Where Stands Functoriality Today?[1]

Robert P. Langlands

1. Introduction. The notion of functoriality arose from the spectral analysis of automorphic forms but its definition was informed by two major theories: the theory of class fields as created by Hilbert, Takagi, and Artin and others; and the representation theory of semisimple Lie groups in the form given to it by Harish-Chandra. In both theories the statements are deep and general and the proofs difficult, highly structured, and incisive. Historical antecedents and contemporary influences aside, both were in large part created by the power of one or two mathematicians. Whether for intrinsic reasons or because of the impotence of the mathematicians who have attempted to solve its problems, the fate of functoriality has been different, and the theory of automorphic forms remains in 1997 as it was in 1967: a diffuse, disordered subject driven as much by the availability of techniques as by any high esthetic purpose.

Preoccupied with other matters I have drifted away from the field, so that I certainly have no remedy to offer. None the less, having had two occasions[2] to address the question posed in the title, I have tried to understand something of the techniques that have led to progress on the questions central to functoriality, their successes and their limitations, as well as the new circumstances in the theory of automorphic forms: problems and notions that once seemed peripheral to me and whose importance I failed to appreciate are now central, either because of their intrinsic importance or because of their accessibility.

Partly as an encouragement to younger, fresher mathematicians to take up the problem of functoriality, for that is one of the purposes of this school, but also partly as an idle reflection as to what I myself might undertake if I returned to it, I would like to respond to the title in broad terms, personal and certainly diffident and uncertain. My own mathematical experience and observation strongly suggest that progress is almost always the result of sustained awareness of the principal issues supplemented by some specific, concrete insight: begged, borrowed, or stolen or, happiest of all, distilled in one's own alembic. I offer no insights.

Initially there were two principal issues: functoriality itself, the relation between automorphic forms on different groups; and the identification of *motivic* L-functions, thus those associated to algebraic varieties over number fields, of which the zeta function, Artin L-functions, and the Hasse-Weil L-functions are the primitive examples, with *automorphic* L-functions, of which the zeta function and Hecke L-functions – of all types – are the first examples. The first issue arose and could be broached in the context of nonabelian harmonic analysis: representation theory and the trace formula. The second arose elsewhere but could also be broached in this context. Three key names[3] here are: Arthur for the trace formula; Kottwitz for the study of the Hasse-Weil zeta-functions of Shimura varieties; and Waldspurger for the solution (in part) of some important attendant problems, transfer and the fundamental lemma, in local harmonic analyis. The methods developed so far are difficult and important and probably essential for the construction of the complete theory but they have limitations, and I see no reason to believe that they alone will suffice.

In parallel to the notions and problems of functoriality arose a theory of special values of motivic L-functions, thus a collection of problems, and methods for dealing with some of them. Whereas progress in functoriality, as a part of the theory of automorphic forms, has been largely analytic, exploiting the more abstract consequences of abelian class-field theory but developing very few *arithmetic* arguments, some of the most incisive progress in the theory of special values has been purely arithmetic. Relations between the analytic theory and the arithmetical have often been uneasy. In Wiles's proof of the Fermat theorem they are fused: although the burden of the proof is borne by the arithmetic, an essential initial element is provided by the analysis. This is a clue that it would be reckless to ignore.

---

[1] First appeared in Representation Theory and Automorphic forms, ed. T. N. Bailey and A. W. Knapp, Proc. Symp. Pure Math., vol. 61 (1997).

[2] In Edinburgh and at a conference to celebrate the 250th anniversary of Princeton University organized by its department of mathematics.

[3] There are of course many more, but my purpose is to provide a brief guide to the subject, not to the literature. Names are mentioned to provide a little color and as implicit suggestions for further reading, nothing more.

The notion of functoriality arose as a strategy – modeled on that of Artin for dealing with abelian Artin L-functions – for establishing the analytic continuation of all automorphic L-functions by showing that each was equal to a special kind of autormorphic L-function attached to $GL(n)$, a $standard$ L-function. Some of these automorphic L-functions can be handled by other methods. Since functoriality is not available in general, one is obliged to resort to them, either to establish critical cases of functoriality[4] or to deal with finer properties of the automorphic L-functions: location of poles; properties of special values. These methods, in which the principal ingredients are theta functions and the Rankin-Selberg integral, are exploited in an unstructured, catch-as-catch-can way; it would be useful, especially in view of the increasing importance of automorphic L-functions and automorphic forms in the analytic theory of numbers,[5] to have a coherent notion of their function.

The analytic theory of automorphic forms can be developed over function fields as well as over number fields, but here, thanks largely to Drinfeld, some constructive, arithmetic elements were introduced very early. It is instructive to compare the proofs of the very little hard evidence we have that a nonabelian class-field theory exists, namely base change for $GL(2)$ or, more generally, $GL(n)$ with respect to cyclic extensions, with the proofs of class-field theory. In class-field theory the principal proofs are preceded by the construction and analysis of basic arithmetic objects, Kummer extensions. For base change in general, the little arithmetic information needed is simply taken from the abelian theory; the proofs themselves are entirely analytic. It is unlikely that a general nonabelian theory will be constructed so cheaply. It appears that over function fields the arithmetic – or diophantine – information is already implicitly at hand in the reductive group, it being possible to approximate the set $G(F)\backslash G(\mathbb{A})$ by points on an algebraic variety. His reflections on function fields also led Drinfeld to purely geometric formulations – in which the constant field of the function field is replaced by the field of complex numbers – of functoriality or rather of the duality in terms of which it is expressed. What influence the geometric study will have on the arithmetic problems is far from clear; on the other hand the very existence of the geometric notions and problems is further evidence that the dual group, in terms of which functoriality is defined, is a natural and not a factitious construct.

