Duality, the Langlands Correspondence, and the Influence of Michael Atiyah

Edward Witten

Lecture at Isaac Newton Institute, September 23, 2021

I met Michael Atiyah in the spring of 1977 when he was lecturing at Harvard. He invited me to visit in Oxford at the end of 1977.

During that visit, Michael was lecturing about gauge theory instantons in four dimensions. With Vladimir Drinfeld, Nigel Hitchin, and Yuri Manin, he had just discovered the ADHM construction of instantons via twistor theory and algebraic geometry. Instantons were an exciting topic in physics at the time, largely because of the work of Gerard 't Hooft in using instantons to solve the "U(1) problem" of the nuclear force. By the end of 1977, we also knew - from an observation by Albert Schwarz - that 't Hooft's work on the U(1) problem could be understood as an application of the Atiyah-Singer index theorem.

So for most of that month, there were very lively discussions about instantons. But toward the end of the month, Atiyah changed the subject. He drew my attention to two physics papers that I had not seen, one by Peter Goddard, Jean Nuyts, and David Olive, and one by Olive with Claus Montonen. He said that he thought there was something deep here and I should go to London to discuss the matter with Olive.

Here is the abstract of the paper of Goddard, Nuyts, and Olive:

Nuclear Physics B125 (1977) 1–28

GAUGE THEORIES AND MAGNETIC CHARGE

P. GODDARD
DAMTP, Cambridge University

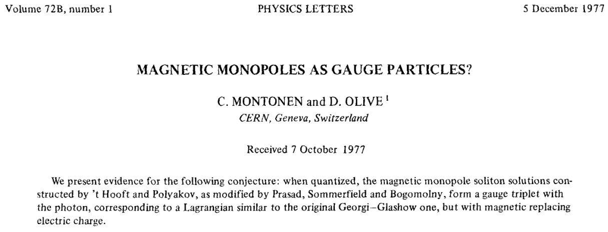
J. NUYTS
Department of Physics, University of Mons

D. OLIVE
CERN, Geneva
and
The Niels Bohr Institute, Copenhagen ★

Received 14 January 1977

If the magnetic field for an exact gauge group H (assumed compact and connected) exhibits an inverse square law behaviour at large distances then the generalized magnetic charge, appearing as the coefficient, completely determines the topological quantum number of the solution. When this magnetic charge operator is expressed as a linear combination of mutually commuting generators of H, the components are uniquely determined, up to the action of the Weyl group, and have to be weights of a new group H^v which is explicitly constructed out of H. The relation between the "electric" group H and the "magnetic" group H^v is symmetrical in the sense that $(H^v)^v = H$. The results suggest that H monopoles are H^v multiplets and *vice versa* and that the true symmetry group is $H \otimes H^v$. In this duality topological and Noether quantum numbers exchange rôles rather as in Sine-Gordon theory. A physical possibility is that H and H^v be the colour and weak electromagnetic gauge groups.

And here is the Montonen-Olive abstract:



Volume 72B, number 1 PHYSICS LETTERS 5 December 1977

MAGNETIC MONOPOLES AS GAUGE PARTICLES?

C. MONTONEN and D. OLIVE [1]
CERN, Geneva, Switzerland

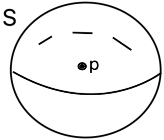
Received 7 October 1977

We present evidence for the following conjecture: when quantized, the magnetic monopole soliton solutions constructed by 't Hooft and Polyakov, as modified by Prasad, Sommerfield and Bogomolny, form a gauge triplet with the photon, corresponding to a Lagrangian similar to the original Georgi–Glashow one, but with magnetic replacing electric charge.

Atiyah explained to me that what Goddard, Nuyts, and Olive (GNO) were defining as the “magnetic gauge group” was actually the same as the Langlands dual group in the context of the Langlands program of number theory. I am pretty sure that Atiyah is the one who made this connection, though as far as I know he never wrote anything about it. Like other physicists of the time, I had never heard of the Langlands program, or of the name Robert Langlands. But Atiyah assured me that the Langlands program of number theory was something very deep and that he thought the physicists must likewise be on to something important.

