PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Magnetic Monopoles in Grand Unified Theories

P. Goddard

Phil. Trans. R. Soc. Lond. A 1982 **304**, 87-95
doi: 10.1098/rsta.1982.0007

Phil. Trans. R. Soc. Lond. A **304**, 87–95 (1982) [87]
Printed in Great Britain

Magnetic monopoles in grand unified theories

By P. Goddard
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, U.K.

Magnetic monopoles occur naturally in grand unified theories as solutions to the classical equations of motion. The occurrence of these objects, some of their properties, and the relation of their existence to charge quantization are reviewed.

1. Introduction

It is an intriguing aspect of grand unified theories that their conventional structure is just such that they possess classical particle-like solutions with the interpretation of magnetic monopoles ('t Hooft 1974; Polyakov 1974). (For more detailed reviews of monopoles in gauge theories see, for example, Coleman (1975*a*) or Goddard & Olive (1978)). However, these monopoles were no part of the 'design specification' of the theories. Indeed they are somewhat of an embarrassment. It seems they would have been produced so copiously in the early Universe that they entail serious difficulties for our understanding of nucleosynthesis and the expansion of the Universe (Preskill 1979; Zeldovich & Khlopov 1979). Of course there have been ingenious attempts to reconcile these difficulties (see, for example, Guth & Tye 1980), but it is not my purpose to give a critical account of these. I shall attempt to review briefly the topological reasons for the occurrence of monopoles and their properties (§ 2), and the bounds that can be set, classically, on the masses of such objects (§ 3), and to give some indication of what has been learnt about such solutions and the tantalizing theoretical possibilities they might offer (§ 4).

Magnetic monopole solutions seem to fit into grand unified theories like a piece of a jigsaw. Although, in the absence of any observation of a monopole, they might seem to be an inconvenience, they are intimately related to one of the frequently cited motivations for unification, namely to provide an explanation for electric charge quantization. Fifty years ago, Dirac (1931) pointed out that to obtain a consistent quantum mechanics for an electrically charged particle in the field of a magnetic monopole, the electric charge would have to be an integral multiple of a constant inversely proportional to the magnetic charge. The electrical charges of the quanta of fields in a gauge theory, possessing magnetic monopole solutions, satisfy this Dirac quantization condition (or the appropriate generalization) relative to the magnetic charges (Corrigan *et al.* 1976; Corrigan & Olive 1976). Obviously, if this had not been so, it would have established a *prima facie* case that the quantum field theory would be inconsistent. Still, it seems to me to be a reflexion of some deep structure that it does work. It is one aspect of the puzzle that this part of it fits together.

2. Topology and monopoles

Some years ago 't Hooft (1974) and Polyakov (1974) pointed out that certain spontaneously broken gauge theories possess static finite-energy solutions carrying non-zero magnetic charge. The Poincaré invariance of the theory means that such a lump of energy can be translated or made to move with any subluminal velocity. Thus it can be thought of as a sort of extended particle. (Such solutions are loosely referred to as solitons, though this does not imply they possess the remarkable and extreme scattering properties sometimes associated with this term.) Being finite rather than point particles, these objects have classically calculable masses, charges and other attributes.

Soliton-like solutions are possessed by all gauge theories based on an original symmetry group G, which is simple or, at least, semi-simple, spontaneously broken by a Higgs mechanism to an 'exact' residual symmetry group H having a local U(1) factor. If this U(1) factor is identified with the electromagnetic gauge group, the solutions carry magnetic charge. The prototype considered by 't Hooft and by Polyakov comprised an SU(2) gauge field interacting with a Higgs field ϕ transforming under the triplet representation. Such a theory is described by a Lagrangian of the form

$$L = -\tfrac{1}{4}G_a^{\mu\nu}G_{a\mu\nu} + \tfrac{1}{2}D^\mu\phi\, D_\mu\phi - V(\phi). \tag{1}$$

A renormalizable SU(2)-symmetric potential giving rise to spontaneous symmetry-breaking has the form

$$V(\phi) = \tfrac{1}{4}\lambda(\phi^2 - a^2)^2, \quad \lambda > 0. \tag{2}$$

[Here $D_\mu \phi = \partial_\mu \phi - eW_\mu \wedge \phi$; $G_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - eW_\mu \wedge W_\nu$.]