2. Basic analytic theory. In order to make the notions as accessible as possible, I shall begin by recalling in somewhat metaphorical terms the analytic theory[6] of automorphic forms on $GL(n)$ and then pass to the simplest kind of automorphic L-functions and the notion of functoriality. Although the theory of automorphic forms applies to all reductive groups and draws many insights and many techniques from this generality, for many, maybe all, serious applications to the theory of numbers, this generality is superfluous; only $GL(n)$ matters. On the other hand, it is unlikely that the theory on $GL(n)$, $n > 2$, would ever have been broached, had it not been suggested by experience with the symplectic and other groups.

The analytic theory is a spectral theory, for $families$ of commuting operators, some of them differential operators and some of them difference operators (Hecke operators). At the analytic level the difficulties connected with the difference operators are of less importance, so that they are suppressed at this stage. [7]

Basic objects I

[4] For example, the existence, established by Tunnell, of an automorphic form attached to the most general octahedral representation.

[5] Since I shall have occasion to discuss none of this, I mention two recent, and quite different, reviews, one by Duke in the Notices of the AMS of February, 1997, and one by Bump, Friedberg, and Hoffstein in the April, 1996 issue of the Bulletin of the AMS.

[6] There are several expositions of this theory available; I leave the choice among them to the reader.

[7] In few contexts except the related one of the representation theory of real reductive groups has there been, I had thought, a successful, nontrivial theory of families of commuting differential operators. It is likely the rigidity of the group structure that permits the construction of the theory for automorphic forms. It has, however, recently been pointed out to me by Siddhartha Sahi that the spectral theory of real reductive groups now appears to be only an instance of more general theories, to which he suggests Macdonald's report in the Séminaire Bourbaki (exp.797, 1995) as an introduction.

The first is the group $GL(n)$, but over two rings: the field $\mathbb{Q}$ of rational numbers and the ring $\mathbb{A}$ of adeles, which will not be defined here. The notion of an adele is just a formal expression of the importance of congruences for number theory. The basic spaces are $GL(n,\mathbb{Q})\backslash GL(n,\mathbb{A})/K$, in which K is the product of a compact subgroup of $GL(n,\mathbb{A})$ with a central subgroup of the same group. What are these spaces in reality? Examples:

1) $n = 1$ —– multiplicative group of $\mathbb{Z}/n\mathbb{Z}$;
2) $n = 2$ —– the upper half-plane divided by the subgroup of matrices in $SL(2,\mathbb{Z})$ that are congruent to I modulo an integer N.

Notice that both these spaces carry the structure of algebraic varieties: the first is of dimension zero; the second is of dimension one – a *modular* curve of level N. This ceases to be so for $n > 2$.

Basic objects II

These are, first of all, the space of square-integrable functions on the space

$$GL(n,\mathbb{Q})\backslash GL(n,\mathbb{A})$$

or what, for our purposes, amounts to the same thing, on the family of spaces

$$GL(n,\mathbb{Q})\backslash GL(n,\mathbb{A})/K,$$

and its decomposition under the action of $GL(n,\mathbb{A})$. This is the spectral theory. The basic ingredients of the spectral theory are the cusp forms on $GL(m,\mathbb{A})$, $m = 1, 2 \ldots$. They, or perhaps better, if due attention is paid to the symmetries of the theory, the irreducible subspaces of cusp forms, are to be thought of as the elementary particles. I observe right off the bat that for each m there are infinitely many and that, although some are related to other objects, there is no question whatsoever of classifying them.

The spectrum on $GL(n,\mathbb{A})$ is obtained by choosing $m_1,\ldots,m_r$ such that $\sum m_i = n$ and putting particles $\pi_1,\ldots,\pi_r$, one for each m_i, together, moving with various velocities (in one-dimension!). Observe that particles of different types can sometimes fuse to form *bound* particles. This is rare and the possibilities are dealt with by Arthur's conjecture, which is very precise and proved in only a few instances. It is closely related to Ramanujan's conjecture.

The basic objects π, thus collections of elementary particles, fused or not, and with relative motion, also have a *local* structure – at each prime number and over the real numbers – that mimics the *global* adelic structure. For all but a finite number of primes these local objects have no internal structure; hence a global elementary particle undergoes complete fission at almost all primes and becomes just n objects with no internal structure (thus by definition, globally or locally, attached to the constant function on $GL(1)$) moving about with different velocities $s_1, s_2, \ldots, s_n$. (To push the metaphor to unwarranted extremes, they obey Bose-Einstein statistics; so the order does not matter!) What matters is in fact the conjugacy class of the matrix

$$A_p(\pi) = \begin{pmatrix} p^{is_1} & 0 & \ldots & 0 & 0 \\ 0 & p^{is_2} & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & p^{is_{n-1}} & 0 \\ 0 & 0 & \ldots & 0 & p^{is_n}. \end{pmatrix}$$

Attach an incomplete L-function to π by

$$\text{(A)} \qquad L(s,\pi) = \prod{}' \frac{1}{\det(I - A_p(\pi)/p^s)}.$$

The product can be completed at the primes where local objects have an internal structure, thus in number-theoretical language where there is ramification, and also at the infinite prime. Classical methods, developed by many people but always resting ultimately on the imbedding of $GL(n)$ in the additive group of $n \times n$ matrices and the use of Fourier analysis, allow one to extend these functions (completed or not) to all of the complex plane as meromorphic functions with a very limited number of poles. For reasons to be explained, I refer to these L-functions as the *standard* L-functions. (This terminology has unfortunately been corrupted.)