Let me first explain a little of what Goddard, Nuyts and Olive had said. Let us consider gauge theory, first for the case of abelian gauge group $G = U(1)$ – in the real world, the gauge group of electromagnetism. There are "electrically charged particles" such as the electron. The most fundamental definition of the electric charge of a particle is that it is a representation of G associated to the particle, although, since every representation of $U(1)$ is the n^{th} power of a fundamental representation, for some integer n, you can think of electric charge as just being an integer. A less abstract way to describe this is to say that electric charge is "quantized" and the charge of any particle is an integer multiple of the charge of an electron.

In $U(1)$ gauge theory, the "gauge field" is a connection A on a unitary complex line bundle $\mathcal{L}$ over spacetime. We write $F = \mathrm{d}A$ for the curvature of A. The quantum wavefunction of the electron is a section ψ of $\mathcal{L}$. Dirac in the 1930's had introduced the idea of a magnetic monopole (as a quantum object; in classical physics the idea was older). Dirac's concept of the monopole was that the position of the monopole is a singularity p in space at which the line bundle $\mathcal{L}$ and connection A are undefined; the complement in $\mathbb{R}^3$ of the point p is contractible to a two-sphere S



and the magnetic charge is

$$m = \int_S c_1(\mathcal{L}),$$

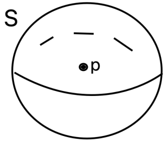
an integer.

If we ask that the connection A should satisfy Maxwell's equations, we can determine what the connection must be for a monopole at rest. The curvature is

$$F = \frac{m}{2} \star \mathrm{d} \frac{1}{|\vec{x}|}.$$

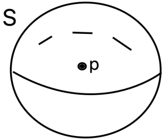
If and only if m is an integer, there is a line bundle $\mathcal{L}$ with connection A over $\mathbb{R}^3 \backslash p$ such that the curvature $F = \mathrm{d}A$ is what I have written; $\mathcal{L}$ and A are unique up to isomorphism.

Now let us discuss nonabelian gauge theory, say with a simple nonabelian gauge group G. Over spacetime, there is now a G-bundle E, with connection A and curvature $F = \mathrm{d}A + A \wedge A$. Electric charge is easy to define. The electric charge of any particle is a representation R of G. What we should mean by magnetic charge is more subtle. The most direct analog of what I said for $G = U(1)$ is to again say that a monopole is a point (or a small region) in space at which E and A are not defined and that magnetic charge classifies the choice of a G-bundle over the surrounding two-sphere S



With this interpretation (which is appropriate in the right context), magnetic charge of G takes values in $\pi_1(G)$.

That is not the answer that GNO wanted. They considered a theory in which, by a Higgs mechanism (the same Higgs mechanism that was famously confirmed by the LHC accelerator at CERN in 2012) the gauge group is effectively reduced (at low energies, or for most practical purposes) to a maximal torus $T \subset G$. If this reduction only occurs away from the point p, then, reasoning as before, we can define a magnetic charge that classifies T-bundles (rather than G-bundles) over a two-sphere



so it takes values in $\pi_1(T)$, which (for a torus T) is the same as $\mathrm{Hom}(U(1), T)$.

On the other hand, once we assume that the gauge symmetry is effectively reduced from G to a maximal torus T, we should classify electric charge by a representation of T. Irreducible representations of the abelian group T are 1-dimensional, so they correspond to elements of $\mathrm{Hom}(T, U(1))$. So in the setup considered by GNO, they had:

- Electric charge: $\mathrm{Hom}(T, U(1))$.
- Magnetic charge: $\mathrm{Hom}(U(1), T)$.

GNO then went on to make the observation that had attracted the interest of Michael Atiyah: For any simple Lie group G, there is another simple Lie group $G^\vee$, with maximal torus $T^\vee$, such that

$$\mathrm{Hom}(T, U(1)) = \mathrm{Hom}(U(1), T^\vee)$$

and reciprocally

$$\mathrm{Hom}(U(1), T) = \mathrm{Hom}(T^\vee, U(1)).$$

In other words, in the GNO interpretation (at least with the gauge symmetry reduced or "Higgsed" to a maximal torus on each side) electric charge of G is magnetic charge of $G^\vee$, and vice-versa. Of course, some of you will notice that, as originally observed by Atiyah, the GNO dual group $G^\vee$ is the same as the Langlands dual group.

The point of the second paper of Montonen and Olive

Volume 72B, number 1 PHYSICS LETTERS 5 December 1977

MAGNETIC MONOPOLES AS GAUGE PARTICLES?