Clearly for finite energy $|\phi| \to a$ as $r \to \infty$, i.e. at spatial infinity, so that $V(\phi) \to 0$. Before 1974, it had been generally assumed that the direction of ϕ could be standardized at infinity,

$$\phi^a(\boldsymbol{x}, t) \to a\,\delta_{a3}, \quad \text{as} \quad r = |\boldsymbol{x}| \to \infty, \tag{3}$$

as well as its length, for the freedom to perform local gauge transformations means that in any region of space we can choose a gauge in which the Higgs field is in the third direction. The problem with this argument is that while it is true locally, problems can arise when one tries to make such an alignment smoothly over a closed surface. The reason is essentially topological (Arafune *et al.* 1975).

Let us consider this point in more detail. At a given time and for each fixed value of r, the direction of ϕ will vary smoothly over the (two-dimensional) sphere $x_1^2 + x_2^2 + x_3^2 = r^2$, unless its direction is ambiguous because ϕ vanishes on the sphere. As $|\phi| \to a$ at infinity, it cannot vanish for suitably large spheres. The difficulty is that it is not always possible to rotate this direction into a fixed direction in such a way that the rotation varies continuously over the sphere. A simple example of when this is impossible is that of ϕ radial at every point of the sphere. (Notice that we have implicitly identified directions in the internal space and real space arbitrarily. Different ways of doing this differ by constant overall gauge rotations, which are not significant.) One cannot find a rotation $R \in SO(3)$ that is a continuous function of the unit vector $\hat{\boldsymbol{x}}$, as $\hat{\boldsymbol{x}}$ varies over all the unit sphere S^2 in three-dimensional space, such that

$$R(\hat{\boldsymbol{x}})\,\hat{\boldsymbol{x}} = \hat{\boldsymbol{z}}, \tag{4}$$

$\hat{\boldsymbol{z}}$ being the fixed unit vector pointing in the third direction.

For each fixed r and t, the direction of $\phi(r\hat{\boldsymbol{x}}, t)$ provides a unit vector continuously depending on $\hat{\boldsymbol{x}} \in S^2$ that defines a map $S^2 \to S^2$. As we vary r (and t) this map will itself change continuously. Maps that can be continuously distorted into one another are said to be *homotopic*. Once r is sufficiently large to have got past all the zeros of ϕ, this map is homotopic to its asymptotic form

$$\hat{\phi}_\infty(\hat{\boldsymbol{x}}) = \lim_{r\to\infty} \phi(r\hat{\boldsymbol{x}})/a. \tag{5}$$

(The t-dependence is suppressed for notational simplicity.) It is fairly clear that homotopy is an equivalence relation, so that maps can be divided into mutually exclusive classes consisting of those that are mutually homotopic. In general labelling the equivalence classes (homotopy classes) and classifying maps into them are a classical topological problem. In our simple case of maps $S^2 \to S^2$ it is not difficult. To any such map $f(\hat{x})$ we can attach an integer or 'winding number', the number of time $f(\hat{x})$ covers the unit sphere S^2 as $\hat{x}$ covers it once. This number clearly varies continuously if we continuously deform f; so, being integral, it remains constant. Thus the winding number is the same for homotopic maps. In fact the converse also holds and two maps $S^2 \to S^2$ are homotopic if and only if they have the same winding number. This winding number is a conserved gauge invariant quantity.† Being integral it is quantized and so it may be thought of as a 'quantum number' of the soliton, although we are still discussing the classical theory.

