}$. Moreover, the coefficients a_n of f lie in the integer ring $\mathcal{O}_f$ of E. It is perhaps worth noting that the ring $\mathbf{Z}[\dots, a_n, \dots]$ generated by the a_n inside E, while a subring of finite index in $\mathcal{O}_f$, is not necessarily equal to $\mathcal{O}_f$.

Using the abelian variety A_f, one constructs representations indexed not by the prime *numbers*, but rather by the non-zero prime *ideals* of $\mathcal{O}_f$. (For details, see [80, Ch. 7].) If λ is such a prime, its residue field $\mathbf{F}_\lambda$ is a finite field, say of characteristic ℓ. The prime field $\mathbf{F}_\ell = \mathbf{Z}/\ell\mathbf{Z}$ is then canonically embedded in $\mathbf{F}_\lambda$. For each λ, one finds a semisimple representation $\rho_{f,\lambda}\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F}_\lambda)$ which is characterized up to isomorphism by the following property: if p is a prime number not dividing ℓN, then $\rho_{f,\lambda}(\mathrm{Frob}_p)$ has trace a_p mod λ and determinant p mod λ.

The assertion concerning the determinant of the matrices $\rho_{f,\lambda}(\mathrm{Frob}_p)$ may be rephrased as the statement that the determinant of the *representation* $\rho_{f,\lambda}$ is the mod ℓ cyclotomic character χ_ℓ. This character is defined by considering the group μ_ℓ of ℓth roots of unity in $\overline{\mathbf{Q}}$; the action of the Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the cyclic group μ_ℓ gives rise to a continuous homomorphism

$$\chi_\ell\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(\mu_\ell).$$

Since μ_ℓ is a cyclic group of order ℓ, its group of automorphisms is canonically the group $(\mathbf{Z}/\ell\mathbf{Z})^* = \mathbf{F}_\ell^*$. We emerge with a map $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{F}_\ell^*$, which is the character in question. The equality

$$\det \rho_{f,\lambda} = \chi_\ell$$

interpreted by viewing both homomorphisms as taking values in $\mathbf{F}_\lambda^*$.

Suppose now that $c \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is the automorphism "complex conjugation." Then the determinant of $\rho_{f,\lambda}(c)$ is $\chi_\ell(c)$. Now c operates on roots of unity by the map $\zeta \mapsto \zeta^{-1}$, since roots of unity have absolute value 1. Accordingly,

$$\det(\rho_{f,\lambda}(c)) = -1;$$

one says that $\rho_{f,\lambda}$ is *odd*.

This parity statement generalizes to modular forms "with Nebentypus" whose weights are not necessarily two. Here is a quick synopsis of the situation; some relevant references are provided in [60]. For integers $k \geq 1$, $N \geq 1$ and characters

$\epsilon\colon (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$, one considers the space $S_k(N, \epsilon)$ of weight-k cusp forms with character ϵ on $\Gamma_0(N)$; we have $S(N) = S_2(N, 1)$. This space is automatically zero unless $\epsilon(-1) = (-1)^k$, so we will assume that this parity condition is satisfied. The space $S_k(N, \epsilon)$ admits an operation of Hecke operators T_n, and we again have the concept of a normalized eigenform in $S_k(N, \epsilon)$. If $f = \sum a_n q^n$ is such a form, the numbers a_n $(n \geq 1)$ and the values of ϵ all lie in a single integer ring $\mathcal{O}_f$. For each non-zero prime ideal λ of $\mathcal{O}_f$, one constructs a semisimple representation

$$\rho_{f,\lambda}\colon\ \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F}_\lambda).$$

Let ℓ again denote the characteristic of $\mathbf{F}_\lambda$. Then for all $p \nmid \ell N$, the trace of $\rho_{f,\lambda}(\mathrm{Frob}_p)$ is again a_p mod λ. The determinant of this matrix is $p^{k-1}\epsilon(p)$ mod λ.

Once the proper definition is made, the determinant of the map $\rho_{f,\lambda}$ becomes the product $\chi_\ell^{k-1}\epsilon$. In writing $\det \rho_{f,\lambda} = \chi_\ell^{k-1}\epsilon$, we use χ_ℓ to denote the mod ℓ cyclotomic character and employ a standard construction to regard ϵ as a map $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{F}_\lambda^*$. The construction in question begins with the map $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to (\mathbf{Z}/N\mathbf{Z})^*$ giving the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the Nth roots of unity. Composing this map with the character ϵ, we obtain a homomorphism $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathcal{O}_f^*$. On reducing this homomorphism mod λ, we obtain the desired variant of ϵ.

Evaluating the formula $\det \rho_{f,\lambda} = \chi_\ell^{k-1}\epsilon$ on $c =$ "complex conjugation", one finds

$$\det(\rho_{f,\lambda}(c)) = (-1)^{k-1}\epsilon(c) = -1.$$

In these equalities, we exploit the fact that $\epsilon(c)$ is another name for $\epsilon(-1)$ and remember the parity condition $\epsilon(-1) = (-1)^k$. The upshot of this is that the representations $\rho_{f,\lambda}$ are always odd, even in the generalized set-up.

In fact, it is possible in this situation to find a normalized eigenform f' of weight two with some character ϵ', along with a maximal ideal λ' of the integer ring for f', so that the representations $\rho_{f,\lambda}$ and $\rho_{f',\lambda'}$ are isomorphic, cf. [67, Th. 2.2 and Cor. 3.2]. (To compare these representations, it is necessary to embed the residue fields of λ and of λ' in a suitably chosen common finite field of characteristic ℓ.) Hence the "generalized set-up" can be reduced to the case of weight two, provided that one considers eigenforms for which the associated characters may be non-trivial. The process of reduction to weight two is a very powerful one in the theory, since the representations arising from forms of this weight are constructed directly from points of finite order on abelian varieties. To the best of my knowledge, the idea of reducing systematically to weight two originated with an unpublished 1968 manuscript of Shimura [78].

Before leaving this topic, we should mention that the cases $k = 1$ and $k > 1$ are quite distinct in flavor. In the former case, the representations $\rho_{f,\lambda}$ may be viewed as the set of reductions of a single continuous representation ρ_f with finite image. For details, see [16].