The particle metaphor, in which what is otherwise referred to as an *automorphic representation* or *automorphic form* is viewed as a collection of particles with structure moving about at various speeds, fits well the assemblage of an automorphic representation ρ on $GL(l)$ and another σ on $Gl(m)$ to construct an automorphic representation $\pi = \rho \oplus \sigma$ on $GL(n)$, $n = l + m$. This is an operation that is well understood in the context of the spectral theory. It is the theory of Eisenstein series. There is another operation that is not well understood, although its existence is conjectured as a part of functoriality. If we have an automorphic repesentation ρ of $GL(l)$ and one σ of $GL(m)$ and if this second is just a collection of m structureless particles moving about at velocities $s_1, \dots, s_m$ then we can construct the automorphic representation $\rho \otimes \sigma$ of $GL(n)$, $n = lm$, as the sum of $\rho(s_1), \dots, \rho(s_m)$. The representation $\rho(s)$ is just ρ with a modified velocity and is constructed by multiplying ρ by the one-dimensional representation $|\det|^{is}$. There is reason to believe that for all ρ and all σ an automorphic representation that deserves, because of its local properties, to be denoted $\rho \otimes \sigma$ can be constructed.

3. Functoriality. Functoriality is a succinct hypothesis that is easy to state once the basic notions are at hand and whose consequences are immediate and serious: the analytic continuation of Artin's L-functions and the general form of Ramanujan's conjecture.

As remarked it is believed that it is possible to construct $\rho \otimes \sigma$ even if σ has internal structure. This would be a basic form of functoriality. What is the general form and how does one arrive at it? Notice, before turning to functoriality, that the basic categories in which it is possible both to add objects of degrees l and m to arrive at one of degree $l + m$ and to multiply them to arrive at one of degree lm are the categories defined by the finite-dimensional representations of a given group (finite, compact, or algebraic). An automorphic representation is, however, almost always an infinite-dimensional representation of a very large group, and the pertinent degree has nothing to do with its dimension.

So functoriality refers to objects and categories that have not yet appeared in this review. We can study automorphic representations π on any (reductive) algebraic group G defined over $\mathbb{Q}$ (even over number fields and function fields). The principal groups are $GL(n)$. To any G we associate a complex group ${}^LG$, its L-group, by a procedure that is now fairly well known, and that I do not describe in any further detail here, except in the context of examples. For a given π, in general an automorphic repesentation of $G(\mathbb{A})$, one can associate to almost all primes p a conjugacy class $A_p(\pi)$ in ${}^LG$. If ρ is a *finite-dimensional holomorphic representation* of the L-group ${}^LG$, we can introduce the, perhaps incomplete, L-function

$$L(s, \pi, \rho) = \prod_p{}' \frac{1}{\det(I - \rho(A_p(\pi)/p^s))}. \tag{B}$$

The study of Eisenstein series yields a large number of Euler products with meromorphic continuation that can be written in this form. Since the L-group of $GL(n)$ is $GL(n, \mathbb{C})$ or at least has $GL(n, \mathbb{C})$ as a quotient, according to the context in which it is appropriate to work, the products (A) are a special case of (B)

Once the Euler products are so expressed, it is very strongly suggested that they can be continued not just for those ρ that arise from Eisenstein series, but for all ρ. Class-field theory, in the form given to it by Emil Artin, immediately suggests a strategy. Given π, an automorphic representation of the group $G(\mathbb{A})$, and ρ, an n-dimensional holomorphic representation of the complex group ${}^LG$, show that there is an automorphic representation Π of $GL(n, \mathbb{A})$ such that

$$\{A_p(\Pi)\} = \{\rho(A_p(\pi))\}$$

for almost all p. Then

$$L(s, \pi, \rho) = L(s, \Pi),$$

and, as already observed, the analytic continuation of the right-hand side is essentially a matter of Poisson summation.

This is the first form of functoriality that suggested itself. Applying it to the trivial group $\{1\}$ over a finite algebraic extension of $\mathbb{Q}$ viewed as a group over $\mathbb{Q}$ by restriction of scalars, so that the (somewhat protean)

L-group becomes the Galois group $\mathrm{Gal}(K/\mathbb{Q})$ of a finite extension, I immediately realized that it would yield the nonabelian class-field theory for which Artin had searched unsuccessfully.

An extension of the first notion of functoriality that suggests itself quickly is that if G and G' are two groups and

$$\phi : {}^LG \to {}^LG',$$

then there is a corresponding map of automorphic forms $\pi \to \pi'$ such that $A_p(\pi')$ and $\phi(A_p(\pi))$ are conjugate for all almost all p. It is this non-formal, deep-lying functoriality to which the title refers.