What has it to do with magnetic charge? To answer this we consider how electromagnetism would be embedded in this SU(2) gauge theory. The electromagnetic gauge group is an exact U(1) gauge group. After the symmetry-breaking resulting from the non-zero value of ϕ in the vacuum, the residual gauge symmetry group is that of rotations about the direction of ϕ which is isomorphic to U(1). If we could choose ϕ to have a fixed direction, $\hat{z}$ say, we could identify electromagnetism with the third component of the gauge field. More generally, in the background of a field configuration for which this is not possible, we would identify it with the component of $G_{\mu\nu}$ in the direction of ϕ. When ϕ is in a vacuum configuration, which it is at large distance, so that

$$\mathrm{D}_\mu \phi = 0, \quad |\phi| = a, \tag{6}$$

we would expect to regain conventional electro-dynamics. Indeed if we define the electromagnetic field tensor by

$$F_{\mu\nu} = \phi \cdot G_{\mu\nu}/a, \tag{7}$$

it satisfies Maxwell's equations, including $\nabla \cdot B = 0$, where equations (6) hold. If we put $A_\mu = \phi \cdot W_\mu / a$, then

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + (1/ea^3)\, \phi \cdot (\partial_\nu \phi \wedge \partial_\nu \phi). \tag{8}$$

Equation (8) gives information about the possible magnetic charges of the solution. Defining the magnetic charge as the flux of the magnetic field $B_i = \frac{1}{2}\epsilon_{ijk} F_{jk}$ out of a large sphere, we see that the first two terms in equation (8) $(\partial_\mu A_\nu - \partial_\nu A_\mu)$ make a zero contribution by Stokes's theorem. Thus if we can take ϕ to be asymptotically constant the field configuration has no net magnetic charge. In those configurations with winding number n non-zero, in which this is not possible, it follows from equation (8) that the magnetic charge

$$g = 4\pi n/e. \tag{9}$$

Equation (9) shows that the topological quantum number n and the magnetic charge are, up to a constant coefficient, $\bar{g} = 4\pi/e$, identical. As a result of its topological origin, magnetic charge is quantized at the classical level and conserved independently of the dynamics (equations of motion), merely on the assumption that the fields evolve continuously in time.

The possible magnetic charges $g = 4\pi n/e$, possessed by the soliton solutions, and the possible electric charges $q = \frac{1}{2}n'e\hbar$, possessed by the quanta of fields that could be coupled into the theory,

† Conservation depends on the fact that ϕ evolves continuously in time, while gauge invariance is a result of the fact that any gauge transformation $R(\hat{x})$ on the unit sphere defines a map $S^2 \to \mathrm{SO}(3)$ that *can* be continuously deformed to a constant, for example $R = 1$; all maps $S^2 \to G$, where G is a Lie group, are topologically trivial.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

satisfy the condition that $qg/2\pi\hbar$ is integral. This is the condition that Dirac (1931) found to be necessary for a consistent quantum theory of an electric charge q moving in the given magnetic field of a monopole of strength g. The Dirac quantization condition can be written

$$\exp(\mathrm{i}qg/\hbar) = 1. \tag{19}$$

The particular configuration considered by 't Hooft (1974) and Polyakov (1974) is static and spherically symmetric,

$$\phi^a = [x^a H(aer)]er^2, \tag{11a}$$

$$W_a^i = -\epsilon_{aij}x^j[1-K(aer)]/er^2, \quad W_a^0 = 0, \tag{11b}$$

in that it is symmetric under simultaneous rotations and gauge transformations. With this *ansatz* the equations of motion amount to

$$\xi^2(\mathrm{d}^2K/\mathrm{d}\xi^2) = KH^2 + K(K^2-1), \tag{12a}$$

$$\xi^2(\mathrm{d}^2H/\mathrm{d}\xi^2) = 2K^2H + (\lambda/e^2)H(H^2-\xi^2), \tag{12b}$$

where $\xi = aer$; the boundary conditions for a finite energy solution are $K-1 \leqslant O(\xi)$, $H \leqslant O(\xi)$ as $\xi \to 0$ and $K \to 0$; $H \sim \xi$, sufficiently fast as $\xi \to \infty$. That these equations possess a solution was first indicated by computation but it has also been rigorously established. This spherically symmetric solution possesses unit magnetic charge and a mass bounded below by $4\pi a/e = \bar{g}a$, as we shall see in § 3.