7. Serre's conjectures

We shall give a brief summary of conjectures made by J.-P. Serre in [75]. A recent article by H. Darmon [15] discusses the conjectures more extensively and emphasizes applications and numerical examples.

Let ℓ be a prime number and let $\mathbf{F}$ be an algebraic closure of the prime field $\mathbf{F}_\ell$. Suppose that $\rho\colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ is an odd continuous representation. It seems natural to ask whether or not ρ is the mod λ representation attached to a suitable normalized eigenform. Since the representations $\rho_{f,\lambda}$ are semisimple by definition, it is necessary to limit our discussion to the case where ρ is semisimple. In fact, the case where ρ is semisimple and reducible is sometimes awkward, so we will assume from now on that ρ is irreducible. However, the excluded case where ρ is reducible is quite interesting; see [8] for some observations in this case.

Let us say then that ρ is *modular* if one can find: (i) a normalized eigenform f in some space $S_k(N, \epsilon)$; (ii) a prime λ dividing ℓ in the ring of integers $\mathcal{O}_f$ associated to f; and (iii) an embedding $\mathbf{F}_\lambda \hookrightarrow \mathbf{F}$ such that ρ is isomorphic to the representation obtained by composing $\rho_{f,\lambda}$ with the inclusion

$$\mathbf{GL}(2, \mathbf{F}_\lambda) \hookrightarrow \mathbf{GL}(2, \mathbf{F})$$

associated with (iii). A weak form of the conjectures made by Serre in [75] is the following statement: *Every irreducible continuous odd representation* $\rho\colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ *is modular.*

This statement was first formulated by Serre in the 1970s for modular forms of level 1 (i.e., on $\mathbf{SL}(2, \mathbf{Z})$). For such forms, the representations $\rho_{f,\lambda}$ are ramified only at ℓ; Serre asked whether an odd irreducible ρ which is unramified outside ℓ is necessarily associated to a normalized eigenform on $\mathbf{SL}(2, \mathbf{Z})$. Tate confirmed this for $\ell = 2$ [87] by showing that there are no such representations. (As Tate remarks at the end of his article, Serre treated the case $\ell = 3$ in a similar manner by exploiting the discriminant bounds of Odlyzko and Poitou.) The case where ρ may be ramified at primes other than ℓ was taken up by Serre in the mid-1980s, when computer calculations by J.-F. Mestre convinced Serre that the conjectured statement was plausible.

The qualitative conjecture to this effect is supplemented in [75] by an intricate recipe which pinpoints the space $S_k(N, \epsilon)$ where one should find an eigenform f giving rise to a specific representation ρ. (As Serre later observed, the recipe for ϵ needs to be modified in certain cases when $\ell = 2$ or 3.) The conjunction of the qualitative statement that ρ is modular and the precise recipe fingering the space $S_k(N, \epsilon)$ is sometimes called the Strong Serre Conjecture. One possible justification for this name is the fact that the "strong" conjecture immediately implies the Taniyama-Shimura conjecture, Fermat's Last Theorem, and a host of other assertions! It has gradually emerged that the qualitative statement and its strong cousin are in fact *equivalent*, at least when $\ell > 2$; see [67] and [17] for a proof of the equivalence. Hence it is now possible to use the singular term "Serre's conjecture" to refer to what was initially a package of interrelated conjectures.

Perhaps I should close this section by expressing the sentiment that the conjecture of Serre, while visibly important, currently seems intractable. Given an irreducible representation ρ with odd determinant, one is at a loss for a strategy which will lead to a proof that ρ is modular. In my Progress in Mathematics lecture, I engaged in a certain amount of philosophical speculation, stressing the parallel between the Taniyama-Shimura conjecture and Serre's conjecture. Each conjecture states that all objects of a certain type are modular; in both cases one has a unidirectional "arrow"—a means of constructing objects from modular forms. Thus one's impulse is to try to attack these conjectures by an appropriate form of "counting." While Serre's conjecture is broader than the Taniyama-Shimura conjecture (the former implies the latter), one might suspect that Galois representations might be easier to count than elliptic curves. To the extent that this is so, one could imagine attacking the geometric conjecture about elliptic curves via the Galois-theoretic conjecture of Serre.

This philosophy is perhaps not far removed in spirit from the strategy used by Andrew Wiles in approaching the Taniyama-Shimura conjecture. Certainly, however, the analogy should not be taken too seriously; in fact, Wiles introduced his approach in 1993 with the statement that it was "orthogonal" to Serre's conjecture.

8. Frey's construction

Like most recent work on Fermat's Last Theorem, the connection between Serre's conjecture and FLT begins with constructions linking solutions to Fermat's equation with elliptic curves. Although Y. Hellegouarch and others had noted such constructions, a decisive step was taken by G. Frey in an unpublished 1985 manuscript entitled "Modular elliptic curves and Fermat's conjecture."

Frey's idea goes as follows. Suppose that there is a non-trivial solution to Fermat's equation $X^\ell + Y^\ell = Z^\ell$. We can assume that the exponent is a prime number different from 2 and 3 and that the solution is given by a triple of relatively prime integers a, b, and c. The equation $y^2 = x(x - a^\ell)(x + b^\ell)$ then defines an elliptic curve A with unexpected properties.

These properties are catalogued in §4.1 of Serre's article [75]: After performing some elementary manipulations, we arrive at a triple (a, b, c) in which b is even and c is congruent to 1 mod 4. Frey's construction yields for this triple an elliptic curve A whose conductor is the product of the prime numbers which divide abc (each occurring to the first power). In particular, A is semistable. On the other hand, the minimal discriminant Δ of A is the quotient of $(abc)^{2\ell}$ by the factor 2^8. From Frey's point of view, the main "unexpected" property of A is that Δ is the product of a power of 2 and a perfect ℓth power, where ℓ is a prime ≥ 5. Frey translated this property into a statement about the Néron model for A: if p is an odd prime at which A has bad reduction, the number of components in the mod p reduction of the Néron model is divisible by ℓ. Frey's idea was to compare this number to the corresponding number for the Jacobian of the modular curve $X_0(N)$, where N is the conductor of A. Frey predicted that a discrepancy between the two numbers would preclude A from being modular. In other words, Frey concluded heuristically that the existence of A was incompatible with the Taniyama-Shimura conjecture, which asserts that all elliptic curves over $\mathbf{Q}$ are modular.