Some progress, even substantial progress, has been made on functoriality since the idea was introduced in the late sixties. None the less we still do not understand it in any serious sense. Since functoriality, once established, entails immediately the analytic continuation of all the functions $L(s, \pi, \rho)$, and all L-functions with Euler products appearing in the theory of automorphic forms are – as yet, but probably for good – of this type, there would seem to be little point in pursuing seriously the various methods introduced for dealing with this or that special case:

(1) Fourier coefficients of Eisenstein series: Langlands, Shahidi;
(2) Hecke's first method;
(3) integration against Eisenstein series: Rankin, Selberg,. . . ;
(4) integration against a product of theta functions and Eisenstein series:
Shimura,. . .

On the other hand, for the moment these methods sometimes yield very important information that is not otherwise available, so that it would be premature to discard them. I am not sure what their ultimate role will be.

4. The trace formula. Functoriality itself can, for the moment, only be established in a very limited number of cases. One technique is to use the trace formula, thus the spectral theory, some local information, and some form – to use once again a colorful language – of the uncertainty principle. The local information – endoscopy, transfer, and the fundamental lemma – has been hard won, by Waldspurger and many others, and is still incomplete. I myself thought about these matters for ten years without making any serious headway. The global arguments, in part an elaborate induction, are being carried out largely by Arthur and demand the massive deployment of technically difficult analysis in a highly structured conceptual context.

In a number of important cases (infinitely many but still of a very special nature) a homomorphism

$$\phi : {}^LG \to {}^LG'$$

is accompanied by a natural map from conjugacy classes in a semi-direct $G' \rtimes \Sigma$, where Σ is a finite cyclic group to conjugacy classes in G. Note the transition from the L-groups to the groups themselves. Then the map $\pi \to \phi(\pi)$ predicted by functoriality is reflected in character identities, and thus in the trace formula.

These character identities usually manifest themselves in relatively simple, easily understood or well understood, contexts – in particular: finite-dimensional representations of G and $G' \rtimes \Sigma$; representations of real groups; and unramified representations. New identities would be greatly appreciated. Once the basic identity is recognized, the difficulty is, first, to establish sufficiently many of the identities in local harmonic analysis suggested by it, and then to apply the trace formula to the two groups $G(\mathbb{A})$ and $G'(\mathbb{A}) \rtimes \Sigma$ simultaneously in order to compare their spectra. Note that the trace formula compares *geometric* information, thus integration over conjugacy classes in $G(\mathbb{Q})$ or $G'(\mathbb{Q}) \rtimes \Sigma$, on one side with spectral information on the other. The uncertainty principle implies that if most terms on one side vanish then all terms on both sides vanish. Various components of the spectrum appear in different ways, and it was, I believe, Arthur's reflections on the identities entailed between the terms corresponding to bound states that led him to the conjecture to which I alluded earlier. I cannot give it here. It is very delicate, and very important, and is established in some cases, but I am not certain that anyone, even Arthur himself, has fully understood its implications. Some investigators have overlooked it to their later embarrassment.

A basic ingredient in Arthur's conjecture is the *generalized* Ramanujan conjecture that affirms that if π is a cusp form for $GL(m)$ then the conjugacy classes $A_p(\pi)$ have eigenvalues of absolute value one. The generalized Ramanujan conjecture is easily shown to be a consequence of functoriality, and partial results for functoriality lead to partial results for the conjecture.

Arthur himself has focussed on two cases: either the group $G' \rtimes \Sigma$ is defined by a group G' given as $GL(n)$ over a cyclic extension of the ground field; or Σ is of order two and its nontrivial element acts as the involution $A \to {}^tA^{-1}$ on $GL(n)$ over the ground field. The first, cyclic base change, is of importance in connection with Artin's conjecture; the second gives the transfer of automorphic forms associated to the standard homomorphisms of the L-groups of orthogonal and symplectic groups into $GL(n)$. There are other possibilities, [8] still largely neglected. Moreover, the relative trace formula of Jacquet perhaps deserves the attention of an ambitious, talented youth. At the same time it is manifest that although it is unlikely that the final goals can be reached without the trace formula and the methods developed by Arthur, they are inaccessible with identities between traces alone. It is natural to suppose that inequalities between traces, combined perhaps with a deeper group-theoretical analysis (even at the level of finite groups), would be the appropriate tool, but no-one has been able to do anything with this.

The two cases of functoriality not yet established that lie nearest at hand are base change for icosahedral extensions and the existence of automorphic forms on $GL(n)$, $n > 4$, attached to automorphic forms on $GL(2)$ via the homomorphisms of dual groups $GL(2, \mathbb{C}) \to GL(n, \mathbb{C})$ defined by representations on symmetric tensors. These are two basic outstanding problems of the subject. A solution of the second would yield all forms of Ramanujan's conjecture for $GL(2)$. The Ramanujan-Petersson conjecture for holomorphic forms has of course been solved through the last of the Weil conjectures, whose proof was seriously influenced by ideas from the theory of automorphic forms, but the analogous problem for Maaß forms as well as the Selberg conjecture remain open.

I add that there are two other methods to establish functoriality in certain cases: converse theorems and, through Howe's conjecture, the oscillator representation. Converse theorems, although at the present stage occasionally extremely useful, sometimes indispensable, are, if class-field theory is taken as our paradigm, philosophically perverse, putting the cart before the horse. The oscillator representation, thus theta series, yields automorphic representations that are, in the sense of Arthur's conjecture, quite degenerate. Although the connection between theta series and functoriality is quite delicate, and therefore quite fascinating, I am not convinced that it is basic. None the less theta series, and therefore automorphic forms of half-integral weight, remain for me a troubling philosophical puzzle. They are obviously important; they are not unrelated to functoriality; yet the notions of functoriality cannot accommodate them.