The structure of the monopole is exponentially damped outside a finite region of space. The structure functions of the gauge and Higgs fields, K and H, respectively, approach their asymptotic forms on a length scale given by the Compton wavelengths $M_{\mathrm{W}}/\hbar$ and $M_\phi/\hbar$, respectively, where $M_{\mathrm{W}} = ae\hbar$ and $M_\phi = (2\lambda)^{\frac{1}{2}}a\hbar$ are the masses associated with the massive quanta of the corresponding fields. The monopole radius can be thought of as the larger of $M_{\mathrm{W}}/\hbar$ and $M_\phi/\hbar$. Substantially outside this radius, the massive fields vanish exponentially leaving fields satisfying Maxwell's equations. Viewed on such a scale, the 't Hooft–Polyakov monopole begins to look like the point monopoles considered by Dirac. Even at such a distance we cannot drop the third term on the right-hand side of equation (8) all over a large sphere for the topological reason we have emphasized: ϕ cannot be continuously rotated to a constant direction all over the sphere. If we exclude some part of the sphere, however small the solid angle it subtends at the centre, we can do it over the rest. Then $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ with the exclusion of the small region. The union of these small regions as the sphere radius is increased forms the Dirac string attached to the point (on this scale) monopole.

Much work has been done in studying generalizations of the case originally considered by 't Hooft and Polyakov, both in the direction of larger gauge groups G and higher magnetic-charge SU(2) solutions. The known solutions obey a generalized inverse square law (Englert & Windey 1976; Goddard *et al.* 1977) for the magnetic part of the gauge field tensor:

$$G_{ij} \sim (1/4\pi r^2)\,\epsilon_{ijk}\hat{x}_k\,\mathscr{G}(\hat{x}) \quad \text{as} \quad r \to \infty, \tag{13}$$

where $\mathscr{G}$ is covariantly constant:

$$\mathrm{D}_i\mathscr{G} = 0. \tag{14}$$

(Here $G_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu + \mathrm{i}e[W_\mu, W_\nu]$ and $W_\mu = W_\mu^a T^a$ where the $\{T^a\}$ are a set of Hermitian generators for the gauge group G.) For these solutions the Dirac quantization condition (equation (10)) generalizes to

$$\exp(\mathrm{i}e\mathscr{G}) = 1. \tag{15}$$

The theories we are considering will, in general, contain a number of Higgs fields $\phi, \psi, \ldots$, transforming according to different irreducible representations of G, all of which are asymptotically covariantly constant: $D_\mu \phi = 0$, $D_\mu \psi = 0, \ldots$. As in the prototype of 't Hooft and Polyakov, it may not be possible to take all or any of the Higgs fields to be asymptotically constant. They spontaneously break G down to their little group $H = \{g \in G\colon g\phi = \phi, g\psi = \psi, \ldots\}$. If this H has a U(1) subgroup that is a factor, at least locally, we can deduce from equation (15) a condition on the 'electric' and 'magnetic' charges associated with this subgroup (which might have the physical interpretation of electromagnetism or weak hypercharge, depending on the context).

Since $[D_i, D_j] = ieG_{ij}$, the magnetic field G_{ij}, and so $\mathscr{G}$, are generators of H. If H has a local U(1) factor with generator T, and

$$\mathscr{G} = gT + \mathscr{G}^\perp, \tag{16}$$

where $\mathscr{G}^\perp$ is orthogonal to T, we obtain (Corrigan & Olive 1976)

$$\exp(\mathrm{i}Qg/\hbar) = k, \tag{17}$$

where $Q = eT\hbar$ is the 'electric' charge operator associated with the U(1) subgroup and k is an element of the subgroup K of H generated by those generators of H orthogonal to T. Thus $k \in Z = \mathrm{U}(1) \cap K$ and H is isomorphic to $\{\mathrm{U}(1) \times K\}/Z$. If K is semi-simple, it must be compact, as G is, and Z is forced to be finite. In this case equation (17) leads to the quantization condition that $qg|Z|/2\pi\hbar$ has to be integral where q is any eigenvalue of Q, and $|Z|$ is the number of elements in Z.