Frey's construction spawned several lines of inquiry, in which mathematicians sought either to prove Fermat's Last Theorem outright, or else to link it to established or emerging conjectures such as the abc conjecture and Szpiro's conjecture. These latter conjectures are treated by such articles as [23], [24], [32], [39], [50] and [56]. From the point of view of Szpiro's conjecture and the abc conjecture, the surprising feature of Frey's curve is the size of its discriminant, rather than any special properties of the discriminant's factorization. For Frey's curve A, the absolute value of the discriminant can be bounded from below by a high power of the conductor of A.

To illustrate the force of Frey's construction, we now sketch the deduction of Fermat's Last Theorem from Serre's conjectures [75]. (A word of caution: these conjectures are still conjectures!) Suppose that a, b and c are relatively prime integers which satisfy Fermat's equation with exponent ℓ. After performing the manipulations mentioned above, we may write down a Frey curve A whose discriminant is the product of a power of 2 and a perfect ℓth power. If ℓ is greater than 3, a theorem of Mazur [46] asserts that the representation $\rho = \rho_{A,\ell}$ defined by $A[\ell]$ is an irreducible representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Because of the hypothesis on the discriminant of A, ρ is unramified outside 2 and ℓ; moreover it is "finite" at ℓ in a sense which is explained in [75]. The recipes of Serre's article require that ρ arises from a normalized eigenform in the space $S(2)$. However, as was mentioned above, this space has dimension 0.

9. Conjecture "epsilon"

A first step toward justifying Frey's heuristic conclusion was taken in August, 1985 by Serre in a letter to J.-F. Mestre [74]. In this letter, the writer formulated two related conjectures about modular forms, which he called C_1 and C_2. (These conjectures predate the conjectures of [75]; they are now special cases of the latter conjectures.) Serre pointed out that Fermat's Last Theorem is a consequence of the Taniyama-Shimura conjecture *together with* the two new conjectures. As it was thought initially that C_1 and C_2 would be easy to establish, the two statements immediately acquired the collective nickname "Conjecture ϵ." One thus had

$$\text{Taniyama-Shimura} + \epsilon \implies \text{Fermat's Last Theorem.}$$

Serre's "ϵ" conjecture is in fact a subtle statement about the mod ℓ Galois representations arising from eigenforms in the various spaces $S(N)$. Specifically, suppose that $f \in S(N)$ is a normalized eigenform, and let λ be a prime ideal in the ring of integers of the field generated by the coefficients of f. Then $\rho_{f,\lambda}$ is a semisimple representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ with values in a finite field, whose characteristic we will call ℓ. This representation is unramified at all primes $p \neq \ell$ which do not divide N; in other words, $\rho_{f,\lambda}$ has the right to be ramified at ℓ and at those primes which divide N. Serre's conjecture concerns the case where $\rho_{f,\lambda}$ is unramified at such a prime: it predicts that this behavior can be attributed to the existence of a normalized eigenform f' of level lower than N which gives rise to $\rho_{f,\lambda}$.

Specifically, suppose that $\rho_{f,\lambda}$ is irreducible, that ℓ is different from 2, and that $\rho_{f,\lambda}$ is unramified at a prime number $p \neq \ell$ which divides N but whose square does not divide N. Serre's conjecture predicts that there is an eigenform $f' \in S(N/p)$, together with a prime ideal λ' in the integer ring of the field of coefficients of f, such that $\rho_{f,\lambda}$ and $\rho_{f,\lambda'}$ are isomorphic. A variant of this conjecture concerns the case $p = \ell$. (Note that, as in a situation discussed earlier, one must embed the residue fields of λ and of λ' in a suitably chosen common finite field of characteristic ℓ before comparing $\rho_{f,\lambda}$ and $\rho_{f,\lambda'}$.)

I proved Serre's level-lowering conjecture in a 1990 article [62], thereby establishing the implication

$$\text{Conjecture of Taniyama-Shimura} \implies \text{Fermat's Last Theorem}$$

which was the goal of Frey's construction. (See also [58] and [63] for expository accounts of this work.) Since the Frey curves associated with Fermat solutions are semistable elliptic curves, I proved that the semistable case of the Taniyama-Shimura conjecture implies Fermat's Last Theorem.

To illustrate the logic used in establishing this implication, we consider a semistable elliptic curve A over $\mathbf{Q}$, together with a prime $\ell \geq 3$ for which the representation $\rho_{A,\ell}$ is irreducible. Suppose that $\rho_{A,\ell}$ is unramified at all primes $p \neq 2, \ell$ and moreover that $\rho_{A,\ell}$ is "finite" at ℓ. Then the results of [62] assert that A cannot be modular. To apply these results to Fermat's Last Theorem, we suppose that A is the Frey curve associated to a hypothetical solution to the degree-ℓ Fermat equation with $\ell > 3$. Then, as was noted above, A is semistable, and $\rho_{A,\ell}$ is an irreducible representation with the indicated ramification and "finiteness" properties. Accordingly, A cannot be modular. Consequently, if all elliptic curves over $\mathbf{Q}$ are modular, then there can be no solution to Fermat's equation.

Further light can be shed on [62] if we focus on the simplest situation in which its results apply. Suppose that A is an elliptic curve over $\mathbf{Q}$ whose conductor is a prime number p. Let ℓ be a prime number different from 2 and p for which $\rho_{A,\ell}$ is irreducible. The representation $\rho_{A,\ell}$ is unramified at all primes other than p and ℓ, and the contribution of [62] is to show that $\rho_{A,\ell}$ is indeed ramified at p if A is modular. To prove this, the one supposes that A is modular, so that $\rho_{A,\ell}$ is connected up with the space of weight-two cusp forms on $\Gamma_0(p)$. It is possible to show that $\rho_{A,\ell}$ is similarly connected with other discrete subgroups of $\mathbf{SL}(2,\mathbf{R})$, coming from indefinite quaternion division algebras over $\mathbf{Q}$. More precisely, there are prime numbers $q \neq p$ such that $\rho_{A,\ell}$ arises from the quaternion algebra over $\mathbf{Q}$ of discriminant pq. The desired result about $\rho_{A,\ell}$ follows from a detailed comparison of the mod p and mod q reductions of the three modular curves $X_0(p)$, $X_0(q)$ and $X_0(pq)$ with the mod p and mod q reductions of the Shimura curve associated with the quaternion algebra of discriminant pq. The latter curve is an analogue of $X_0(pq)$ in which the group $\Gamma_0(pq)$ is replaced by the group of norm-1 elements in a maximal order of the quaternion algebra of discriminant pq.