C. MONTONEN and D. OLIVE [1]
CERN, Geneva, Switzerland

Received 7 October 1977

We present evidence for the following conjecture: when quantized, the magnetic monopole soliton solutions constructed by 't Hooft and Polyakov, as modified by Prasad, Sommerfield and Bogomolny, form a gauge triplet with the photon, corresponding to a Lagrangian similar to the original Georgi–Glashow one, but with magnetic replacing electric charge.

was to propose a role for all this in quantum dynamics.

Quantum gauge theory is a generalization of classical gauge theory that depends on an extra parameter, the gauge coupling e (a single parameter in the case of a simple gauge group). Quantum gauge theory reduces to classical gauge theory when e is small. Quantum gauge theory is quite difficult to understand when e is not small.

Gauge theories with different gauge groups are not equivalent classsically. But Montonen and Olive proposed that they can be equivalent quantum mechanically. More specifically, they proposed a particular example of a pair of gauge theories, one with gauge group G, one with gauge group $G^\vee$, which they suggested would be equivalent quantum mechanically, but not classically. The equivalence was supposed to be such that electric charge in the first theory is mapped to magnetic charge in the second, and vice-versa. Because of this, Montonen-Olive duality is often called electric-magnetic duality. The duality was supposed to invert the coupling constant

$$e^{\vee\,2} = \frac{4\pi}{e^2}$$

(where $e^\vee$ is the gauge coupling of the $G^\vee$ theory) and since e and $e^\vee$ cannot be simultaneously small, this would account for why the duality is not visible classically.

It is going to be hard to properly convey how wild the Montonen-Olive conjecture seemed at that time. Quantum gauge theory was only understood for fairly small e, and the region of large e where quantum effects are big was really *terra incognita*. The Montonen-Olive conjecture seemed like a jump not just beyond what we understood but beyond what we could possibly understand. But anyway Montonen and Olive had some striking possible evidence for their conjecture, namely they showed that certain classical formulas for the masses of various particles – particles with electric and/or magnetic charge – satisfied relations which, if valid quantum mechanically, would follow from their duality conjecture.

I took Atiyah's advice and arranged to talk to Olive in London, but by the time I got there, I was pretty skeptical. What bothered me about the GNO paper was that the matching between electric and magnetic charge only worked after Higgsing, that is, reduction to a maximal torus. I felt a fundamental statement should not depend on this Higgsing. What bothered me about the Montonen-Olive paper was that their technical assumption of vanishing of the "Higgs potential" in their theory was, by all standard logic, not consistent quantum mechanically.

The point that bothered me in the GNO paper was not resolved until almost 30 years later, but Olive and I actually succeeded in resolving my technical objection to the Montonen-Olive paper. Their technical assumptions are not valid in the ordinary or "bosonic" gauge theory that they assumed, but they actually are valid in a supersymmetric version of the theory. A supersymmetric theory is roughly a theory of differential forms on a function space rather than functions on a function space. And we showed that with at least $N = 2$ supersymmetry, the classical mass formulas of Montonen and Olive are actually valid quantum mechanically. (Supersymmetric gauge theories in four spacetime dimensions are classified by the amount of supersymmetry, $N = 1, 2$ or 4. The $N = 4$ theory is unique for a given gauge group; a theory with $N = 2$ is not completely determined by the gauge group but it is highly constrained. The $N = 1$ theory is less highly constrained and is closer to realistic physics. Very roughly, $N = 1, 2$, and 4 correspond to complex numbers, quaternions, and octonions.)

The resulting paper was probably the most significant one I had written up to that point:

Volume 78B, number 1 PHYSICS LETTERS 11 September 1978

SUPERSYMMETRY ALGEBRAS THAT INCLUDE TOPOLOGICAL CHARGES

E. WITTEN [1]
Harvard University, Society of Fellows, Cambridge, MA, USA

and

D. OLIVE
The Blackett Laboratory, Imperial College of Science and Technology, London, England

Received 10 May 1978

We show that in supersymmetric theories with solitons, the usual supersymmetry algebra is not valid; the algebra is modified to include the topological quantum numbers as central charges. Using the corrected algebra, we are able to show that in certain four dimensional gauge theories, there are no quantum corrections to the classical mass spectrum. These are theories for which Bogomolny has derived a classical bound; the argument involves showing that Bogomolny's bound is valid quantum mechanically and that it is saturated.