In the simpler examples of grand unified theories the original simple gauge group G is broken by a pair of Higgs fields ϕ, ψ to an exact symmetry group H. If $\mathscr{M}_0$ denotes the set of values in (ϕ, ψ)-space at which the combined Higgs potential attains its minimum value, zero, the topological considerations applied to the prototype generalize to show that the homotopy class of the map $S^2 \to \mathscr{M}_0$, defined by the asymptotic values of the Higgs fields, is a conserved quantity. We shall assume that G acts transitively on $\mathscr{M}_0$ (i.e. any two points of it are related by a gauge transformation) so that all vacuum values of the Higgs fields are physically equivalent. Then, if G has been taken to be simply connected, there is a well known theorem that the homotopy classes of maps $S^2 \to \mathscr{M}_0$ are in one : one correspondence with those of maps from S^1, the circle, to H; the collection of the former classes is denoted by $\Pi_2(\mathscr{M}_0)$ and the latter by $\Pi_1(H)$. When H has the local structure of $\mathrm{U}(1) \times K$ with K semi-simple, $\Pi_1(H)$ is labelled by the integers and there is just one topological quantum number.

Let us consider the specific case of the SU(5) theory (Georgi & Glashow 1974). This is broken by a Higgs field ϕ in the 24-dimensional adjoint representation to a group with the local structure of $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$, and then further broken by a Higgs field ψ in the five-dimensional fundamental representation to $\mathrm{SU}(3) \times \mathrm{U}(1)$. After the first breaking the $\mathrm{SU}(2) \times \mathrm{U}(1)$ is the gauge group of the Salam–Weinberg model and after the second the U(1) is the gauge group of electromagnetism. The ϕ-field is responsible for the masses of the superheavy gauge bosons, and the ψ-field for the masses of weak interaction W and Z boson masses and the fermion masses. In this model all the topological structure is associated with the ϕ-field. The breaking it causes introduces the U(1) factor with which the magnetic quantum numbers are associated; it is not difficult to see that, as far as its topological properties are concerned, the asymptotic map defined by the Higgs field might as well be projected on its ϕ-value alone. We shall consider this model in more detail in the next section.

3. Classical bounds on masses

Let us consider again the prototype. 't Hooft (1974) estimated the mass of the monopole to be $4\pi a/e$ multiplied by a function of λ/e^2 which varied from just over 1 at $\lambda/e^2 = 0.1$ to about 1.44 at $\lambda/e^2 = 10$. In fact $4\pi a/e = ag$ provides a rigorous lower bound for the mass (Bogomol'nyi 1976; Coleman *et al.* 1977) as its energy is

$$\frac{1}{2}\int \mathrm{d}^3x\{\mathscr{E}^2+\mathscr{B}^2+(\mathrm{D}_i\,\phi)^2+(\mathrm{D}_0\,\phi)^2+2V(\phi)\}, \tag{18}$$

where $\mathscr{E}_i = G_{0i}$ and $\mathscr{B}_i = \frac{1}{2}\epsilon_{ijk}\,G_{jk}$. This clearly exceeds

$$\frac{1}{2}\int \mathrm{d}^3x[\mathscr{B}^2+(\mathrm{D}_i\,\phi)^2] \geqslant \pm\int \mathrm{d}^3x(\mathscr{B}_i\cdot\mathrm{D}_i\,\phi) = \pm\int \mathrm{d}S_i(\phi\cdot\mathscr{B}_i) \tag{19}$$

taken over a large sphere, as $\mathrm{D}_i\cdot\mathscr{B}_i = 0$ by the Bianchi identities. But this last expression differs from the magnetic charge only by a factor of a, proving that the energy of any solution is bounded below by $a|g|$. This implies that the ratio of the mass of the monopole to that of a heavy gauge boson

$$M_{\mathrm{m}}/M_W \geqslant 4\pi/e^2 = 1/\alpha, \tag{20}$$

where α denotes the fine-structure constant, before any quantum corrections have been taken into account.

To saturate this lower bound, often called the Bogomol'nyi bound, some severe conditions have to be satisfied. Clearly we must have

$$\mathscr{E} = \mathrm{D}_0\,\phi = V(\phi) = 0 \quad \text{and} \quad \mathscr{B}_i = \pm\,\mathrm{D}_i\,\phi.$$

Hence saturation is only possible in the Bogomol'nyi–Prasad–Sommerfield (B.P.S.) limit of vanishing Giggs potential (Bogomol'nyi 1976; Prasad & Sommerfield 1975). In this limit it is possible to establish many exact and elegant results about the theory.