10. Wiles's strategy

Suppose that A is an elliptic curve over $\mathbf{Q}$. To verify the Taniyama-Shimura conjecture for A is to link A to modular forms. In an approach inspired by Serre's conjecture, one might begin by considering the representations $\rho_{A,\ell}$ obtained from the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $A[\ell]$, when ℓ is a prime number. If one could show that an infinite number of these representations are modular (in the broadest possible sense), one would go on to prove that A is modular. Alas, as was indicated above, it is not clear how to translate this approach into a proof.

We mentioned previously that Wiles's approach to the Taniyama-Shimura conjecture is "orthogonal" to one based on consideration of the varying $\rho_{A,\ell}$. Here is the nub of the idea: One first fixes a prime ℓ and considers the family of groups $A[\ell^\nu]$ for $\nu = 1, 2, \ldots$. The resulting sequence of representations

$$\rho_{A,\ell^\nu} \colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{Z}/\ell^\nu\mathbf{Z})$$

may be packaged as a single representation

$$\rho_{A,\ell^\infty} \colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{Z}_\ell),$$

where $\mathbf{Z}_\ell$ is the ring of ℓ-adic integers, i.e., the projective limit of the rings $\mathbf{Z}/\ell^\nu\mathbf{Z}$. To prove that A is a modular elliptic curve, it suffices to show that ρ_{A,ℓ^∞} is modular in an appropriate sense. Indeed, the trace of $\rho_{A,\ell^\infty}(\mathrm{Frob}_p)$ coincides with the rational integer b_p for all $p \nmid \ell N$, where N is the conductor of A. As soon as one finds an eigenform f in $S(N)$ whose eigenvalues are related to the traces of $\rho_{A,\ell^\infty}(\mathrm{Frob}_p)$, one has essentially proved that A is modular.

Needless to say, if ρ_{A,ℓ^∞} is modular, then so, in particular, is $\rho_{A,\ell}$. Relating ρ_{A,ℓ^∞} to modular forms is thus no easier than the formidable task of proving that $\rho_{A,\ell}$ is modular! On the other hand, to prove that A is modular by the ℓ-adic method, we need only work with a *single* prime ℓ. The approach of [92] capitalizes on the fact that the finite groups $\mathbf{GL}(2,\mathbf{F}_2)$ and $\mathbf{GL}(2,\mathbf{F}_3)$ are solvable. This circumstance enables one to apply deep results of Langlands [40] and Tunnell [89] to prove that $\rho_{A,\ell}$ is modular for $\ell \leq 3$, cf. [68, §2.3]. (The relevant results of Langlands are those concerning the theory of base change à la Saito-Shintani. For expositions of these results, see [25] and [26].)

Wiles's basic idea is to prove that that if ℓ is a prime for which $\rho_{A,\ell}$ is modular, then ρ_{A,ℓ^∞} is automatically modular (and hence A is a modular elliptic curve). In thinking about the jump from $\rho_{A,\ell}$ to ρ_{A,ℓ^∞}, ignores A as much as possible—the aim is to prove results about ℓ-adic representations which can be applied to ρ_{A,ℓ^∞}.

11. The language of deformations

We now introduce the machinery which underlies the jump from the modularity of $\rho_{A,\ell}$ to the modularity of ρ_{A,ℓ^∞}. We suppose for simplicity that A is a semistable elliptic curve over $\mathbf{Q}$, and we let $\ell \geq 3$ be a prime number for which $\rho_{A,\ell}$ is both modular and irreducible. Choosing a basis of $A[\ell]$, we regard $\rho_{A,\ell}$ as taking values in the matrix group $\mathbf{GL}(2,\mathbf{F}_\ell)$. As was suggested above, we seek to establish the modularity of ρ_{A,ℓ^∞} by a method which treats simultaneously *all* lifts of $\rho_{A,\ell}$ which are plausibly modular.

In this context, lifts are continuous homomorphisms

$$\tilde{\rho}\colon\ \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2,B),$$

where B is a complete local Noetherian $\mathbf{Z}_\ell$-algebra with residue field $\mathbf{F}_\ell$. They are constrained to lift $\rho_{A,\ell}$ in the obvious sense. Namely, we require that $\rho_{A,\ell}$ coincide with the composite of $\tilde{\rho}$ and the homomorphism $\mathbf{GL}(2,B) \to \mathbf{GL}(2,\mathbf{F}_\ell)$ induced by the residue map $B \to \mathbf{F}_\ell$. (his depiction ignores a technical wrinkle: it might to necessary later on to replace $\mathbf{Z}_\ell$ by the integer ring of a finite extension of the ℓ-adic field $\mathbf{Q}_\ell$.) The lifts which are "plausibly modular" are those which obey a set of local properties. The word "local" is meant to allude to the restrictions of $\tilde{\rho}$ to the subgroups $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ obtained as decomposition groups for prime numbers p (or, more precisely, for primes of $\overline{\mathbf{Q}}$). Wiles imposes conditions on these restrictions which lift conditions already satisfied by $\rho_{A,\ell}$. These conditions are summarized in §3.4 of [68]; we shall evoke them below as well. There is flexibility and tension implicit in the choice of these conditions. They should be broad enough to be satisfied by ρ_{A,ℓ^∞} and tight enough to be satisfied only by lifts that can be related to modular forms. Roughly speaking, in order to prove the modularity of all lifts satisfying a fixed set of conditions, you need to specify in advance a space of modular forms S so that the normalized eigenforms in S satisfy the conditions and such that, conversely, all lifts satisfying the conditions are plausibly related to forms in S. It is intuitively clear that this program will be simplest to carry out when the conditions are the most stringent and progressively harder to carry out as the conditions are relaxed.

A theme which emerges rapidly is that there are at least two sets of conditions of special interest. Firstly, one is especially at ease when dealing with the most stringent possible set of conditions which are satisfied by $\rho_{A,\ell}$; this leads to what Wiles calls the "minimal" problem. Secondly, one needs at some point to consider some set of conditions which allows treatment of the lift ρ_{A,ℓ^∞}—this lift is, after all, our main target. It would be natural to consider the most stringent such set. The two sets of conditions may coincide, but there is no guarantee that they do; in general, the second set of conditions is more generous than the first.

In [92], Wiles provides a beautiful "induction" argument which enables him to pass from the minimal set of conditions to a non-minimal set. Heuristically, this argument requires keeping tabs on the set of those normalized eigenforms whose Galois representations are compatible with an incrementally relaxing set of conditions. As the conditions loosen, the set of forms must grow to keep pace with the increasing number of lifts. The increase in the number of lifts can be estimated from above by a local cohomological calculation. A sufficient supply of modular forms is then furnished by the theory of congruences between normalized eigenforms of differing level. This latter theory may be viewed as a vast generalization of what went on in the author's article [61].

At the risk of distorting slightly the theory, I will index the shifting set of local conditions by a finite set of prime numbers Σ which contains the set Σ_0 of primes at which $\rho_{A,\ell}$ is ramified. This latter set may be obtained concretely as the set of primes which divide the discriminant of the number field obtaining by adjoining to $\mathbf{Q}$ the coordinates of all points in $A[\ell]$. It is not hard to see that Σ_0 contains ℓ and is contained in the union of $\{\ell\}$ and the set of primes at which A has bad reduction. Indeed, this union is the set of primes Σ^* at which ρ_{A,ℓ^∞} is ramified, according to the well-known criterion of Néron-Ogg-Shafarevich [77]. In this perspective, the minimal set of conditions corresponds to the choice $\Sigma = \Sigma_0$, while a set of conditions broad enough to include ρ_{A,ℓ^∞} is obtained by choosing $\Sigma = \Sigma^*$.

As promised, we shall now give the flavor of the conditions that one imposes on the lifts $\tilde{\rho}$ of type Σ. Firstly, we demand that $\tilde{\rho}$ be unramified outside Σ. Secondly, we ask that $\tilde{\rho}$ have the same qualitative behavior at each prime $p \in \Sigma$ as the representation $\rho_{A,\ell}$. (The imposition of this condition can be traced back to Gouvêa's article [27].) Since ℓ lies in Σ_0, a condition is imposed on $\tilde{\rho}$ locally at the prime ℓ—this condition requires that $\tilde{\rho}$ be "ordinary" if A has ordinary or multiplicative reduction at ℓ and that $\tilde{\rho}$ be "flat" if A has supersingular reduction at ℓ, cf. [68]. For convenience, a final (global) condition may be imposed on $\tilde{\rho}$, to the effect that the determinant of $\tilde{\rho}$ be the composite of the ℓ-adic cyclotomic character

$$\tilde{\chi}_\ell \colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{Z}_\ell^*$$

and the structural map $\mathbf{Z}_\ell^* \to A^*$. This supplementary condition has the effect of allowing one to work with the spaces $S(N)$ rather than with spaces of modular forms on groups of the form $\Gamma_1(N)$.

To prove that all lifts $\tilde{\rho}$ of type Σ are modular, one first passes to equivalence classes with respect to the relation in which two representations with values in $\mathbf{GL}(2, B)$ are equivalent whenever they are conjugate by an element of $\mathbf{GL}(2, B)$ which maps to the identity matrix in $\mathbf{GL}(2, \mathbf{F}_\ell)$. The equivalence classes of lifts are called *deformations*—the terminology is borrowed from algebraic geometry, where deformation theory has been developed extensively. The idea of introducing deformation theory into the subject of Galois representations is due to Mazur [47]. As F. Gouvêa reports in his recent survey [29], the deformation viewpoint has evolved considerably since [47] first appeared. In particular, a 1993 article by R. Ramakrishna [59] introduces foundational tools which Wiles requires in [92]; Ramakrishna shows that deformation theory can be applied to study the family of lifts which satisfy the local conditions to which we have been alluding.

More precisely, the work of Mazur and Ramakrishna proves that there is a *universal* deformation of type Σ. This is a lift

$$\rho_\Sigma\colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, R_\Sigma)$$

which is characterized by the property that for each lift

$$\tilde{\rho}\colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, B)$$

of type Σ there is a unique homomorphism $\omega\colon R_\Sigma \to B$ of local $\mathbf{Z}_\ell$-algebras so that the deformation defined by $\tilde{\rho}$ agrees with the one obtained from ρ_Σ and ω. A common initial impression is that R_Σ is a relatively mysterious object whose existence stems from an abstract representability theorem—one comes to grips with it only gradually.

Wiles seeks to compare R_Σ with a concrete ring $\mathbb{T}_\Sigma$, which he defines directly as a completion of a classical ring of Hecke operators. A theorem of Carayol constructs a Galois representation

$$\rho'_\Sigma\colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbb{T}_\Sigma)$$

with the property that the trace of $\rho'_\Sigma(\mathrm{Frob}_p)$ is the Hecke operator T_p, for all but finitely many primes p. One thinks of ρ'_Σ as the universal *modular* deformation of type Σ. The problem is then to prove that ρ_Σ and ρ'_Σ coincide, so that all deformations of type Σ are modular. To come to grips with this problem, Wiles begins with the canonical homomorphism

$$\varphi_\Sigma\colon R_\Sigma \to \mathbb{T}_\Sigma$$

which results from the universality of ρ_Σ. It is relatively easy to show that φ_Σ is surjective; the coincidence of ρ_Σ and ρ'_Σ means that φ_Σ is an *isomorphism*, cf. [68, §4.2].

My reaction to Wiles's 1993 announcement was astonishment that one could prove the modularity of Galois representations by working directly with φ_Σ. Despite conjectures by Mazur [51, p. 85] and Gouvêa [27, p. 108] to the effect that φ_Σ is an isomorphism in the ordinary case, and Theorem 2 of [20], I was not prepared for the revelation that φ_Σ could be studied fruitfully.

12. Gorenstein and complete intersection conditions

The conjecture that φ_Σ is an isomorphism is proved in [88, 92]. The proof relies on standard notions of commutative algebra which are discussed in [43] and in [41]. The argument of Wiles and Taylor-Wiles proceeds from the definition of $\mathbb{T}_\Sigma$ as a completion of the ring generated by the Hecke operators T_n acting on a specific space of classical cusp forms. In particular, $\mathbb{T}_\Sigma$ is free of finite rank over $\mathbf{Z}_\ell$. It has been known for some time that Hecke rings such as $\mathbb{T}_\Sigma$ look special from the vantage point of commutative ring theory. For example, $\mathbb{T}_\Sigma$ tends to be Gorenstein, which means that the $\mathbb{T}_\Sigma$-module $\operatorname{Hom}(\mathbb{T}_\Sigma, \mathbf{Z}_\ell)$ is free of rank 1 over $\mathbb{T}_\Sigma$. (The Gorenstein property was first noted in a special case by Mazur [45, Ch. II, §15], and then established in ever-widening generality by others, including the author.) Chapter 2 of [92] includes a section which summarizes and improves on the known Gorenstein assertions; it proves, in particular, that $\mathbb{T}_\Sigma$ is Gorenstein.

Using commutative algebra techniques, [92] presents a number of conditions each of which is sufficient to show that φ_Σ is an isomorphism. As stated in the Introduction above, Wiles became aware while studying the problem that φ_Σ is an isomorphism whenever the ring $\mathbb{T}_\Sigma$ is a complete intersection ring. (This implication is proved from another point of view in [49].) Later, judging that it would be difficult to prove directly that $\mathbb{T}_\Sigma$ is a complete intersection ring, Wiles focused on a numerical inequality [68, Th. 5.2] and showed that φ_Σ is an isomorphism whenever it is satisfied.

At the time of his 1993 Cambridge lectures, Wiles believed that he had proved the numerical inequality through the construction of a "geometric Euler system," thereby generalizing work of M. Flach [20]. Later analysis showed that the construction envisaged by Wiles was flawed. Interestingly, it is not yet clear whether it can be modified so as to yield an Euler system with the desired properties. At the minimum, one feels that this avenue of inquiry is likely to remain extremely active. In particular, Mazur's course notes [49] extract new information from Flach's original construction.

At the time of this revision, the arguments of [88] and [92] are being studied and internalized by the mathematical community. Followup work is already beginning to appear. For example, we mentioned above that the main theorem of [92] has been strengthened by a recent manuscript of F. Diamond [18]. Also, the arguments given in [88] and [92] have been shortened somewhat by simplifications due to G. Faltings; these are explained in an appendix to [88].

13. Toward the Taniyama-Shimura conjecture

We conclude with a short survey of results which have been obtained by Wiles [92], Taylor-Wiles [88], and Diamond [18]. Two theorems which have the flavor

$$\rho_{A,\ell} \text{ modular } \implies \rho_{A,\ell^\infty} \text{ modular}$$

are presented in [92]. In each of the theorems, the prime ℓ is taken to be odd, and the representation $\rho_{A,\ell}$ is required to be irreducible.

One of the theorems (namely, [92, Th. 4.8]) has no application to Fermat's Last Theorem, since it does not apply to semistable elliptic curves. This theorem does apply to in situations where $\rho_{A,\ell}$ is isomorphic to the representation obtained from the ℓ-division points of a complex multiplication elliptic curve A' over $\mathbf{Q}$. In these cases, A need not have complex multiplication itself; it is merely "linked mod ℓ" to a CM curve. For a discussion of this theorem, including an explicit determination of the family of curves which can be linked to a fixed curve A', see [69].

The theorem in [92] which applies to Fermat's Last Theorem is the one whose proof depends on the new work of Taylor-Wiles [88]:

Theorem 1. *Suppose that A is a semistable elliptic curve over $\mathbf{Q}$. Let ℓ be an odd prime. Assume that the representation $\rho_{A,\ell}$ is both irreducible and modular. Then A is a modular elliptic curve.*

Fermat's Last Theorem may then be proved by combining the author's theorem [62] with the following result, which may be viewed as a highly non-obvious corollary of Theorem 1.

Theorem 2. *Let A be a semistable elliptic curve over $\mathbf{Q}$. Then A is a modular elliptic curve.*

Wiles deduces Theorem 2 from Theorem 1 by an ingenious argument, which we will now describe. (The argument, which is presented in [92, Ch. 5], has been sketched in [68].) Let A be a semistable elliptic curve, and consider the representation $\rho_{A,3}$. If this representation happens to be irreducible, then it is also modular, by the the results of Langlands and Tunnell which were cited above. Thus Theorem 1, proves that A is modular.

What happens if $\rho_{A,3}$ is reducible? In this case, we examine $\rho_{A,5}$. If this latter representation is reducible as well, then Wiles shows directly that A is modular. Hence we can, and do, suppose that $\rho_{A,5}$ is irreducible. Wiles shows then that one can find a second semistable elliptic curve A' whose mod 5 representation is isomorphic to that of A and whose mod 3 representation is irreducible (cf. [68, Appendix B]). Two applications of Theorem 1 then suffice to show that A is modular. Indeed, applying the theorem to A' with $\ell = 3$, we find that A' is modular. In particular, the irreducible representation $\rho_{A',5}$ is modular. Since this representation coincides with $\rho_{A,5}$, we may apply the theorem to A with $\ell = 5$ to conclude that A is modular, as desired.

A preprint of F. Diamond [18] generalizes Theorem 1 to the case where A is an elliptic curve over $\mathbf{Q}$ whose conductor is not divisible by ℓ^2 (i.e, one which is semistable at ℓ):

Theorem 3. *Let ℓ be an odd prime number. Suppose that A is an elliptic curve over $\mathbf{Q}$ which is semistable at ℓ. Assume that the representation $\rho_{A,\ell}$ is both irreducible and modular. Further, if $\ell = 3$, assume that $\rho_{A,3}\left(\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}\left(\sqrt{-3}\right))\right)$ is non-abelian. Then A is a modular elliptic curve.*

The condition concerning $\mathrm{Gal}\left(\overline{\mathbf{Q}}/\mathbf{Q}\left(\sqrt{-3}\right)\right)$ occurs already in Wiles's work. However, in the situation of Theorem 1, Wiles proves that $\rho_{A,3}\left(\mathrm{Gal}\left(\overline{\mathbf{Q}}/\mathbf{Q}\left(\sqrt{-3}\right)\right)\right)$ is irreducible and non-abelian when $\ell = 3$; in other words, the assumption relative to $\mathrm{Gal}\left(\overline{\mathbf{Q}}/\mathbf{Q}\left(\sqrt{-3}\right)\right)$ has been omitted from Theorem 1 because it may be proved unconditionally.

Using a variant of the Wiles argument we have just sketched, Diamond deduces the following generalization of Theorem 2.

Theorem 4. *Suppose that A is an elliptic curve over $\mathbf{Q}$ which is semistable both at 3 and at 5. Then A is modular.*

References

1. A. Ash and R. Gross, *From Gauss to Langlands: a context for Wiles's achievement*, in preparation.
2. A. O. L. Atkin and J. Lehner, *Hecke operators on $\Gamma_0(m)$*, Math. Annalen **185** (1970), 134–160.
3. C. Batut, D. Bernardi, H. Cohen and M. Olivier, *GP/PARI*, Available by anonymous ftp from `megrez.math.u-bordeaux.fr` or `math.ucla.edu`, in the directory `/pub/pari`.
4. B. J. Birch and W. Kuyk, eds., *Modular functions of one variable IV*, Lecture Notes in Math., vol. 476, Springer-Verlag, Berlin and New York, 1975.
5. W. Bosma and H. W. Lenstra, Jr., *Complete systems of two addition laws for elliptic curves* (to appear).
6. N. Boston, *A Taylor-made plug for Wiles' proof*, College Math. J. (to appear).
7. N. Boston and A. Granville, *Review of* [90], American Math. Monthly (to appear).
8. K. M. Buzzard, *The levels of modular representations*, Cambridge University thesis, 1995.
9. H. Carayol, *Sur les représentations ℓ-adiques associées aux formes modulaires de Hilbert*, Ann. scient. Éc. Norm. Sup., 4ᵉ série **19** (1986), 409–468.
10. J. H. Coates and S. T. Yau, eds., *Elliptic curves and modular forms*, proceedings of a conference held in Hong Kong, December 18–21, 1993, International Press, Cambridge, MA and Hong Kong (to appear).
11. I. Connell, *Apecs (arithmetic of plane elliptic curves)—a program written in Maple*, Available by anonymous ftp from `math.mcgill.ca` in the directory `/pub/apecs`.
12. G. Cornell and J. Silverman, eds., *Arithmetic Geometry*, Springer-Verlag, Berlin and New York, 1986.
13. D. Cox, *Introduction to Fermat's Last Theorem*, American Math. Monthly **101** (1994), 3–14.
14. H. Darmon, *The Shimura-Taniyama conjecture (d'après Wiles)*, Russian Math Surveys (to appear).
15. ———, *Serre's conjectures*, in [55].
16. P. Deligne and J.-P. Serre, *Formes modulaires de poids* 1, Ann. scient. Éc. Norm. Sup., 4ᵉ série **7** (1974), 507–530.
17. F. Diamond, *The refined conjecture of Serre*, in [10].
18. ———, *On deformation rings and Hecke rings* (to appear).
19. G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), 349–366.
20. M. Flach, *A finiteness theorem for the symmetric square of an elliptic curve*, Invent. Math. **109** (1992), 307–327.
21. J.-M. Fontaine, *Il n'y a pas de variété abélienne sur* $\mathbf{Z}$, Invent. Math. **81** (1985), 515–538.
22. J.-M. Fontaine and B. Mazur, *Geometric Galois representations*, in [10].
23. G. Frey, *Links between stable elliptic curves and certain diophantine equations*, Annales Universitatis Saraviensis **1** (1986), 1–40.
24. ———, *Links between elliptic curves and solutions of $A - B = C$*, Journal of the Indian Math. Soc. **51** (1987), 117-145.
25. S. Gelbart, *Automorphic forms and Artin's conjecture*, Lecture Notes in Math **627** (1977), 241–276.
26. P. Gérardin and J. P. Labesse, *The solution to a base change problem for* $\mathbf{GL}(2)$ *(following Langlands, Saito, Shintani)*, Proceedings of Symposia in Pure Mathematics **33 (2)** (1979), 115–133.
27. F. Q. Gouvêa, *Deforming Galois representations: controlling the conductor*, Journal of Number Theory **34** (1990), 95–113.
28. ———, *"A marvelous proof"*, American Math. Monthly **101** (1994), 203–222.
29. ———, *Deforming Galois representations: a survey*, in [55].