But I drew the wrong conclusion from it. I thought that we had explained the formulas of Montonen and Olive without the radical assumption of their duality, so that there was no real evidence for the duality.

Even if I had drawn the right conclusion, I don't think the time was ripe to understand Montonen-Olive duality, so I doubt I would have made too much progress in that period. But a few noteworthy things were done in those early days.

For one thing, Hugh Osborn took a closer look at the quantization of the monopole and discovered that only the $N = 4$ theory (not the more general $N = 2$ theories that Olive and I had considered) has the right properties for Montonen-Olive duality:

Volume 83B, number 3,4 PHYSICS LETTERS 21 May 1979

TOPOLOGICAL CHARGES FOR $N = 4$ SUPERSYMMETRIC GAUGE THEORIES AND MONOPOLES OF SPIN 1

Hugh OSBORN
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England

Received 1 February 1979

The central charges in the supersymmetry algebra of the $N = 4$ supersymmetry gauge theory are obtained. When spontaneous symmetry breaking is imposed it is shown that the spins of the topological monopole states should be identical to those of the massive elementary particles.

Later I will explain what happened for $N = 2$.

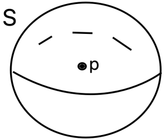
At the time, I remained skeptical about Montonen-Olive duality, but in any event it also seemed clear – and on this I was surely correct – that even if this duality is correct, it was not going to be possible to make significant progress in understanding it using methods available at the time. However, before going on, I should point out one basic fact that was important later. For the case of an abelian theory with gauge group $U(1)$ – to be precise, a "free" theory without charged fields – the duality actually is correct, and in this case supersymmetry is not required. For $U(1)$, Montonen-Olive duality is a quantum version of Hodge duality for harmonic forms. In other words, Maxwell's equations $\mathrm{d}F = \mathrm{d}\star F = 0$ in vacuum say that the curvature F of a $U(1)$ gauge field is a harmonic form. This condition is invariant under Hodge duality

$$F \to \star F$$

and for gauge group $U(1)$, Montonen-Olive duality is a quantum version of this.

But anyway, for a number of years there was not much progress with Montonen-Olive duality. During this period, Atiyah was working – among other things – on monopoles. I would like to explain a little about that work.

Dirac had considered a monopole in $U(1)$ gauge theory, and in that description the monopole is a singularity:



The equations we know in physics are not complete and usually, a "singularity" in some description of a physics problem is really a region in space or spacetime where the given level of description breaks down and a more complete description is needed. For example, electromagnetism is described classically – and even in Quantum Electrodynamics – as a $U(1)$ gauge theory, but in the Standard Model it turns out that a more complete description requires the gauge group $U(2)$ – with a "Higgs mechanism" that accounts for why the theory can be approximated for many purposes as a $U(1)$ gauge theory.

In the early 1970's, 't Hooft and A. M. Polyakov had shown that if one replaces $U(1)$ gauge theory by an $SU(2)$ gauge theory with "Higgsing" to $U(1)$ – meaning that in low energy observations, the gauge group is effectively reduced to $U(1)$ – then a completely smooth description of the monopole is possible. In this description (and for small e!), the monopole corresponds to a solution of a nonlinear partial differential equation.

The simplest equation describing monopoles is the "Bogomolny equation" (introduced by E. B. Bogomolny) which is an equation for a connection A on a G-bundle $E \to \mathbb{R}^3$ and a zero-form ϕ (the "Higgs field") with values in the adjoint bundle $\mathrm{ad}(E)$. The equation is

$$F = \star \mathrm{d}_A \phi.$$

This equation is a sort of three-dimensional cousin of the instanton equation in four dimensions. It has many fascinating properties, many of which were explored by Atiyah and Hitchin in their book *The Geometry and Dynamics of Magnetic Monopoles* (1988). Among other things, the moduli spaces of solutions of the Bogomolny equations are hyper-Kahler manifolds. For magnetic charge 2, the "reduced" moduli space (factoring out trivial symmetries) is four-dimensional and Atiyah and Hitchin described it explicitly. This is the Atiyah-Hitchin manifold, a four-dimensional hyper-Kahler manifold with $SO(3)$ symmetry.

Another development of the same period, also influenced by Atiyah, was Donaldson theory of smooth four-manifolds, constructed using instantons of nonabelian gauge theory. Atiyah recommended understanding Donaldson theory as a problem for physicists. By 1988, it was clear that Donaldson theory could be understood as a "twisted version" of $N = 2$ super Yang-Mills theory. Remember that $N = 2$ super Yang-Mills theory is almost a candidate for a theory that could have Montonen-Olive duality. It satisfies the criterion of Olive and me, but not the more refined criterion of Osborn. The relation of $N = 2$ super Yang-Mills theory to Donaldson theory raised the question of whether progress with the $N = 2$ theory would lead to progress with four-manifolds. I was pessimistic about this, however.

By the early 1990's, the landscape concerning Montonen-Olive duality was changing. The main change, to me, was that colleagues – Ashoke Sen, John Schwarz, Mike Duff, Paul Townsend, Chris Hull, and unfortunately too many others to properly credit here – were developing conjectures analogous to Montonen-Olive duality, but for string theory rather than quantum field theory. The arguments were highly suggestive. To me, there was no smoking gun, and I did not see how one could hope to make solid progress. But I was not as skeptical as I had been in 1977.

The smoking gun was submitted to the arXiv on February 7, 1994:

ELSEVIER

16 June 1994

PHYSICS LETTERS B

Physics Letters B 329 (1994) 217–221

Dyon–monopole bound states, self-dual harmonic forms on the multi-monopole moduli space, and SL(2,Z) invariance in string theory

Ashoke Sen [1]

Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India

Received 3 March 1994
Editor: H. Georgi

Abstract

Existence of SL(2,Z) duality in toroidally compactified heterotic string theory (or in the $N = 4$ supersymmetric gauge theories), that includes the strong–weak coupling duality transformation, implies the existence of certain supersymmetric bound states of monopoles and dyons. We show that the existence of these bound states, in turn, requires the existence of certain normalizable, (anti-)self-dual, harmonic forms on the moduli space of BPS multi-monopole configurations, with specific symmetry properties. We give an explicit construction of this harmonic form on the two monopole moduli space, thereby proving the existence of all the required bound states in the two-monopole sector.

Sen had understood that Montonen-Olive dualty of $N = 4$ super Yang-Mills theory implies the existence of multi-monopole bound states. He showed that such bound states would correspond to L^2 harmonic forms of middle dimension on the reduced monopole moduli spaces – the hyper-Kahler manifolds that had been studied by Atiyah and Hitchin. For the two-monopole case, Atiyah and Hitchin had described the reduced moduli space explicitly, and their description was simple enough that Sen could explicitly demonstrate the existence of the requisite L^2 harmonic form.

This paper had an electrifying impact for me. For one thing, I viewed it as the first fundamentally new test of Montonen-Olive duality since the early days. After it passed this test, I really did not seriously doubt any more whether it was true. Moreover, Sen had shown that the tools at hand really were sufficient to make real progress. This inspired us all to try harder.

In roughly the preceeding year, Nati Seiberg had made significant progress on the "dynamics" (the quantum behavior) of $N = 1$ super Yang-Mills theory. $N = 1$ super Yang-Mills is closer to physics, but farther from Montonen-Olive duality, than $N = 2$. Nati invited me to become involved in trying to understand the quantum behavior for $N = 2$. Remember, $N = 2$ is the theory which is almost correct for Montonen-Olive duality. The Sen paper and certain qualitative considerations focused our attention on the possible role of electric-magnetic duality for understanding the dynamics of $N = 2$. What we eventually discovered (about five months after Sen's paper) is that $N = 2$ is governed by a sort of effective field theory version of Montonen-Olive duality. Because of a Higgs mechanism, the $N = 2$ theory can be generically described at low energies by an effective $U(1)$ theory. $U(1)$ gauge theory does have electric-magnetic duality, as I explained before. Combining these statements with some qualitative considerations and some known facts about $N = 2$ super Yang-Mills theory, Seiberg and I arrived at a description of this theory.

From a physical point of view, the main insight in our work is that although the theory can be *generically* described at low energies as a free $U(1)$ theory with massive charges and monopoles, there are special points in field space at which the monopoles behave as light (or even massless) elementary particles. This was rather dramatic, and was in the spirit of Montonen and Olive: a particle – the monopole – which for weak coupling is understood as an extended object, a solution of the Bogomolny equations, can behave for strong coupling like a pointlike elementary particle.

Applied to the twisted version of the $N = 2$ theory that is related to Donaldson theory, our results implied that the four-manifold information contained in the Donaldson invariants can alternatively be obtained by counting solutions of what are now generally called the Seiberg-Witten equations. The Seiberg-Witten equations are equations for a $U(1)$ connection A and a charged spinor field M:

$$F^+ = \overline{M}\Gamma M, \qquad \not{D}M = 0.$$

Here M is the massless monopole. In this description, M doesn't look like a solution of the Bogomolny equations, which it was in the book of Atiyah and Hitchin.

Many other developments followed, and it became clear in the mid 1990's that electric-magnetic duality is not just true, but is very important in the general understanding of both quantum field theory and string theory. However, the relation to the Langlands correspondence remained unclear.

Before going on, I have to backtrack and explain another thread in the story: Starting around 1990, Alexander Beilinson and Vladimir Drinfeld had formulated a "geometric" version of the Langlands correspondence, for an ordinary Riemann surface (rather than a number field). From the beginning, they made heavy use of ingredients familiar in physics, such as current algebra (affine Kac-Moody algebras). I made many efforts to understand what they were saying, but in those years I found it impossible. It seemed that ingredients familiar in physics had been scrambled up in a totally unfamiliar way.

What especially perplexed me were the geometric Hecke transformations (originally introduced, I believe, in the 1970's by M. S. Narasimhan and S. Ramanan) that are completely central in geometric Langlands. The setting is as follows. $E \to C$ is a holomorphic $G_{\mathbb{C}}$ bundle over a Riemann surface C ($G_{\mathbb{C}}$ is a simple complex Lie group). A Hecke modification of E at a point $p \in C$ is a holomorphic $G_{\mathbb{C}}$ bundle $E' \to C$ together with an isomorphism φ between E and E' over $C \backslash p$:

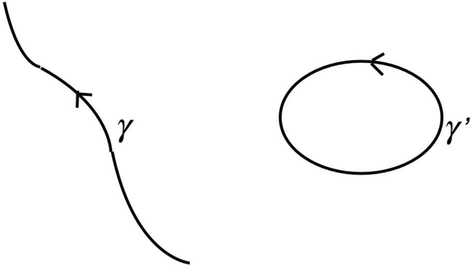
$$\varphi : E|_{C\backslash p} \cong E'|_{C\backslash p}.$$

For example, if $G_{\mathbb{C}} = \mathbb{C}^*$, then E is a holomorphic line bundle $\mathcal{L} \to C$, and an example of a Hecke modification of $\mathcal{L}$ at p is $\mathcal{L}' = \mathcal{L}(p) = \mathcal{L} \otimes \mathcal{O}(p)$, which by definition is isomorphic to $\mathcal{L}$ away from p. If $G_{\mathbb{C}} = SL(2, \mathbb{C})$, we can represent E by a rank 2 holomorphic vector bundle. For any local decomposition of E as a sum of line bundles $\mathcal{L}_1 \oplus \mathcal{L}_2$, an example of a Hecke modification of E would be $E' = \mathcal{L}_1(p) \oplus \mathcal{L}_2(-p)$.

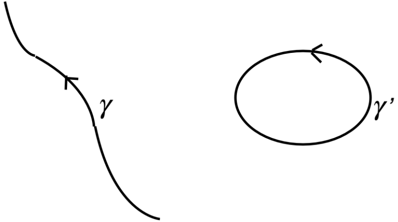
There are different types of Hecke modification possible at a given point p. For example, for $G_{\mathbb{C}} = \mathbb{C}^*$, instead of $\mathcal{L} \to \mathcal{L}(p)$, we could have taken $\mathcal{L} \to \mathcal{L}(np) = \mathcal{L} \otimes \mathcal{O}(p)^n$, for any integer n. An important input in the geometric Langlands program is the geometric Satake correspondence (developed by Lusztig, Drinfeld, Ginzburg, Mirkovic, and Vilonen, among others) which says that the possible types of Hecke modification for G are labeled by irreducible finite dimensional representations of the dual group $G^\vee$. (For example, for $G_{\mathbb{C}} = \mathbb{C}^*$, the dual group is $G_{\mathbb{C}}^\vee = \mathbb{C}^*$ again. Irreducible representations of $G_{\mathbb{C}}^\vee$ are labeled by an integer, and this is the integer n in $\mathcal{L} \to \mathcal{L}(np)$.)

For years, I puzzled over how such geometric Hecke transformations could possibly be related to physics – as I presumed they had to be if geometric Langlands was to have a physical interpretation. Eventually it turned out that there is a simple answer in terms of basic ingredients of quantum gauge theory, namely electric-magnetic duality between Wilson and 't Hooft operators, and the Bogomolny equations, which had been studied so thoroughly by Atiyah and Hitchin as well as a number of physicists. (This answer was one of the main ideas in the paper I wrote in 2006 with Anton Kapustin on gauge theory and geometric Langlands. There won't be time today to explain the other ideas.)

A "line operator" in quantum field theory, in general, is a modification of a theory along an embedded one-manifold γ. γ could be either a closed loop or an "open" curve with endpoints on the boundary of spacetime:



The most simple line operator is the Wilson operator, which is simply the holonomy of a connection, taken in some representation R of the gauge group. For a closed loop

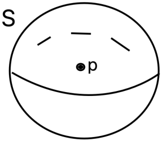


one defines the Wilson operator to be the trace, in the representation R, of the holonomy of a connection A

$$W_R(\gamma) = \mathrm{Tr}_R \mathrm{Hol}_\gamma A$$

(which physicists write as $\mathrm{Tr}_R P \exp \oint_\gamma A$). For an open curve, the Wilson operator needs a slightly longer discussion, which I omit. Note that there is a Wilson operator for every irreducible representation R.

The Wilson operator is widely used in physics to probe confinement of quarks, and around 1980 this motivated 't Hooft to try to define the electric-magnetic dual of a Wilson operator. The definition that he gave was adequate for his application, though it turns out that for geometric Langlands, we need a more precise definition. The original definition actually uses the naive definition of magnetic charge that we discussed at the beginning. On a four-manifold M, a one-manifold γ is of codimension 3. 't Hooft imagined doing gauge theory with a G bundle $E \to M \backslash \gamma$ that is only defined on the complement of γ. Since the "link" of a 1-manifold in four dimensions is a two-sphere S^2, there is a local invariant in this situation



which is the topological type of a G-bundle on this S^2, classified by $\pi_1(G)$. The 't Hooft operator as originally defined by 't Hooft is classified by this invariant.

However, Kapustin (2005) gave a more precise definition. His discussion differed from what 't Hooft had said in that he imposed a more precise condition on what is happening near the singularity, rather than just specifying the topology of the situation.

The idea is to ask that near γ, the fields are asymptotic to a specified singular solution of the Yang-Mills equations. The singularity considered is as follows. For a local model of a codimension 3 singularity, we can take the behavior at the origin in $\mathbb{R}^3$. For $G = U(1)$, the charge 1 "Dirac monopole" solution is a connection A_0 with curvature

$$F_0 = \frac{1}{2} \star \mathrm{d}\frac{1}{|\vec{x}|}.$$

For any G, with Lie algebra $\mathfrak{g}$, pick a homomorphism $\rho : \mathfrak{u}(1) \to \mathfrak{g}$ and ask that, near $\vec{x} = 0$, the connection A should be asymptotic to

$$A_\rho = \rho(A_0).$$

An 't Hooft operator of type ρ supported on γ is defined by the recipe "do gauge theory with this kind of singularity along γ."

Since we only care about A_ρ up to conjugation, we can think of ρ as a homomorphism to a maximal torus $\mathfrak{t}$:

$$\rho : \mathfrak{u}(1) \to \mathfrak{t}.$$

Of course we only care about ρ up to a Weyl transformation.

To summarize:

(1) Wilson operators in G gauge theory are classified by representations of G, which in turn are classified by highest weights. A highest weight is a homomorphism $\rho_{\mathrm{el}} : T \to U(1)$.

(2) 't Hooft operators in G gauge theory are classified by singularities that are labeled by homomorphisms $\rho_{\mathrm{mag}} : U(1) \to T$.

Now we can state the duality between Wilson and 't Hooft operators: in the Montonen-Olive duality between theories with gauge groups G or $G^\vee$, Wilson operators of G correspond to 't Hooft operators of $G^\vee$, and vice-versa, using the natural correspondence between

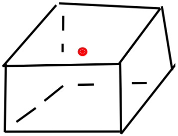
$$\rho : T \to U(1)$$

and

$$\rho^\vee : U(1) \to T,$$

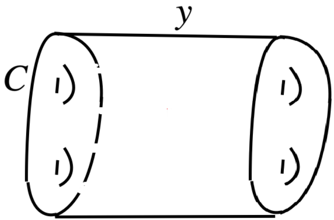
and the fact that G and $G^\vee$ have the same Weyl group. This is the same as the original GNO duality that Atiyah had pointed out to me in 1977, except that by restating it in terms of line operators rather than particles, it is not necessary to invoke a Higgs mechanism.

But what does this have to do with geometric Hecke transformations? Here we have to consider the Bogomolny equations, roughly because they govern a time-independent situation in a four-dimensional supersymmetric gauge theory. The Bogomolny equations, as we discussed earlier, are in three dimensions. In three dimensions, a codimension three singularity is at an isolated point,



So we have to solve the Bogomolny equations with an isolated singularity of Dirac monopole type. (The Bogomolny equations with this type of singularity were first studied by Kronheimer, 1986, followed by Pauly, 1990.)

In describing a "spectral curve" that governs solutions of the Bogomolny equations on $\mathbb{R}^3$, Atiyah and Hitchin used a decomposition $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R}^2$, and viewed $\mathbb{R}^3$ as a two-parameter family of copies of $\mathbb{R}$. Here we will instead use the same decomposition and view $\mathbb{R}^3$ as a one-parameter family of copies of $\mathbb{R}^2$. Actually, we can replace $\mathbb{R}^2$ by any Riemann surface C. So we will study the Bogomolny equations on $\mathbb{R} \times C$, viewed as a one-parameter family of copies of C. The picture is like this, where we parametrize $\mathbb{R}$ with a real variable y:



As explained by Atiyah and Bott, any connection A on a G-bundle $E \to C$ (where C is a Riemann surface) determines a holomorphic $G_{\mathbb{C}}$ bundle $E_{\mathbb{C}}$: one simply takes the $(0,1)$ part of A to define a holomorphic structure on the complexification of E. So any gauge field A on $C \times \mathbb{R}$ determines a 1-parameter family of holomorphic bundles $E_y \to C$, for $y \in \mathbb{R}$. However, if A obeys the Bogomolny equation

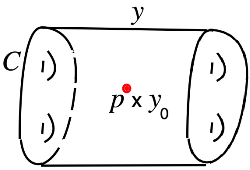
$$F = \star \mathrm{d}_A \phi$$

(for some ϕ) then E_y, as a holomorphic bundle over C, is canonically independent of y. That is true because 2 of the 3 real components of the Bogomolny equation can be put in the form

$$0 = \left[\frac{D}{Dy} + \mathrm{i}\phi, \frac{D}{D\bar{z}} \right].$$

This condition tells us the gauge transformation we have to make to show that the complex structure of E_y is independent of y. (The third Bogomolny equation can be viewed as a moment map condition, similarly to what Atiyah and Bott said in discussing two-dimensional Yang-Mills theory.)

Now suppose that there is an 't Hooft/Dirac singularity at some point $p \times y_0 \in C \times \mathbb{R}$:



Then E_y is still canonically independent of y except at $y = y_0$. And even when we go across $y = y_0$, the bundle E_y is still canonically independent of y if we restrict it to $C \backslash p$. So in other words, what is happening at $y = y_0$ is that the bundle E_y undergoes a Hecke modification at the point p. The "type" of Hecke modification is determined by the particular 't Hooft/Dirac singularity, associated to $\rho : U(1) \to T$. Electric-magnetic duality relates this to some $\rho^\vee : T^\vee \to U(1)$ and thus to a representation of the dual group. This relation between geometric Hecke modifications and representations of the dual group is the same one that is claimed in the geometric Langlands correspondence.

The moduli space of solutions of the Bogomolny equations coincides with the moduli space of Hecke modifications as defined in algebraic geometry; this follows from the “moment map” interpretation of the “third” Bogomolny equation.

In short, I have explained some of the main ideas in the gauge theory interpretation of the geometric Langlands correspondence. I have tried to describe how the story began with the work of Goddard, Nuyts and Olive and then Montonen and Olive in the 1970's. I have said something about how the story unfolded in parallel with other developments in mathematics and physics, and I have explained something about the influence on this process of Michael Atiyah.