In generalizations to grand unified theories one would expect qualitatively similar results for monopole masses. The details of the derivation of the Bogomol'nyi bound for such theories have been considered by Scott (1980*a*, *b*) who obtained

$$M_{\mathrm{m}}/ \geqslant M_{\mathrm{W}}^{\phi}\,\tfrac{3}{8}\alpha \tag{21}$$

in the SU(5) theory and in different versions of the SO(10) model. Here M_{W}^{ϕ} denotes the mass of a 'superheavy' gauge boson which acquires its mass when the adjoint representation Higgs field breaks G to a group containing $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$, locally. It is believed that $M_{\mathrm{W}}^{\phi} \geqslant 10^{14}$ GeV so that the magnetic monopoles have masses of the order of 10^{16} GeV.

Let us consider the derivation of the inequality (21). In any theory with an adjoint-representation Higgs field, we can derive an analogue of (19), for the mass of a monopole,

$$M_{\mathrm{m}} \geqslant \pm\int \mathrm{d}S_i\,\phi^a\mathscr{B}_i^a, \tag{22}$$

ignoring the effect of the other Higgs field ψ which makes a non-negative contribution to the mass. Taken alone, the Higgs field ϕ breaks G to the little group of ϕ that is generated by those generators of G that commute with $\phi^a T^a$, the direction in the Lie algebra of G picked out by ϕ. Now $\phi^a T^a$ is itself a generator of this group and so generates a local U(1) factor of the little group, since it commutes with all the generators of the little group by construction. This preferred generator does not have the interpretation of charge but, in the SU(5) model, of weak hypercharge; so let us denote its direction by $\hat{Y}$:

$$\phi^a = a\hat{Y}^a, \quad (\hat{Y}^a)^2 = 1, \tag{23}$$

a denoting the asymptotic length of the Higgs field as usual. Hence the inequality (22) becomes

$$M_{\mathrm{m}} \geqslant a|g_y| \tag{24}$$

where g_y denotes the weak magnetic hypercharge of the monopole, defined as in equation (16), by

$$\mathscr{B} = g_y \hat{Y} + \mathscr{B}^{\perp} \tag{25}$$

where $\mathscr{B}^{\perp}$ is perpendicular to Y. It remains to determine the smallest non-zero value of $|g_y|$.

If ϕ breaks G to a group having the local structure of $\mathrm{U}(1) \times K$, we have seen that equation (17) implies the quantization condition that $eg_y y|Z|/2\pi$ be integral where y is any eigenvalue of $\hat{Y} = \hat{Y}^a T^a$, e is the coupling constant for the grand unified group G, and $|Z|$ is the number of elements of $Z = \mathrm{U}(1) \cap K$. This is a subgroup of the centre $Z(K)$ of K. In any representation y has to be an integral multiple of some number y_0. Further $|Z|$ must divide $|Z(K)|$ so that it is of the form $|Z(K)|/N$ for some integer N which in fact divides both $|Z(K)|$ and $|Z(G)|$. Hence $eg_y(Ny_0)|Z(K)|/2\pi$ is integral. Now in the adjoint representation, y has to be a multiple of $\bar{y}$ that can be shown to equal $Ny_0/|Z(G)|$. Thus we expect the lowest value of g_y, $\bar{g}_y$ to satisfy (Goddard & Olive 1981 *a*)

$$e\bar{y}\bar{g}_y = 2\pi|Z(G)|/|Z(K)|. \tag{26}$$

Now the masses of the vector mesons produced by the vacuum value of ϕ are $M_{\mathrm{W}}^{\phi} = ae\bar{y}\hbar$. Hence

$$M_{\mathrm{m}}/M_{\mathrm{W}}^{\phi} \geqslant 2\pi|Z(G)|/|Z(K)|\,e^2\bar{y}^2\hbar. \tag{27}$$