30. B. Hayes and K. A. Ribet, *Fermat's Last Theorem and modern arithmetic*, American Scientist **82** (1994), 144–156.
31. W. R. Hearst III and K. A. Ribet, *Review of "Rational points on elliptic curves" by Joseph H. Silverman and John T. Tate*, Bulletin of the AMS **30** (1994), 248–252.
32. M. Hindry, *"a, b, c", conducteur, discriminant*, Publications mathématiques de l'Université Pierre et Marie Curie, Problèmes diophantiens (1986–87).
33. A. Jackson, *Update on proof of Fermat's Last Theorem*, Notices of the AMS **41** (1994), 185–186.
34. ———, *Another step toward Fermat*, Notices of the AMS **42** (1995), 48.
35. A. W. Knapp, *Elliptic curves*, Math. Notes, vol. 40, Princeton Univ. Press, Princeton, NJ, 1992.
36. S. Lang, *Introduction to modular forms*, Springer-Verlag, Berlin and New York, 1976.
37. ———, *Abelian varieties*, Springer-Verlag, Berlin and New York, 1983.
38. ———, *The Taniyama-Shimura file*, Available directly from S. Lang, Yale Math. Department.
39. ———, *Old and new conjectured diophantine inequalities*, Bull. AMS **23** (1990), 37–75.
40. R. P. Langlands, *Base change for* $\mathbf{GL}(2)$, Annals of Math. Studies, vol. 96, Princeton University Press, Princeton, 1980.
41. H. W. Lenstra, Jr., *Complete intersections and Gorenstein rings*, in [10].
42. W.-C. W. Li, *Newforms and functional equations*, Math. Annalen **212** (1975), 285–315.
43. H. Matsumura, *Commutative ring theory*, Cambridge University Press, Cambridge (U.K.), 1986.
44. P. A. van Mulbregt and J. H. Silverman, *Elliptic curve calculator*, Available by anonymous ftp from `gauss.math.brown.edu` in the directory `/dist/EllipticCurve`.
45. B. Mazur, *Modular curves and the Eisenstein ideal*, Publ. Math. IHES **47** (1977), 33–186.
46. ———, *Rational isogenies of prime degree*, Invent. Math. **44** (1978), 129–162.
47. ———, *Deforming Galois representations*, Galois groups over $\mathbf{Q}$, MSRI Publications, vol. 16, Springer-Verlag, Berlin and New York, 1989, pp. 385–437.
48. ———, *Number theory as gadfly*, Am. Math. Monthly **98** (1991), 593–610.
49. ———, *Very rough course notes for Math 257y, parts I–III*, to appear as "Galois deformations and Hecke curves".
50. ———, *Questions about number*, New directions in mathematics (to appear).
51. B. Mazur and J. Tilouine, *Représentations galoisiennes, différentielles de Kähler et ⟪conjectures principales⟫*, Publ. Math. IHES **71** (1990), 9–103.
52. T. Miyake, *On automorphic forms on* $\mathbf{GL}_2$ *and Hecke operators*, Annals of Math. **94** (1971), 174–189.
53. ———, *Modular forms*, Springer-Verlag, Berlin and New York, 1989.
54. V. K. Murty, *Introduction to Abelian varieties*, CRM monograph series, vol. 3, American Mathematical Society, Providence, 1993.
55. ———, ed., *Elliptic curves, galois representations and modular forms*, CMS Conference Proceedings, American Mathematical Society, Providence (to appear).
56. J. Oesterlé, *Nouvelles approches du "théorème" de Fermat*, Astérisque **161/162** (1988), 165–186.
57. A. Ogg, *Elliptic curves and wild ramification*, American J. of Math. **89** (1967), 1–21.
58. D. Prasad, *Ribet's Theorem: Shimura-Taniyama-Weil implies Fermat*, in [55].
59. R. Ramakrishna, *On a variation of Mazur's deformation functor*, Compositio Math. **87** (1993), 269–286.
60. K. A. Ribet, *The* ℓ*-adic representations attached to an eigenform with Nebentypus: a survey*, Lecture Notes in Math **601** (1977), 17–52.
61. ———, *Congruence relations between modular forms*, Proc. Int. Cong. of Mathematicians 1983, pp. 503–514.
62. ———, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. Math. **100** (1990), 431–476.
63. ———, *From the Taniyama-Shimura conjecture to Fermat's Last Theorem*, Annales de la Faculté des Sciences de l'Université de Toulouse **11** (1990), 116–139.
64. ———, *Abelian varieties over* $\mathbf{Q}$ *and modular forms*, 1992 Proceedings of KAIST Mathematics Workshop, Korea Advanced Institute of Science and Technology, Taejon, 1992, pp. 53–79.

65. ———, *Wiles proves Taniyama's conjecture; Fermat's Last Theorem follows*, Notices of the AMS **40** (1993), 575–576.
66. ———, *Modular elliptic curves and Fermat's last theorem: a lecture presented at George Washington University, Washington, DC, August 1993*, Selected Lectures in Mathematics, American Mathematical Society, Providence, 1993.
67. ———, *Report on mod ℓ representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, Proceedings of Symposia in Pure Mathematics **55 (2)** (1994), 639–676.
68. K. Rubin and A. Silverberg, *A report on Wiles' Cambridge Lectures*, Bulletin of the AMS **31** (1994), 15–38.
69. ———, *Families of elliptic curves with constant mod p representations*, in [10].
70. J-P. Serre, *Une interprétation des congruences relatives à la fonction τ de Ramanujan*, Séminaire Delange-Pisot-Poitou (1967–68, n° 14).
71. ———, *Facteurs locaux des fonctions zêta des variétés algébriques (définitions et conjectures)*, Séminaire Delange-Pisot-Poitou (1969–70, n° 19).
72. ———, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), 259–331.
73. ———, *A course in arithmetic*, Graduate Texts in Math., vol. 7, Springer-Verlag, New York, Heidelberg and Berlin, 1973.
74. ———, *Lettre à J.-F. Mestre (13 août 1985)*, Contemporary Mathematics **67** (1987), 263–268.
75. ———, *Sur les représentations modulaires de degré 2 de* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, Duke Math. J. **54** (1987), 179–230.
76. ———, *Algebraic groups and class fields*, Graduate Texts in Math., vol. 117, Springer-Verlag, New York, Heidelberg and Berlin, 1988.
77. J-P. Serre and J. Tate, *Good reduction of abelian varieties*, Ann. of Math. **88** (1968), 492–517.
78. G. Shimura, *An ℓ-adic method in the theory of automorphic forms.*
79. ———, *A reciprocity law in non-solvable extensions*, Journal für die reine und angewandte Mathematik **221** (1966), 209–220.
80. ———, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, 1971.
81. ———, *On elliptic curves with complex multiplication as factors of the Jacobians of modular function fields*, Nagoya Math. J. **43** (1971), 199–208.
82. ———, *Class fields over real quadratic fields and Hecke operators*, Ann. of Math. **95** (1972), 131–190.
83. ———, *On the factors of the jacobian variety of a modular function field*, J. Math. Soc. Japan **25** (1973), 523–544.
84. J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Math., vol. 106, Springer-Verlag, Berlin and New York, 1986.
85. ———, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Math., vol. 151, Springer-Verlag, Berlin and New York, 1994.
86. J. T. Tate, *Algorithm for determining the type of a singular fiber in an elliptic pencil*, Lecture Notes in Math **476** (1975), 33–52.
87. ———, *The non-existence of certain Galois extension of* $\mathbf{Q}$ *unramified outside 2*, Contemporary Mathematics **174** (1994), 153–156.
88. R. L. Taylor and A. Wiles, *Ring theoretic properties of certain Hecke algebras*, Annals of Math. (to appear).
89. J. Tunnell, *Artin's conjecture for representations of octahedral type*, Bull. AMS (new series) **5** (1981), 173–175.
90. M. vos Savant, *The world's most famous math problem: the proof of Fermat's last theorem and other mathematical mysteries*, St. Martin's Press, New York, 1993.
91. A. Weil, *Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Math. Annalen **168** (1967), 165–172.
92. A. Wiles, *Modular elliptic curves and Fermat's Last Theorem*, Annals of Math. (to appear).

UC Mathematics Department, Berkeley, CA 94720-3840 USA
E-mail address: ribet@math.berkeley.edu