The present paper, for which its author was awarded a Fields medal, had achieved, even before publication, considerable fame and the proceedings of the International Mathematical Congress in Hyderabad will contain two accounts of it, one by the author itself and one by James Arthur, a laudation delivered at the presentation of the prize. Both accounts are extremely instructive, and I refer the reader to them, as well as to two instructive presentations of the fundamental lemma on ArXiv, one by Thomas Hales, one by David Nadler

There is a great deal to be said about the fundamental lemma, about its origins, about the methods used to prove it and the developments that preceded the proof itself, and about its consequences or possible consequences, much more than could be accommodated in a normal review. No-one is yet familiar with all this material. As a consequence, a good deal has been written about the lemma that, in my view, is misleading. I am convinced that anyone who wants to contribute to the central problems in the contemporary theory of automorphic representations, or, better, to *functoriality* and matters related to it, will need a better grasp of all these matters than any one person possesses at present. I shall try here to clarify this assertion, although this will entail a risk, not only of false prophecy but also of revealing my own ignorance. I understand the origins of the lemma; I believe I have as much insight into its possible consequences as anyone; but the proof itself, which exploits difficult tools and concepts from both modern algebraic geometry and topology, contains a very great deal of which I have only an uncertain understanding. The reader should take what I say about geometry or topology with a grain of salt.

The origins of the lemma are in the theory of Shimura varieties and in the theory of harmonic analysis on reductive groups over $\mathbb{R}$. This second source is analytic and algebraic, the theory of the spectral decomposition of invariant distributions on real reductive groups, a theory that we owe almost in its entirety to Harish-Chandra, although the basic idea, that the pertinent eigenfunctions are characters, was introduced in the context of finite groups by Dedekind and Frobenius. What was imposed on our attention by the theory of Shimura varieties and the trace formula, was the understanding that for reductive algebraic groups there are two different notions of conjugation invariance: invariance and stable invariance. These are a result of two different kinds of conjugacy in, say $G(\mathbb{R})$, but more generally in $G(F)$, where F is a local field, archimedean or nonarchimedean. One is conjugacy in $G(F)$ itself, the other is conjugacy in $G(\bar{F})$, where $\bar{F}$ is the (separable) algebraic closure of F. It was only as we began the study of the zeta-functions of Shimura varieties with the help of the trace formula that the importance of the distinction, its consequences, and the attendant difficulties were recognized. They led to the fundamental lemma.

The issue, at first, is less the fundamental lemma, which can take diverse forms, than its consequences, not only for Shimura varieties but more importantly for harmonic analysis, both local and global. With the fundamental lemma, it is possible to create a theory of endoscopy that reduces invariant harmonic analysis, even various forms of twisted-invariant harmonic analysis, on arbitrary reductive groups to stably invariant harmonic analysis on quasi-split groups. It is the latter in which the notion of functoriality is best expressed, and it is functoriality, still to a large extent conjectural, that is the source of the arithmetic power of representation theory and harmonic analysis. Specific forms of functoriality have already been used in the course of establishing Fermat's theorem and other conjectures of

considerable interest to arithmeticians.

The fundamental lemma, once proved, offers two methods to attack functoriality: the first more immediate; the second much more encompassing. Although more limited, the first is of great importance, as it has offered to Arthur reasons for developing the general trace formula, which thanks to him, has been given a chance to demonstrate the enormous power of nonabelian harmonic analysis, of which the trace formula is an expression, for arithmetic. The lemma allows global, and presumably also local, transfer of stable characters from the endoscopic groups H for a given group G provided with a twisting, perhaps trivial, to the group groups G itself. The best reference for this type of theorem will be Arthur's book *The Endoscopic Classification of Representations: Orthogonal and Symplectic Groups*. It promises to increase greatly the confidence of mathematicians at large in the notion of functoriality, even though the functoriality yielded directly by endoscopy is limited. I add that, in my view, the central issue in endoscopy is the theory with no twisting.

After the introduction of endoscopy, there were a good many years during which I did not pay much attention to the attempts to develop it, by Waldspurger, Hales, and others on one hand, and by Goresky, Kottwitz, and MacPherson on the other. These contributions not only made possible the final proof of the lemma in the hands of Laumon and then Ngô, but introduced ideas that will, I expect, play a major role in the continuing attack on functoriality.

The principal tool of Harish-Chandra in the development of harmonic analysis on real reductive groups and then, later, of Shelstad's treatment of endoscopy were the bi-invariant differential operators on the group. The spectral decomposition amounts to a spectral decomposition of this family of commuting operators on $L^2(G(\mathbb{R}))$. This is a local theory. Although a great deal of effort has been spent on nonarchimedean fields, the theory has not reached the same stage, in good part because the spectral theory could not be reduced to one for a commutative family. My impression on studying the work of Waldspurger, Laumon and Ngô, without yet in any sense mastering it, is that the cohomology theory of perverse sheaves may offer a substitute, so that the possibilities offered by Waldspurger's reductions have by no means been exhausted.

Without any real knowledge of perverse sheaves as I began the study of Ngô's proof, and the earlier work with Laumon, and still only superficially informed, I am struck by the advantages of working with them. At the coarsest of levels, the orbital integrals provide over $\mathbb{R}$ or $\mathbb{C}$ the transfer that is dual to the transfer of characters from Cartan subgroups H of G, or better, although the theory has not been properly developed in this form even over $\mathbb{R}$, the transfer of characters implied by functoriality. Something similar will, I suppose, be true for nonarchimedean fields, but it will be more delicate because some irreducible characters are not associated to a Cartan subgroup, for example, those associated to representations of the local Galois groups as tetrahedral representations. What, in my view, is taking place in Waldspurger's analysis, although I have yet to examine it with sufficient care, or even any care, is a reduction of the local analysis to the study of orbital integrals on Lie algebras, not over a local field, but over a finite field, or, better expressed, in the context of algebraic geometry over a finite field. The asymptotic behavior described by the germs of Shalika becomes at this level, a question of direct images of perverse images

and their support, thus a behavior that is strictly geometric and strictly within the range of behavior encountered already in the study of these sheaves. I can imagine that the geometric information available through this translation might replace Harish-Chandra's study of the orbital integrals and their jumps to characters of G. Something similar to the jump conditions that Harish-Chandra met, and even something more subtle, might appear. I imagine that, when examining the possible behavior of the direct images with care, one will find behavior that can only be explained with the help of local Galois groups that admit surjective homomorphisms to relatively complex solvable groups. These matters will have to be studied on their own.

This kind of local information will be necessary if the program proposed for the utilisation of the stable trace formula, — a formula available only after the fundamental lemma has been established —is to succeed in establishing functoriality. It is to be utilised in combination with the Poisson formula on the Steinberg-Hitchin base, an affine object introduced by myself with Frenkel and Ngô. The introduction of the Poisson formula was suggested by Ngô's use of the Hitchin base,

None of this explains the reasons for the success of Ngô nor for the earlier partial successes of Goresky-Kottwitz-MacPherson and Laumon-Ngô. Moreover, with the exception of Arthur's laudation, little attention has been paid in various expositions to the needs of specialists of the theory of automorphic representations, thus of those to whom the lemma itself is of the most interest and who may, like me, have little, if any, familiarity, with stacks, perverse sheaves, or equivariant cohomology. So it may be worthwhile for me to have attempted to describe some glimpses of understanding that I have had while trying to penetrate their thoughts. I still have a long way to go and I am not certain that these glimpses are not will-o-the-wisps. Waldspurger and one or two others may have clearer notions of the possibilities than I.

The fundamental lemma itself appears in the context of orbital integrals, thus integrals over the conjugacy classes $\{g^{-1}\gamma_G g\}$ defined by elements $\gamma = \gamma_G \in G(F)$, F a local field, for the present nonarchimedean. For γ_G semisimple and regular, the conjugacy classes within the stable conjugacy class of γ_G are parametrized, in essence, by the elements of the abelian group $H^1(\mathrm{Gal}(\bar{F}/F), T)$, T the centralizer of γ_G. If κ is a character of this group, we may form $\sum \kappa(\gamma'_G)\mathcal{O}_G(\gamma'_G, f_G)$, where the sum over conjugacy classes is to be interpreted as a sum over $H^1(\mathrm{Gal}(\bar{F}/F), T)$, and f_G is the unit element of the Hecke algebra over G. Associated to κ is an endoscopic group, thus a quasi-split reductive group H and to γ a stable conjugacy class $\{\gamma_H\}$ in H, for which we can form a stable sum $\mathcal{O}^{\mathrm{st}}_H = \sum \mathcal{O}_H(\gamma'_H, f_H)$, where f_H is the unit element in the Hecke algebra of H. The fundamental lemma, in its simplest and earliest formulation, is the equality of these two sums, up to a well-defined constant factor that will necessarily depend on the choice of Haar measure on G and H.

After Waldspurger's reduction, a new, but similar equality appears with integrals over a set determined by an element γ, again often semisimple and regular, of the Lie algebra $\mathfrak{g}$ of G (or H) over F', again a local field but of positive characteristic, the ring of formal power series over a finite field k. Not having followed the developments over the years, I find the transition from one context to the other abrupt. My intuition is often brought up short. In addition, the proof of the fundamental lemma, like early proofs in local class

field theory and occasionally elsewhere, is an argument from a global statement to a local statement, so that the function field F of a complete nonsingular curve X over k of which F' is a completion at some place v is introduced. G is replaced by a group over this new F and γ by an element of the Lie algebra $\mathfrak{g}$ of G, or more precisely by a section of the Lie-algebra bundle defined by a G-bundle over X, a section that is allowed to have poles of large but finite order at a certain number, again large but finite, of points. It is in this difficult, especially for those not sufficiently conversant with the notions of modern algebraic geometry, context that the proof functions.

I was first disoriented by the appearance of Picard varieties in this context. They seemed to be of the usual type, thus closely related to abelian varieties. It was only after some time, when I noticed that the point of departure was the first cohomology group of a torus — thus a multiplicative group — the centralizer of γ, and that it was entirely possible that the transition from the local field F to the function field F of X and from Galois cohomology to etale cohomology or other cohomologies might entail the appearance of Picard varieties, that I began to feel more at ease. Galois cohomology groups have not been for me geometric objects. As descriptions of families of line bundles, thus of cohomology groups with values in $GL(1)$ or, possibly, other abelian algebraic groups, Picard varieties (or stacks) may be representable — whether by varieties or by stacks — and thus subject to study by the usual methods of algebraic geometry. Once reoriented, I found it much easier to follow, at least superficially, the presentations by Ngô and others of the geometrical proofs of the fundamental lemma, in the final form as well as in the earlier forms.

There are nevertheless in Ngô's proof and in the reflexions of other authors that preceded it several notions of which my grasp is tenuous, equivariant cohomology on the one hand and the — apparently — related notion of stacks on the other. Some aspects of the structure of the proof are quite clear. At a given place of X that is defined over k, in particular at the place with which we began, the orbital integrals, both for G and for H, can be interpreted as counts, although the count is a weighted count because centralizers of the elements γ interfere. One of the functions of stacks and equivariant cohomology, for those who understand them, is to take this weighting into account. That said, thanks to the passage to a global context, in the sense of algebraic geometry, thus to the passage to X and bundles over X, the counting, or rather the equality of two different counts asserted by the fundamental lemma, is replaced, in the spirit of the Weil conjectures and the Lefschetz formula, by an isomorphism of cohomology groups. The global count is, however, a sum over the points of X of local counts, so that, a global equality once established in general, it is necessary to return to X and to the section of the $\mathfrak{g}$-bundle that replaced the original γ, and to make choices that allow us to isolate the local contribution with which we began. Most of the effort is expended on the proof of the global cohomological statement — in the context of perverse sheaves for the etale cohomology and in the context of stacks.

I found it difficult to discover and keep firmly in mind the nature of the local count. There are at least two parameters at hand: the point of X and the point γ, which is now a section φ of the Lie algebra of a G-bundle E on X the total order of whose poles is controlled by a divisor D. The family $\mathcal{M}$ of these *Hitchin pairs*, (E, φ), is an essential element of the theory. The family of the classes in the Lie algebra of the group in question, G or one of its endoscopic groups H — as the case may be — is the Hitchin base, a

designation now familiar, thanks to Ngô, to a wide mathematical audience. The count is made over this base. Rather, the count is made, for both G and H, after a projection to this base. The domain of the projection is, to a first approximation, a scheme whose points are, first, a G-bundle on the given base X and, secondly, the section γ. So, implicit in the discussion is, I suppose, the existence of moduli spaces or stacks and an understanding of the cohomology of perverse sheaves defined on them. Most of this, and much else, I have to take on faith at present.

The Hitchin base is, as an algebraic variety over k, an affine space. The count on the fiber is made indirectly, through the direct images of the cohomology of the fiber. This fiber has, I believe, two important features: one that it shares with the usual Picard varieties, namely an action of a very large connected group, sometimes an abelian variety; this large group is defined over the Hitchin base. If I understand correctly the explanation in Ngô's Hyderabad lecture, an important consequence is that the action of the full group, a Picard group (rather stack!) $\mathcal{P}$ in the sense of Ngô, on the cohomology of the fibers is defined through a discrete quotient, denoted $\pi_0(\mathcal{P})$ by Ngô, a possibility that is certainly plausible from a topological point of view. This discrete quotient is closely related to the Galois cohomology groups $H^1(\mathrm{Gal}(\bar{F}/F), T)$ with which we began. These things are well explained in Ngô's Hyderabad lecture, where it is also explained that the local discrete quotients can be patched together, but in the etale topology, to form a sheaf of abelian groups. It is somewhat comforting, and perhaps not altogether incorrect, if we think of this as a patching in the etale topology of the various $H^1(\mathrm{Gal}(\bar{F}/F), T)$, defined for widely varying tori T. In any case this allows the discrete quotient and its characters to be introduced globally, something that was done in a different manner in the original formulation of the lemma.

The result is a sheaf over the Hitchin base that permits an action of the group $\mathcal{P}$. Since $\mathcal{P}$ acts on the fibres over the base, its action defines an action on the direct image of the cohomology on the Hitchin base, an action that factors through $\pi_0(\mathcal{P})$. Consequently the direct image can be decomposed as a direct sum with respect to the characters κ of $\pi_0(\mathcal{P})$. The principal theorem of Ngô, at least in connection with the fundamental lemma, is to establish that each component of the direct sum is isomorphic to a similar component for an endoscopic group H over X, a group defined by the character κ

There is a fluidity in the development of the proof that Ngô captures in his various expositions. Ideas appear, reveal themselves as suggestive, but ultimately inadequate, and then reappear in a different, often more difficult, guise. It is probably impossible to understand the final proof without some feeling for these initial stages: for equivariant cohomology in all its guises and, above all, for the geometry of the Hitchin fibration. I certainly have a long way to go, but I find the relatively concrete form in which this fibration is used by Laumon-Ngô in the proof of a special case of the fundamental lemma a helpful guide to the general case.

Since the Hitchin fibration and its properties are basic, a word or two about its construction may not be inappropriate. For a vector bundle, thus for a $GL(n)$-bundle, one can associate to the section γ, or better to the point a in the Hitchin base, a matrix valued function on X, and to each point $x \in X$, the n points in an n-dimensional space given by its eigenvalues. As x varies these points trace a curve, an n-fold covering $Y_a = Y_\gamma$ of X.

With γ we can introduce, at least in favorable circumstances, more: for each point x and each of the eigenvalues, a line, the eigenspace corresponding to the eigenvalue. Thus the section γ defines a line-bundle on Y_a. There are questions that arise at the points where the eigenvalues are multiple, but we do see line-bundles on the horizon and therefore, perhaps, abelian varieties and cohomology groups in degree 1, groups related to those with which endoscopy began. The abelian varieties are a sign that in the new context, these cohomology classes appear as line bundles that give rise to representable functors, whose points can be described geometrically. The Hitchin fibration, as defined by Ngô, provides similar constructions for a general group. Even in the original form, the eigenvalues associated to γ define at each point of X a diagonal matrix, but as the order of the eigenvalues is not prescribed, it is in fact only the conjugacy class of this diagonal matrix that is determined.

At the level of groups we cannot, so far as I know, ordinarily find a map from conjugacy classes to matrices that is inverse to that from matrices to conjugacy classes, but at the level of Lie algebras, low characteristics aside, we can. For example, for the group $SL(2)$, the conjugacy class is given, at least at the regular elements by the determinant, a, and the representative matrix for this class can be taken to have diagonal elements 0 and off-diagonal elements 1 and a. There are, I believe, various such lifts. Ngô uses the one associated to the name of Kostant. Our original description of the spectral curve Y_a was deliberately vague about its form at those points where eigenvalues coincide and it is best here to pass over in silence the difficulties they entail in Ngô's definitions. They entail technical difficulties that I have not yet made any attempt to understand. Indeed, I am not much beyond the introduction to his paper. In any case, what results is a lift not only of the regular conjugacy classes of the Lie algebra to the Lie algebra itself, but an abelian group over these lifts. It is closely related to the centralizer of the lifts and yields a fibration in groups over the Hitchin base. The dimension of the fibers is the rank of G. In the definition of the Picard variety (stack) relevant to the Hitchin fibration and to Ngô's analysis, the bundles associated to this fibration in groups replace the line bundles of the classical theory. I have to remind myself constantly that there are two parameters at play in this fibration: the base a, given by the class of γ, and a point x of X, at which γ is essentially an element in the Lie algebra of G, say over the residue field or over the coordinate ring at x.

As already observed, the argument for the proof of the fundamental lemma proceeds in two stages; first for a fixed a and all of X, but fortunately only for well-chosen a; secondly for a suitable X and a suitable point x of X. We have already described the projection at the first stage, from the total space of the Hitchin fibration to the Hitchin base, and the decomposition of the direct image according to the characters κ of the Picard stack.

There is an equality of sheaves over X to be proven at the first stage. There are two issues in the proof of the equality: the support of the relevant direct images; the equality on this support. An endoscopy group is so defined that there is a morphism of the Hitchin base $\mathcal{A}_H$ to $\mathcal{A}_G$. So we can compare the direct image of a sheaf on $\mathcal{A}_H$ with a sheaf on $\mathcal{A}_G$. The sheaf on $\mathcal{A}_G$ is defined by the part of the direct image of the sheaf associated to the character κ. For H, one does the same thing, but the character for H is taken to be trivial. If H is associated to κ, it has first to be shown that the direct image of the κ-component for G is supported on the image of the Hitchin base for H. This is, in

principle, a consequence of the definitions, but it is not an easy consequence. Indeed the final proof is tremendously daunting.

Those of us with less than adequate facility with the concepts can best begin with the theorem for unitary groups proved by Laumon-Ngô, because in their Annals paper, not only does the Picard stack appears in its primitive form in terms of the spectral curve Y_a but, in addition, the proof of the necessary *homotopy lemma*, which is used to deal with the problem of support, appears to be at an altogether different level of difficulty than the *support theorem* of the paper under review. In the Annals paper, both G and H are unitary groups. Since a unitary group is a form of $GL(n)$, the concept of spectral curve has a more immediate geometric content and there is a more direct relation between the Hitchin fibrations of G and H that appears to simplify the arguments considerably.

I have already adumbrated the final step of the proof. If the curve X and an element a of the Hitchin base are given, they define locally at any point x of X the elements for the original statement of the fundamental lemma for the Lie algebra, an element of the local Lie algebra and a group $G(\mathcal{O}_x)$. Moreover the equality of a κ-component of the direct image at a with a direct image for an endoscopic group H can be interpreted, thanks to the Grothendieck-Lefschetz theorem as an equality of the product over the points of X of two counts, one for κ-components on G, one for H. If we can choose x, X, and a so that they reproduce any arbitrarily given local data and if X and a are also chosen such that the fundamental lemma is true at all points $x' \neq x$ of X, we can cancel all terms in the product but those at x and deduce the desired equality at that point. We cannot, apparently, expect to choose X such that the fundamental lemma is utterly obvious away from X, but it can be so chosen that it is accessible to direct computation. To establish the existence of X and a with the necessary properties requires, both in the final paper and in the earlier paper on unitary groups, very sophisticated algebro-geometrical methods. It is also important for its existence that the poles of the section defining a are allowed to grow in number.

For the unitary groups, the very last step, the deduction of the fundamental lemma outside of x from the properties of X and a appears almost an elementary exercise in geometry over finite fields. This is not so in general. Further struggles with perverse sheaves await the reader.

It is certain that, for the majority of specialists in nonabelian harmonic analysis and representation theory, thus, in particular, for specialists in the theory of automorphic representations and the associated arithmetic, certainly for me, it will take more than a few weeks, or even a few months, to assimilate the techniques from contemporary algebraic geometry that are required for the proof of the fundamental lemma. How long it might take geometers to understand fully the questions posed by the arithmetic and the analysis, I hesitate to guess. This might be easier. Certainly representation theory has a briefer and, in some respects, narrower history, but it is less familiar to the majority of mathematicians. Time will tell.

Since, as I intimated at the beginning of this review, the fundamental lemma is an essential and fundamental contribution to a theory that will not be developed by specialists in algebraic geometry alone, there will be a need for further, more accessible expositions of the methods of this paper and those that preceded it, with examples, even very simple

examples, and with considerably more explanation of the geometric intuition implicit in the abstract theory. An index to definitions and symbols would also be welcome! The present paper is 168 pages long and these pages are large and very full. An exposition genuinely accessible not alone to someone of my generation, but to mathematicians of all ages eager to contribute to the arithmetic theory of automorphic representations, would be, perhaps, four times as long, thus close to 700 pages. It would, I believe, be worth the effort.