Let us consider specifically $G = \mathrm{SU}(5)$, $K = \mathrm{SU}(3) \times \mathrm{SU}(2)$. Then $|Z(G)| = 5$, $|Z(K)| = 6$, forcing $N = 1$. In the five-dimensional fundamental representation of G, we can take $\hat{Y}$ to be the diagonal matrix with entries $\sqrt{\frac{3}{5}}(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, \frac{1}{2}, \frac{1}{2})$, normalizing so that $\mathrm{tr}\,(T^aT^b) = \frac{1}{2}\delta_{ab}$ in this representation. Then $y_0 = \frac{1}{6}\sqrt{\frac{3}{5}}$ and $\bar{y} = \frac{5}{6}\sqrt{\frac{3}{5}}$ in agreement with our general considerations. Thus

$$M_{\mathrm{m}}/M_{\mathrm{W}}^{\phi} \geqslant 4\pi/e^2\hbar. \tag{28}$$

This last expression is not just $1/\alpha$ because e is the grand unified coupling constant rather than the electromagnetic one, e_0. Comparing $\mathrm{D}_\mu = \partial_\mu + \mathrm{i}eW_\mu^a T^a$ with the electromagnetic covariant derivative $\partial_\mu + \mathrm{i}QA_\mu/\hbar$ where Q has diagonal entries $e_0\hbar(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, 1, 0)$ in the fundamental representation, we have $A_\mu = 2\,\mathrm{tr}(Q\cdot W_\mu^a T^a)/[2\,\mathrm{tr}(Q^2)]^{\frac{1}{2}}$ and $e\hbar\,\mathrm{tr}(Q, W_\mu) = \mathrm{tr}(Q^2)\,A_\mu$, so that $2\,\mathrm{tr}(Q^2) = e^2\hbar^2$. Hence $e_0^2 = \frac{3}{8}e^2 (= e^2\sin^2\theta_W$ where θ_W is the Weinberg angle) and we obtain the inequality (21).

Comments

Electric charge quantization is one of the most mysterious precise results in nature. The two principal theoretical explanations of this, that through the existence of magnetic monopoles and that following from grand unification are seen to be intimately related in a way that is probably not fully understood. In the unification argument electric charge is quantized because it is the generator of a compact U(1) local factor of the exact symmetry group H remaining after all the spontaneous symmetry-breaking has taken place. Olive has emphasized that this happens automatically if H has the local structure of $\mathrm{U}(1) \times K$, and both the colour group K and the original symmetry group G are semi-simple. There are elementary counter-examples where charge quantization might fail if H has more than one U(1) local factor. It is also precisely this situation in which the quantization condition (17) leads to quantization of electric charge through the presence of magnetic monopoles (Olive 1980). So the arguments come together in some apparently preordained way.

Since the papers of 't Hooft and Polyakov much has been learnt about monopoles in gauge field theories and their properties. In particular, in the B.P.S. limit of vanishing Higgs potential, exact solutions (Prasad & Sommerfield 1975; Taubes 1981) exist to equations (21), which are closely related to the self-duality equations satisfied by instantons. Exact spherically symmetric solutions have been constructed for more general gauge groups G (Wilkinson & Bais 1979), and static solutions with magnetic charge equal to any multiple of $4\pi/e$ have been constructed for SU(2). (Ward 1981 *a*, *b*; Forgacs *et al.* 1981; Prasad & Rossi 1981; Corrigan & Goddard 1981.) Physically this is a peculiar limit. The only vestige of the Higgs potential is the asymptotic boundary condition forcing the Higgs fields to take vacuum values, in $\mathscr{M}_0$, at spatial infinity. The long-range structure is not that of a point monopole as the mass associated with the Higgs field M_ϕ has vanished; in this case the size of a B.P.S. monopole is infinite. This produces a long-range attractive force that can exactly compensate for the Coulombic repulsion between monopoles of like charge (Manton 1977), enabling them to be balanced exactly to form static multimonopole solutions. The physical relevance of these solutions may be difficult to assess, though they should tell us much about the non-Abelian magnetic quantum numbers of monopoles. The history of non-Abelian gauge theories is dominated by elegant developments whose significance was not completely obvious, or was not widely recognized for some time. So it would seem worthwhile to learn more about these classical solutions.