Philosophically perverse or not, converse theorems have been very useful and there are eminent mathematicians who attach a great deal of importance to them; so it is amusing to see how to express them in the context of our metaphor. As observed, the constant function on $GL(1, \mathbb{A})$ is to be thought of as the particle without structure. The scattering matrix for the interaction of this structureless particle with an elementary particle π on $GL(n, \mathbb{A})$ is a quotient of standard L-functions

$$\Lambda(s, \pi)/\Lambda(s + 1, \widetilde{\pi}).$$

Here the letter Λ is used rather than L to indicate that the function is a product over all primes, including those at infinity, so that $\xi(s, \pi)$ is the product of $L(s, \pi)$ with some Gamma functions and some elementary functions. Thus the standard L-functions for π describe the interaction of π with a structureless particle.

To describe the interaction of an elementary particle π on $GL(n, A)$ with another elementary particle π' on $GL(m, A)$ one needs the L-functions associated to the tensor-product representation ρ of the L-group $GL(n, \mathbb{C}) \times GL(m, \mathbb{C})$ of $GL(n, \mathbb{A}) \times GL(m, \mathbb{A})$ in $GL(mn, \mathbb{C})$. Usually one writes

C) $$L(s, \pi \times \pi', \rho) = L(s, \pi \times \pi').$$

The usual converse theorems assert that if for each m, $1 \leq m < n$, there is a family of functions associated to all π' on $GL(m, \mathbb{A})$ with the analytic properties of the collection (C) then these families are defined by an elementary particle π. Expressed with our metaphor, the assertion is that if elementary particles in degree less than n are scattered by an unknown object that seems to be an elementary particle of degree n then it is indeed such an elementary particle.

[8] Rogawski's study of the unitary group in three variables is an example.

An obvious question is whether the Riemann hypothesis is satisfied for the standard L-functions $L(s,\pi)$, and thus for all L-functions. It is generally supposed that this is so, although so far as I know the general hypothesis has not been examined numerically. The intimate connection, already observed, between the L-functions $L(s,\pi)$, in particular the zeta function itself which corresponds to the trivial representation of $GL(1,\mathbb{A})$, and the Eisenstein series appearing in the spectral theory has suggested, first, I believe, to Selberg, that the spectral theory might be used to establish the Riemann hypothesis. Selberg crushed any simple hopes by constructing, in the context of $GL(2)$ (or $SL(2)$), nonarithmetic groups for which the analogue of the hypothesis is false. But the zeta function appears in the spectral analysis of all $GL(n)$, $n = 1,\ldots$, and these are very rigid objects. Moreover, Ribet, in a paper to which we shall come, and others have successfully applied information about automorphic forms on $GL(n)$, $n > m$, to the arithmetic study of automorphic functions attached to $GL(m)$, and it is not utterly far-fetched to imagine that something similar could be done analytically. (For different implementations of such a strategy, see the review of Bump, Friedberg, and Hoffstein.) I myself have no idea, but I do remark that if π is a cusp form on $GL(m)$ then $L(s,\pi)$ appears in the spectral analysis of all $GL(mn)$, $n = 1,2,\ldots$, so that there is at least a certain coherency in the suggestion.

5. Hasse-Weil L-functions. As already observed, the principle of functoriality is more or less the suggestion that the collection of automorphic representations behaves like the collection of (finite-dimensional) representations of a group. For automorphic representations themselves, the putative group is large and elusive, so that it is best not to think much about it. The corresponding local objects can, however, be given concrete meaning. For example, representations of a real reductive group $G(\mathbb{R})$ can be classified by homomorphisms into the L-group ${}^LG$ of an extension

$$\mathbb{C}^\times \to W \to \{1,\sigma\}.$$

The element σ is to be thought of as complex conjugation so that $\sigma z\sigma^{-1} = \bar{z}$, $z \in \mathbb{C}^\times$. In addition, the appropriate choice for σ^2, an element of $\mathbb{C}^\times$, is -1.

Then, for example, the representations of $GL(n,\mathbb{R})$ are classified, in an appropriate sense, by the continuous homomorphisms of W into the L-group of $GL(n)$ which is nothing but $GL(n,\mathbb{C})$. Only homomorphisms such that the image of $\mathbb{C}^\times$ is diagonalizable are allowed. The diagonal elements are then characters of $\mathbb{C}^\times$ and thus of the form

$$z \to z^r(z\bar{z})^s,$$

where r is an integer and s a complex number. A very important class of homomorphisms, and thus of representations of $GL(n,\mathbb{R})$, is determined by demanding that in these diagonal characters the complex number s always be an integer, so that the character is of the form

$$z \to z^p\bar{z}^q.$$

The presence of the two integers p and q suggests a connection with Hodge theory; so I say that representations of $GL(n,\mathbb{R})$ of this kind are of *Hodge* type.

We can then distinguish a very important class of automorphic representations π, those whose component at infinity is of Hodge type. (I have not stressed it, but an automorphic representation π of any group $G(\mathbb{A})$ can be expressed as a tensor product $\bigotimes \pi_v$, where v runs over the places, finite and infinite of $\mathbb{Q}$, thus $v = \infty, 2, 3, 5, \ldots$.) We recall, in addition, that using the Weil zeta-functions attached to varieties over finite fields it is possible to attach to a projective variety V over $\mathbb{Q}$ (more generally over finite extensons of $\mathbb{Q}$ and function fields), a global zeta function, the Hasse-Weil zeta-function and that this zeta function can be expressed as a quotient of products of Euler products

$$\text{(D)} \qquad \frac{\prod_{i=2j+1,0\le j<\dim(V)} L^{(i)}(s,V)}{\prod_{i=2j,0\le j\le\dim(V)} L^{(i)}(s,V)}.$$

It is generally supposed that each factor $L^{(i)}(V,s)$ that appears here, either in the numerator or in the denominator, is equal to a standard automorphic L-function $L(s,\pi)$ for a suitable $\pi = \pi(i,V)$. This π will necessarily be of Hodge type at infinity.

The group-theoretical procedure used by Artin to factor the zeta function of a finite extension of a number field into a product of Artin L-functions can be extended to a factorization of the functions $L^{(i)}(s,V)$ although the properties of this factorization are far from understood. There are very few varieties for which the analytic continuation of (D) can be established, and then only by proving first that each factor is equal to an $L(s,\pi)$. For varieties of dimension zero, the function $L^{(0)}(s,V)$ can be dealt with, but its factors, the Artin L-functions, can be analytically – as opposed to meromorphically – continued only in few cases, again by showing that they are equal to an $L(s,\pi)$. The classical abelian reciprocity laws and the Taniyama conjecture are both examples of this technique.

6. New and different problems. There is, in addition, a converse hypothesis: every standard automorphic L-function $L(s,\pi)$ defined by a representation π such that π_∞ is of Hodge type appears as a "factor" of a Hasse-Weil L-function.[9] This is not the time to attempt to give a precise meaning to the term "factor". For these factors, arising as they do from geometry and arithmetic, there is an entirely new complex of questions, all concerning their values at integers, especially negative integers, that transcend the simpler problem of analytic continuation and that are not related to the Riemann hypothesis.

There are, following Kato, three "phases" to the problem, to which the names of many, many people are attached: Tate, Birch, Swinnerton-Dyer, Shimura, Deligne, Beilinson, Iwasawa, Bloch, Kato, Fontaine, Perrin-Riou,. . . .

(1) When divided by appropriate factors – periods of integrals on the associated varieties as well as "regulators" – the values $L(n,\pi)$, $n\in\mathbb{Z}$, are algebraic. At integers n where the function has zeros or poles the order is predicted, and the appropriate power of $z-n$ is removed before calculating the *nonzero* value of the function thus obtained.

(2) These algebraic values lie in well-defined number fields and satisfy congruences that permit the introduction of p-adic L-functions.

(3) The values themselves, when normalized to be algebraic, or the values of the p-adic L-functions, are themselves very important, and are equal to numbers defined by diophantine data, thus implicitly by the solutions of diophantine equations.

For the zeta function itself at even positive or odd negative integers phase (1) is classical and elementary. For example $\zeta(2)=\pi^2/6$ is an equality easily proved. Phase (2) is not so difficult, but phase (3) is very deep.

We are now faced with a number of questions, almost all embarrassing. First of all, to what extent can we establish that the factors of Hasse-Weil zeta-functions are indeed equal to standard automorphic L-functions. There is the classical case of the zero-dimensional varieties defined by abelian extensions of number fields, in which the methods are deep and, within their express limitations, general. There is also the Eichler-Shimura theory for quotients of the upper half-plane by subgroups of the modular group, but this is to some extent simply part of the theory of complex multiplication and thus of class fields. The theory for Shimura varieties, to which I shall return and which is a generalization of the theory of Eichler-Shimura, is also, in spite of the very many difficult and deep theorems that have been established by Kottwitz and many others, still largely a part of the theory of complex multiplication. Thus, as with the work of Arthur, still in progress, the work of Kottwitz and his collaborators may leave us, even when all clearly formulated questions asked at present are answered, with one more. Where do we go from here?

The issue is even more complex. Apparently we can only effect the analytic continuation of the factors of the Hasse-Weil zeta-function if we first exhibit them as automorphic L-functions, and it certainly only makes sense to inquire about particular properties of their values at negative integers after they have been analytically continued. On the other hand, analytic continuation is effected uniformly for automorphic L-functions $L(s,\pi)$, without inquiring whether π_∞ is of Hodge type. Thus in order to make sense of the values, we have to pass to a set of functions for which the values no longer are expected to have any special properties. This appears to me somewhat paradoxical!

On closer examination, matters are seen to be even more turbid. A Shimura variety S is at first a complex algebraic variety that can be represented as

$$G(\mathbb{Q})\backslash G(\mathbb{A})/K, \tag{E}$$

[9] A third basic outstanding problem is to establish this for Maaß forms of Hodge type on $GL(2,\mathbb{A})$.

a set already introduced. The group G here is certainly not arbitrary, but there are several possibilities. It can for example be the symplectic group in any number of variables or a unitary group. The theory of complex multiplication then permits the definition of this complex variety as a variety over $\mathbb{Q}$ (or over a specific well-defined number field). The first goal of the theory is to express the Hasse-Weil zeta-function of what is now diophantine object as a product not of standard L-functions but of automorphic L-functions $L(s,\pi,\rho)$ associated to G itself or to one of its endoscopic groups H. (Then ${}^LH \subset {}^LG$ and ρ is a representation of ${}^LH$.) Of course functoriality predicts that $L(s,\pi,\rho)$ will be a standard L-function but it may be possible to achieve the first goal without proving this and even without being certain that $L(s,\pi,\rho)$ can be analytically continued.

What sometimes happens is that by one or the other of the special techniques listed (Hecke-Rankin-Selberg-Shimura) the functions $L(s,\pi,\rho)$ can be analytically continued, and then special values appear as integrals over a subset of (E) and can be interpreted as periods. It appears that accidental, conditional phenomena are being used to establish general principles, a philosophically disagreeable circumstance that may or may not be intrinsic to class-field theory and all its hypothetical generalizations. Even though at present these special techniques yield very little of what is foreseen, it is not entirely out of the question that they will remain an essential part of the final argument. It is useful to keep an open mind.

7. Shimura varieties. The original purpose of expressing the Hasse-Weil zeta-functions of Shimura varieties as a product of automorphic L-functions was formed shortly after the introduction of these L-functions and for no other reason than to establish their utility and even necessity.

The techniques used to express the functions $L^{(i)}(s,S)$ as a product of functions $L(s,\pi,\rho)$ are similar to those used by Arthur, but there are additional ingredients. The product of functions $L^{(i)}(s,S)$ or rather its logarithm is expressed by spectral data. The trace formula converts the sum of spectral data into a sum of geometric data. On the other hand the Hasse-Weil zeta-function, or rather its logarithm, can also be expressed by geometric data, the number of points on the variety S over finite fields. Thus the primary problem is to relate the set of points on S with coordinates in a given finite field to conjugacy classes in $G(\mathbb{Q})$. This, it turns out, can be done, by no means easily, with the help of the theory of complex multiplication. There are two additional difficulties, not yet entirely overcome: combinatorial problems and the singularities of the completions of the variety S. By a stroke of luck, the combinatorial problems are the same as those of the fundamental lemma, and in so far as that lemma has been established they are solved. The difficulties arising from the singularities form a chapter, perhaps several chapters, in the study of intersection cohomology, and are being treated as such.

There is, so far as I know, no reason to believe that all motivic L-functions appear as factors of L-functions associated to Shimura varieties and the natural vector bundles on them, so that Shimura's reformulation (but I may misunderstand the somewhat obscure history) of the Taniyama conjecture pertains to an anomaly. Generalized in the most naive way (as no-one has ever suggested was appropriate!) the reformulation, sometimes known as the modular-curve conjecture, would ask that motivic L-functions appear as factors of the L-functions attached to Shimura varieties, whereas Taniyama's original conjecture generalized would only ask that they be identified as standard L-functions.

If the focus is on the L-functions that are attached to the Shimura varieties themselves, then the surprise, but not the disappointment, is that by and large the automorphic L-functions with which the motivic L-functions are identified are not those whose analytic continuation can be effected by the provisional methods at our disposition. For example, for the projective symplectic group $PSp(2n)$ that defines the Siegel varieties, which are the best known of the higher-dimensional Shimura varieties, the L-group is the spinor group attached to the orthogonal group in $2n+1$ dimensions and the principal factors of the zeta function are of the form $L(s,\pi,\sigma)$, where σ is the spinor representation. Except for $n=1$ and $n=2$, these are scarcely accessible at present.

8. Constructive elements. In the most striking case so far in which it has been established that a factor of a Hasse-Weil zeta-function is a standard L-function, the proof by Wiles of the suggestion of Taniyama as modified by Shimura, the two aspects of $GL(2)$ – carrier of the standard L-functions and carrier of Shimura varieties – are entwined in very curious ways that I do not understand. In some sense the argument begins with an object given by the analytic theory, thus one obtained by examining $GL(2)$ in its first guise where no deformation is possible. Then it quickly abandons $GL(2)$ in this attire, and passes to a modular curve, thus to the Shimura variety attached to $GL(2)$, and begins a p-adic deformation argument.

These arguments reflect, I suppose, Wiles's experience with the conjectures of phase (3). For the layman the significance of those in phase (1) is sometimes easier to grasp. For an elliptic curve E over $\mathbb{Q}$ defined by an equation like $y^2 = ax^3 + bx^2 + cx + d$, Taniyama's suggestion, as I interpret it, is that $L^{(1)}(s, E)$ is a standard automorphic L-function for $GL(2)$. Shimura, again in my interpretation, goes further and states that there is a surjective map of a modular curve C of a certain level N onto E, and Weil allows us to predict the level. Apparently (although I have never found the time to carry out such computations) assured of all this – by the conjectures as theorems – we can with time and effort calculate the map $C \to E$ and thus the π for which $L^{(1)}(s, E) = L(s, \pi)$. The representation π in hand we can also calculate the value or the order of vanishing of $L(s, \pi)$ for any s. Then the conjectures of phase (1) at the integer $s = 1$, in the special case of an elliptic curve a part of the Birch-Swinnerton-Dyer conjecture, predict the rank of the group of $\mathbb{Q}$-rational points on E. What happens, however, so far as I can see is that, when this conjecture is proved, it is proved by exhibiting concretely points, the Heegner points, from which $\mathbb{Q}$-rational points can be explicitly constructed. Thus the proofs seem to give more information, even more appealing information, than the conjectures themselves.

Certainly phase (1) is, for me at least, fraught with more easily comprehended consequences than phases (2) and (3). Examined attentively, however, it has extremely problematic aspects, and I cannot resist posing a question to which some readers may have an answer. The Beilinson conjecture, a part of phase (1), includes the Tate conjecture explicitly. The Tate conjecture is certainly closely related to the Hodge conjecture and to its general forms. Does it imply the Hodge conjectures; or do the Hodge conjectures contain an analytic element? Can it be proved independently of the Hodge conjecture, or does phase (1) also contain an implicit analytic element? It is perhaps well to remind ourselves, when passing recklessly, gay of heart and light of foot, over phase (1) and on to phases (2) and (3) that we are abandoning questions whose very nature we have as yet failed to understand.

Even for a point over $\mathbb{Q}$, thus for the zeta function, phases (2) and (3) are extremely difficult. Their origins lie, I believe, in divisibility properties of class numbers of cyclotomic fields, thus in the work of Kummer, the founder of the theory of cyclotomic fields and therefore, in some sense, also the founder of class-field theory. If I understand correctly, phase (3) is for a point essentially the main conjecture of Iwasawa theory and was only solved a little more than ten years ago. It appears to be no accident that more than one of the principal contributors to its solution also contributed to the resolution of Fermat's theorem.

To deal with phase (2) or with phase (3) even for a point, it seems absolutely necessary to appeal to the theory of automorphic forms on $GL(2)$ – more precisely to the study of modular curves – thus, so to speak, to move from $GL(0)$ to $GL(2)$, and in particular (recall the remarks on the Riemann hypothesis) to exploit the theory of Eisenstein series not over $\mathbb{C}$ but over p-adic fields. One reason is that some of the arithmetic objects, whose existence is predicted by the conjectures of phase (3), have to be constructed explicitly, and this can only be done with the aid of the explicitly constructed diophantine objects already at hand, the modular curves.

For example, the value of the zeta function at $1 - 2m$, $m = 1, 2, \ldots$, is expressed in terms of a Bernoulli number as $(-1)^m B_m/2m$. Let p be a prime number and μ_p a p-th root of unity. Let $\mathfrak{A}$ be the ideal class group of $\mathbb{Q}(\mu_p)$ and consider $\mathfrak{C} = \mathfrak{A}/\mathfrak{A}^p$. The Galois group $\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q})$ acts on $\mathfrak{C}$ and $\mathfrak{C}$ decomposes as

$$\mathfrak{C} = \sum_{i \pmod{p-1}} \mathfrak{C}_i,$$

where

$$\mathfrak{C}_i = \{\mathfrak{c} \in \mathfrak{C} | \mathfrak{c}^\sigma = \mathfrak{c}^{\chi(\sigma)^i} \,\forall\, \sigma \in \mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q})\},$$

and

$$\mu_p^\sigma = \mu_p^{\chi(\sigma)}.$$

Suppose k is even and $2 \le k \le p - 3$. As what I suppose is one of the simplest expressions of phase (3), p divides $\zeta(1 - k)$ if and only if the group $\mathfrak{C}_{1-k}$ is not trivial. Since this group is determined by the existence of abelian extensions of $\mathbb{Q}(\mu_p)$ with various properties, the implication

$$\mathfrak{C}_{1-k} \neq \{1\} \text{ implies } p \mid B_k,$$

proved by Herbrand entirely within the context of class-field theory and Kummer extensions, is a constraint on the existence of algebraic numbers with prescribed properties. On the other hand, the converse implication requires the construction of abelian extensions of $\mathbb{Q}(\mu_p)$ with prescribed properties, and this construction is not carried out with Kummer extensions directly. Rather[1]0 the construction utilizes the arithmetic theory of automorphic forms for $GL(2)$, exploiting among other things the appearance of zeta functions and Dirichlet L-functions as Fourier coefficients of Eisenstein series. Thus it is not so far in spirit from the techniques used by Shahidi for the study of automorphic L-functions. Difficult theorems for the smaller group are proved as consequences of easier theorems for the larger.

9. Conclusion. I have certainly not been able to digest the material pertinent to the questions broached in this review, and may never succeed in doing so. The question as to where we go from here, in so far as it applies to Hasse-Weil zeta-functions, and thus to Shimura varieties, especially in view of their connections with the problems evoked in Kato's three phases, is certainly not one I would venture to answer. The possibilities intimated by the recent intermingling of the analytic and the algebraic techniques are manifold but still, to me at least, obscure.

For functoriality too, it is best to be prudent. Once the distinction between the automorphic representations whose component at infinity is Hodge and the general class of all automorphic representations is made, there is one point to be observed. There is less to be established for the general class, and it will probably need to be established analytically and directly because the deformation provided by, say, a p-adic theory is unlikely to be available. On the other hand, the Artin L-functions, thus the factors of the Hasse-Weil zeta-functions of zero-dimensional varieties, are difficult to distinguish from elements of this general class. To repeat myself in different words: the Hasse-Weil zeta-functions of zero-dimensional varieties will probably have to be treated directly, without the help of the auxiliary algebraic techniques such as deformation that may be available for the zeta functions of positive dimension.

In some sense the immediate future of functoriality is less bright than that of Shimura varieties. Beyond the work of Arthur and its possible extensions there is a clearly visible horizon that we have no idea how to cross. When discouraged by this reflection I recall, among other things, that the notion of endoscopy, which has nourished representation theory and harmonic analysis for almost two decades, arose not as an internal development of the analytic theory but from the study of the zeta functions of Shimura varieties.

There is certainly no lack of problems, large and small, any one of which may offer a clue.

[1]0 The relevant paper by Ribet appears in Inv. math., vol. 34, 1976.