In theories in one space and one time dimension, soliton solutions have lead to extremely interesting structures. For example in the famous case of the sine–Gordon and Thirring models one has distinct formulations of equivalent theories, with the soliton-like states of the former being the quanta of the fundamental field in the latter (Skyrme 1961; Coleman 1975 *a*, *b*). The dual way in which electric and magnetic charge appear in the Dirac conditions and generalizations has lead to speculations that there might be dual formulations of spontaneous broken gauge theories, with the solitons carrying magnetic charge and the quanta of the fields carrying electric charge in one formulation, and with these roles reversed in the other (Goddard *et al.* 1977; Montonen & Olive 1977; Bais 1978; Goddard & Olive 1981 *b*). Whatever the fate of these ideas, there seems to be much of interest to be learnt about monopole-like solutions to gauge field theories.

I am very grateful to Edward Corrigan and David Olive for many stimulating discussions on this subject over the last five years.

References (Goddard)

Arafune, J., Freund, P. G. O & Goebel, C. J. 1975 *J. math. Phys.* **16**, 433.
Bais, F. A. 1978 *Phys. Rev.* D **18**, 1206.
Bogomol'nyi, E. B. 1976 *Soviet J. nucl. Phys.* **24**, 449.
Coleman, S. 1975 *a* *Phys. Rev.* D **11**, 2088.
Coleman, S. 1975 *b* *New phenomena in subnuclear physics* (Proc. 1975 Int. School of Physics 'Ettore Majorana') (ed. A. Zichichi), p. 297. New York: Plenum.
Coleman, S., Parke, S., Neveu, A. & Sommerfield, C. M. 1977 *Phys. Rev.* D **15**, 554.
Corrigan, E. & Goddard, P. 1981 *Communs math. Phys.* **80**, 575.

Corrigan, E. & Olive, D. 1976 *Nucl. Phys.* B **110**, 237.
Corrigan, E., Olive, D., Fairlie, D. B. & Nuyts, J. 1976 *Nucl. Phys.* B **106**, 475.
Dirac, P. A. M. 1931 *Proc. R. Soc. Lond.* A **133**, 60.
Englert, F. & Windey, P. 1976 *Phys. Rev.* D **14**, 2728.
Forgacs, P., Horvath, Z. & Palla, L. 1981 *Physics Lett.* B **99**, 232.
Georgi, H. & Glashow, S. L. 1974 *Phys. Rev. Lett.* **32**, 438.
Goddard, P., Nuyts, J. & Olive, D. 1977 *Nucl. Phys*, B **125**, 1.
Goddard, P. & Olive, D. 1978 *Rep. Prog. Phys.* **41**, 1357.
Goddard, P. & Olive, D. 1981 *a* *Nucl. Phys.* B (To be published.)
Goddard, P. & Olive, D. 1981 *b* *Nucl. Phys.* B (To be published.)
Guth, A. & Tye, S. H. 1980 *Phys. Rev. Lett.* **44**, 631.
't Hooft, G. 1974 *Nucl. Phys.* B **79**, 276.
Manton, N. S. 1977 *Nucl. Phys.* B **126**, 525.
Montonen, C. & Olive, D. 1977 *Phys Lett.* B **72**, 117.
Olive, D. 1980 *Unification of the fundamental particle interactions* (ed. S. Ferrara *et al.*) p. 451. New York: Plenum.
Polyakov, A. M. 1974 *JETP Lett.* **20**, 194.
Prasad, M. K. & Rossi, P. 1981 *Phys. Rev. Lett.* **46**, 806.
Prasad, M. K. & Sommerfield, C. M. 1975 *Phys. Rev. Lett.* **35**, 760.
Preskill, J. P. 1979 *Phys. Rev. Lett.* **43**, 1365.
Scott, D. M. 1980 *a* *Nucl. Phys.* B **171**, 95.
Scott, D. M. 1980 *b* *Nucl. Phys.* B **171**, 109.
Skyrme, T. H. R. 1961 *Proc. R. Soc. Lond.* A **262**, 237.
Taubes, C. 1981 *Communs math. Phys.* **80**, 343.
Ward, R. S. 1981 *a* *Communs math. Phys.* **79**, 317.
Ward, R. S. 1981 *b* *Physics Lett.* B **102**, 136.
Wilkinson, D. & Bais, F. A. 1979 *Phys. Rev.* D **19**, 2410.
Zeldovich, Ya. B. & Khlopov, M. Yu. 1979 *Physics Lett.* B **79**, 239.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES