Is there a theorem for the quantization of magnetic charge?

Yu. D. Usachev

P. N. Lebedev Physics Institute, Moscow
Fiz. Él. Chast. Atom. Yad., 4, 225-238 (January-March 1973)

The theorem for the quantization of magnetic charge is analyzed. The difficulties which arise in various proofs of this theorem are pointed out. Arguments against a theorem for quantization of magnetic charge are presented, and the consequences of this position are analyzed relative to experimental attempts to find monopoles.

INTRODUCTION

More than forty years have passed since Dirac[1] first suggested the existence of an isolated magnetic pole (monopole). Since that time a great many studies, both theoretical and experimental, have been devoted to the magnetic charge problem, and the question is still far from being resolved. There has been no experimental observation of magnetic charge, nor any definitive theoretical proof that it cannot exist.

During the past forty years there have been no significant changes in the physical form of the magnetic charge in any of the theoretical studies. Experiment, in turn, has experienced little progress which could result in other than pessimistic forecasts. The regular failure to observe Dirac monopoles in newly completed accelerators has produced a twofold result. First, it has shifted the lower limit on the monopole mass, and second, the upper limit on the cross section for producing monopoles has been decreased. Experiments conducted on monopoles of cosmic origin have removed the lower limit on the flux, reducing it to a very low value. It is now difficult to decide whether the available experimental data prove the absence of such monopoles in nature, or whether the data should stimulate higher precision in new experimental studies.

The main prediction of the Dirac theory relative to monopole properties is that it must have a very large magnetic charge g (g = 68.5e), and therefore a high ionizing power. If such a particle were to appear in a free state, it could not remain undetected by experiment. Nevertheless, as we have already mentioned, experiment quickly leads to pessimistic predictions.

An analysis of the present theoretical and experimental positions prompts at least three possible explanations of the failure to observe magnetic monopoles.

1. The monopole does not exist in nature. This point of view is impossible to accept for at least two reasons. First, an experimental "proof" of the absence of monopoles is practically impossible since there is always the possibility of more precise experiments in the future and, second, present theory places no restrictions on the existence of magnetic charge.

2. Because of a very large coupling constant the monopole does not appear in an isolated state; rather, it always occurs as a monopole–antimonopole pair, which subsequently decays. The newest experimental studies are pursuing monopole observation through the various products resulting from this decay; this would be an indirect proof of the existence of magnetic charge.

3. The actual physical form of the monopole does not correspond to that predicted by Dirac theory. The first difference which comes to mind is that the magnetic charge may be different. If it is assumed that this charge is much smaller than predicted by Dirac theory[1] (g = 68.5e), or by Schwinger[2] (g = 137e), it is clear that the proof of this assumption would entail a real revolution in the experimental methods used to detect the magnetic charge. In essence this would mean that we are not looking for the magnetic charge in the right places.

The third explanation for the "nonobservation" of the monopole is very radical, and suggests a breakdown of the Dirac or Schwinger theorems for the magnitude of the magnetic charge.

According to Dirac we have

$$eg/\hbar c = k/2 \qquad (k = 0,\ \pm 1,\ \pm 2\ \ldots); \tag{1}$$

according to Schwinger,

$$eg/\hbar c = k \qquad (k = 0,\ \pm 1,\ \pm 2\ \ldots). \tag{2}$$

The present paper is devoted to a detailed discussion of this third possibility.

The proofs of Theorems (1) and (2) fall into three categories: 1) proofs based on the requirement that the wave functions be single-valued; 2) proofs obtained through group-theoretical considerations; 3) the Saha–Wilson proof,[3] which is somewhat different from the other two.

We will consider all three types of proofs, and will present arguments which we think enable us to give up Theorems (1) and (2). In any case, we will point out some serious difficulties which arise in proofs of these theorems.[4,5]

1. CLASSICAL THEORY OF THE MAGNETIC CHARGE

There are two initial premises in the theory of magnetic charge.

1. A stationary magnetic charge is the source of a Coulomb magnetic field:

$$\mathbf{H} = g\mathbf{r}/r^3 = -g\,\mathrm{grad}\,(1/r), \tag{3}$$

where g is the magnetic charge. Attempts to completely symmetrize the properties of electrical and magnetic charges can naturally use Eq. (3).

2. A nonrelativistic electron moving in the field of a stationary magnetic charge will experience a Lorentz force:

$$d\mathbf{p}/dt = (e/c)\,[\mathbf{v},\ \mathbf{H}] = -eg\,[\mathbf{v}, \mathrm{grad}\,(1/r)]/c. \tag{4}$$

This equation reflects the fact that the electron is "indifferent" to the source of the magnetic field in which it moves, either one set up by the usual electron currents and described by the usual vector potential A_μ, or a field due to magnetic monopoles, Eq. (3). In both instances the Lorentz force, expressed in terms of H, will have the form of Eq. (4).

We hasten to point out that Eqs. (3) and (4) are postulates of the theory. They are reasonable to the degree to which we wish to preserve the "equality" of electric and magnetic charges.

In the following discussion we shall adhere closely to the logic of Dirac's first paper,[1] in which one of the particles (either the electric or the magnetic particle) is assumed at rest. The relativistic generalization, which Dirac carried out in a second paper,[1] does not change the essential features of the argument, but it greatly complicates the calculations. For our purposes, which deal with the physical difficulties in the theory of the Dirac monopole, it is sufficient to treat just the case in which a monopole is at rest at the origin; therefore we shall not require a relativistic generalization of Eqs. (3) and (4), although such generalizations are easily obtained.

Without the monopoles one can obtain the electrodynamics by varying the Lagrangian:

$$\mathcal{L} = \mathcal{L}_f + \mathcal{L}_i,$$

where $\mathcal{L}_f$ is the Lagrangian of the free motion and $\mathcal{L}_i$ is the interaction Lagrangian ("minimal" interaction):

$$\mathcal{L}_i = j_\mu A_\mu. \qquad (5)$$

It is then easily seen that the equations of motions for the particles, as obtained through the usual variational procedure, inevitably have the form

$$\frac{d\mathbf{p}}{dt} = e\left\{-\operatorname{grad} A_0 - \frac{1}{c}\cdot\frac{\partial \overline{\mathbf{A}}}{\partial t} + \frac{1}{c}[\mathbf{v}, \operatorname{curl} \mathbf{A}]\right\}. \qquad (6)$$

As usual, by identifying

$$\mathbf{E} = -\operatorname{grad} A_0 - \frac{1}{c}\cdot\frac{\partial \mathbf{A}}{\partial t}; \qquad (7a)$$

$$H = \operatorname{curl} \mathbf{A}, \qquad (7b)$$

we obtain the usual electrodynamics without monopoles, because (7b) gives

$$\operatorname{div} \mathbf{H} = \operatorname{div} \operatorname{curl} \mathbf{A} \equiv 0, \qquad (8)$$

while according to Eq. (3) we find that with monopoles

$$\operatorname{div} \mathbf{H} = 4\pi g \delta(\mathbf{r}). \qquad (9)$$

The question then arises, how can we satisfy Eq. (9) and still keep the interaction minimal [i.e., preserve Eq. (5)]?

Dirac solved this problem by giving the vector potential a special form (in the following, this special vector potential will be denoted by B):

$$\mathbf{B} = \frac{g}{r}\cdot\frac{[\mathbf{n}, \mathbf{r}]}{r - \mathbf{nr}}. \qquad (10)$$

In this equation n is a unit vector (unobservable, according to Dirac) with arbitrary direction, and it has no effect on the physical consequences of the theory.

Upon computing curl B using the usual differentiation rules, we find

$$\operatorname{curl} \mathbf{B} = -g\mathbf{r}/r^3 = g \operatorname{grad}(1/r). \qquad (11)$$

It is quite evident that this equality is absurd, as is verified by direct application of the divergence operator to both sides of the equation. The error evidently entered because we did not take into account the singularity in B along the direction specified by the vector n (along the "string" as Dirac calls it). Thus, to correctly calculate B at points on the string and infinitely close to it, we must use Stokes theorem for an infinitely small contour which encloses the singularity:

$$\mathbf{N} \operatorname{curl} \mathbf{B} = \lim_{\Delta S\to 0} \oint_L \mathbf{B}\cdot d\mathbf{r}/\Delta S, \qquad (12)$$

N is the unit vector normal to the infinitely small surface ΔS, which is bounded by the contour L. Carrying out these operations in the case where n is directed along the positive z axis ($\mathbf{n} \equiv \mathbf{k}$) we find, in contrast to Eq. (11) (see ref. 6), that

$$\mathbf{H} = \operatorname{curl} \mathbf{B} = g \operatorname{grad}(1/r) + \mathbf{k}\cdot 4\pi g\theta(z)\,\delta(x)\,\delta(y). \qquad (13)$$

In the general case[5] we have

$$\mathbf{H} = \operatorname{curl} \mathbf{B} = g \operatorname{grad}(1/r) + \mathbf{n} 8g\theta(\mathbf{nr})\,\delta[r^2 - (\mathbf{nr})^2]. \qquad (14)$$

As is clear from Eq. (14) the magnetic field is divided into two parts: the Coulomb field $g\cdot\operatorname{grad}(1/r)$ and the singular field on the string. The total flux of such a field through a closed surface surrounding the origin is zero:

$$\operatorname{div} \mathbf{H} \equiv 0, \qquad (15)$$

because the flux through the closed surface due to the Coulomb magnetic field is completely compensated by the flux of the singular field supplied through the string.

The equations of motion obtained with $\mathcal{L}_i = \mathbf{jB}$ by using the variational principle, where B is defined by Eq. (10), take the form

$$dp/dt = e/c\,[\mathbf{v}, \operatorname{curl} \mathbf{B}] = eg/c\,[\mathbf{v}, \{\operatorname{grad}(1/r) + 8\mathbf{n}\theta(\mathbf{nr})\,\delta(r^2 - (\mathbf{nr})^2)\}]. \qquad (16)$$

If the first term accurately corresponds to the Coulomb character of the magnetic field sought, then the second term will, first, destroy the original postulate (4) of the theory and, second, lead to nonphysical δ-function impulses in the Lorentz force.

The subsequent stages in the development of a theory for Dirac monopoles consist of attempts to obtain a purely Coulomb nature for H for the monopoles, while simultaneously removing the nonphysical terms in the Lorentz force. Dirac accomplished this within a quantum mechanical framework.

2. QUANTUM THEORY OF DIRAC MONOPOLES

Following Dirac, we introduce a special principle (Dirac's veto), which forbids the electron from falling on the string, i.e., the electron wave function is set equal to zero on the string:

$$\psi=0 \ \text{(on the string)}. \tag{17}$$

If electrons do not fall on the string, so that they do not sense the magnetic flux conducted along it to the origin, then we have

$$\operatorname{div}\mathbf{H}=g\operatorname{div}\operatorname{grad}(1/r)=-4\pi g\delta(\mathbf{r}) \tag{18}$$

in contrast to Eq. (15). Thus, the magnetic flux through a closed surface, composed of a sphere and an infinitely thin cylinder cut out of the sphere and running along the string, is now nonzero.

We now see that the condition in Eq. (17) "fixes" the Lorentz force at which the δ-function pulses disappear when Eq. (17) is fulfilled.

Let us now introduce Dirac's relation (1). We shall use the requirement of gradient invariance. Assume that the state of an electron scattered by a monopole at rest at the origin is described by a Schrödinger equation in which the interaction Hamiltonian contains the vector potential of Eq. (10). We replace the vector potential of Eq. (10), which depends on the unit vector $\mathbf{n}_1$:

$$\mathbf{B}(\mathbf{n}_1)=\frac{g}{r}\cdot\frac{[\mathbf{n}_1,\mathbf{r}]}{r-\mathbf{n}_1\mathbf{r}},$$

by a potential $\mathbf{B}(\mathbf{n}_2)$, which depends in the same way on another unit vector $\mathbf{n}_2$ ($\mathbf{n}_2\neq\mathbf{n}_1$) and which does not alter the physical consequences of a gradient-invariant theory if

$$\mathbf{B}(\mathbf{n}_1)-\mathbf{B}(\mathbf{n}_2)=g\operatorname{grad}\chi. \tag{19}$$

Then the wave function of the scattered electron must be altered by a phase factor:

$$\psi(\mathbf{B}(\mathbf{n}_1))=\exp\left(-\frac{ieg}{\hbar c}\chi\right)\psi(\mathbf{B}(\mathbf{n}_2)). \tag{20}$$

Equation (19) is used to calculate the magnitude of $g\chi$:

$$g[\chi(\mathbf{r})-\chi(\mathbf{r}_0)]=\int_{(L)}^{\mathbf{r}}d\mathbf{r}'\,[\mathbf{B}(\mathbf{n}_1)-\mathbf{B}(\mathbf{n}_2)]. \tag{21}$$

The right side contains a curvilinear integral over some path L from the initial point r_0 to the final point at r. Equation (21) is valid, of course, only when Eq. (19) is valid. We shall discuss this validity below.

We now ask, what will be the phase difference χ when the path L is closed, i.e., when the points r and $\mathbf{r}_0$ are the same? In other words we wish to evaluate the integral

$$g\Delta\chi=g[\chi_1(\mathbf{r})-\chi_2(\mathbf{r})]=\oint d\mathbf{r}'\,[\mathbf{B}(\mathbf{n}_1)-\mathbf{B}(\mathbf{n}_2)], \tag{22}$$

where $\chi_1(\mathbf{r})$ and $\chi_2(\mathbf{r})$ are two values of the wave function phase χ at a single point in space after going around some closed loop and returning to the starting point.

We use Stokes' Theorem and Eq. (14):

$$\oint_L d\mathbf{r}'\,[\mathbf{B}(\mathbf{n}_1)-\mathbf{B}(\mathbf{n}_2)]=\int_S d\mathbf{S}\operatorname{curl}[\mathbf{B}(\mathbf{n}_1)-\mathbf{B}(\mathbf{n}_2)]$$

$$=8g\int_S[\mathbf{n}_1\theta(\mathbf{n}_1\mathbf{r})\,\delta(r^2-(\mathbf{n}_1\mathbf{r})^2)-\mathbf{n}_2\theta(\mathbf{n}_2\mathbf{r})\,\delta(r^2-(\mathbf{n}_2\mathbf{r})^2)]\,d\mathbf{S}. \tag{23}$$

Here L is the closed contour and S is the area enclosed by this contour. It is clear from Eq. (23) that if S is not penetrated by either of the two strings $\mathbf{n}_1$ or $\mathbf{n}_2$, the phase difference is $\Delta\chi=0$. In the opposite case one can show that

$$|\Delta\chi|=4\pi. \tag{24}$$

One could also consider this in a different light. If we follow Dirac and exclude from consideration the region occupied by the strings, then Eq. (19) would hold everywhere. But now χ is a single-valued potential in a doubly connected region. A calculation of the phase difference in going around a closed path enclosing a string gives Eq. (24). Mathematically, the two treatments are completely equivalent.

Dirac made the natural assumption that the wave-function phase difference arising when a closed contour is traveled is always some multiple of 2π. This means that traveling a closed path to the original point of departure changes the wave function by the insignificant factor $\exp(-i2\pi k)\equiv 1$. Therefore, for $|\Delta\chi|=4\pi$ we have

$$\exp(-ieg4\pi/\hbar c)=\exp(-i2\pi k) \tag{25}$$

or

$$eg/\hbar c=k/2 \qquad (k=0,\pm1,\pm2\ldots).$$

The absolute value of k depends on the number of circuits one makes around the loop, and its sign depends on the direction of travel.

We now summarize the main postulates of Dirac's theory.

1. The vector potential $\mathbf{B}(\mathbf{n})$ of Eq. (10), which is singular on the string, is used to describe the motion of an electron in the field of a monopole. The use of Eq. (10) in the interaction Lagrangian $\mathcal{L}_i=\mathbf{jB}(\mathbf{n})$, together with the usual variational procedure, leads to the correct expression for the Lorentz force everywhere except on the string.

2. In order to obtain a purely Coulomb magnetic field B and get rid of some nonphysical terms in the Lorentz force, the electron wave function ψ is set equal to zero on the string.

3. The string, which is determined by the unit vector

$\mathbf{n}$ in Eq. (10), is not observable because it does not admit of any physical interpretation.

We will examine these three postulates from the point of view of their physical compatibility.

3. CRITIQUE OF THE DIRAC THEORY

On the basis of Eq. (14), we have

$$\mathrm{curl}\,[\mathbf{B}(n_1)-\mathbf{B}(n_2)] = 8g\,[\mathbf{n}_1\theta(\mathbf{n}_1\mathbf{r})\,\delta(r^2-(\mathbf{n}_1\mathbf{r})^2)-\mathbf{n}_2\theta(\mathbf{n}_2\mathbf{r})\,\delta(r^2-(\mathbf{n}_2\mathbf{r})^2)], \quad (26)$$

which contradicts the original assumption (19) for derivation of Eq. (1), because curl grad $\chi \equiv 0$. In Eq. (26) we see that curl$[\mathbf{B}(\mathbf{n}_1) - \mathbf{B}(\mathbf{n}_2)]$ is not zero on the strings. At first glance it appears that this incompatibility with Eq. (19) is of no importance because the difference occurs just in a one-dimensional region. However, a more detailed analysis shows that this is not the case.

Let us first examine this question from a physical point of view. We ask, what does the gradient transformation of Eq. (20) imply for the Schrödinger equation? By changing the vector potential $\mathbf{B}(\mathbf{n}_1)$ in the Schrödinger equation to $\mathbf{B}(\mathbf{n}_2)$, we see that in the first case the observed magnetic field is determined by the equation $\mathbf{H}_1$ = curl · $\mathbf{B}(\mathbf{n}_1)$ while the field is $\mathbf{H}_2$ = curl $\mathbf{B}(\mathbf{n}_2)$. If $\mathbf{B}(\mathbf{n}_1) - \mathbf{B}(\mathbf{n}_2)$ in Eq. (19) is really the gradient, then $\mathbf{H}_1 \equiv \mathbf{H}_2$, but at the same time we see from Eq. (26) that

$$\mathrm{curl}\,\mathbf{B}(n_1)-\mathrm{curl}\,\mathbf{B}(n_2)=\mathbf{H}_1-\mathbf{H}_2\neq 0. \quad (27)$$

In other words, the transition from the Schrödinger equation with the vector potential $\mathbf{B}(\mathbf{n}_1)$ to the equation with the vector potential $\mathbf{B}(\mathbf{n}_2)$ implies that the real magnetic field changes. In this case the field along the string $\mathbf{n}_1$, $\mathbf{H}_1' = \mathbf{n}_1\theta(\mathbf{n}_1\mathbf{r})\,\delta(r^2-(\mathbf{n}_1\mathbf{r})^2)$, corresponding to the vector potential $\mathbf{B}(\mathbf{n}_1)$ in the first Schrödinger equation, disappears, but at the same time a magnetic field arises along the string $\mathbf{n}_2$ in the amount $\mathbf{H}_2' = \mathbf{n}_2\theta(\mathbf{n}_2\mathbf{r})\,\delta(r^2-(\mathbf{n}_2\mathbf{r})^2)$, which corresponds to the vector potential $\mathbf{B}(\mathbf{n}_2)$ in the second Schrödinger equation (of course, the Coulomb parts are not altered during these changes). It is evident that no gradient transformation can change the real magnetic field. Therefore, if one finds such a transformation which changes $\mathbf{B}(\mathbf{n}_1)$ to $\mathbf{B}(\mathbf{n}_2)$, then such a transformation cannot be termed a gradient transformation. Conversely, if Eq. (19) is valid, we must find that $\mathbf{n}_1 = \mathbf{n}_2$.

Thus, is it possible to represent the difference $\mathbf{B}(\mathbf{n}_1) - \mathbf{B}(\mathbf{n}_2)$ in the form of a gradient? The answer to this question is very important. A positive answer would mean that the Dirac theory is not contradictory and Eq. (1) is a necessary consequence of that theory. A negative answer would cast doubt on the consistency of not only the original premises of the Dirac theory, but it would also call into question Eq. (1) concerning magnetic charge quantization, which is a consequence of the assumption expressed in Eq. (19) that the theory is gradient-invariant.

We now present some mathematical considerations which refute the possibility of representing the difference $\mathbf{B}(\mathbf{n}_1) - \mathbf{B}(\mathbf{n}_2)$ as the gradient of some scalar function $g\chi(\mathbf{r})$. One can show that the vector potential $\mathbf{B}(\mathbf{n})$ in Eq. (10) can be uniquely represented in the form

$$\mathbf{B}(\mathbf{n}) = -g\,\mathrm{curl}\,\{\mathbf{n}\ln(r-\mathbf{nr})\}, \quad (28)$$

i.e., it does not contain a gradient part in the equivalent expansion

$$\mathbf{B}(\mathbf{n}) = -\frac{1}{4\pi}\,\mathrm{grad}\int\frac{\mathrm{div}\,\mathbf{B}'d^3r'}{|\mathbf{r}-\mathbf{r}'|}+\frac{1}{4\pi}\,\mathrm{curl}\int\frac{\mathbf{B}'d^3r'}{|\mathbf{r}-\mathbf{r}'|},$$

because div $\mathbf{B} \equiv 0$ everywhere, as can be shown by using the corresponding integral theorem. A somewhat modified expression for $\mathbf{B}(\mathbf{n})$ was obtained by Schwinger[2] (the difference is that Schwinger treats two strings rather than just one).

It follows immediately from Eq. (28) that Eq. (19) must be impossible. We can also show this in another way. By directing $\mathbf{n}_1$ along the x axis and $\mathbf{n}_2$ along the y axis, one can perform the integrations along various paths in Eq. (21). It is easily shown that the integral in Eq. (21) depends on the path, so that quite different values are obtained for $\chi(\mathbf{r})$ when different paths are used. This, of course, agrees with Eq. (28) and contradicts Eq. (19). The fact that Eq. (21) cannot be used to determine a scalar function $\chi(\mathbf{r})$ also renders Eq. (20) empty.

We should keep in mind that the validity of our previous discussion is certified only with the help of sufficiently accurate transformations with generalized functions, for which the criterion is the application of integral theorems. Equation (14) is a case in point. It is clear from this equation that curl $\mathbf{B}(\mathbf{n})$ and g grad$(1/r)$ agree everywhere in space except on the string. However, the integral of curl $\mathbf{B}$ over a closed path which encompasses the string is not zero, while the integral of g grad$(1/r)$ over any closed path whatsoever is zero. Therefore the field curl $\mathbf{B}$ remains a solenoidal field, while the field g grad$(1/r)$ is a gradient field. And for the same reason the difference curl $\mathbf{B}(\mathbf{n}_1)$ − curl $\mathbf{B}(\mathbf{n}_2)$ cannot be written as the gradient of some scalar function $g\chi(\mathbf{r})$.

This conclusion casts doubt on the validity of the magnetic charge quantization theorem of Eq. (1), which means that its smallest value

$$|g| = 137\,|e|/2 \quad (29)$$

is questionable. We note that there are many different variants of the predictions of Eq. (1), all based on the requirement that the wave function phase be single-valued. However, the variant presented above appears to us to be the most rigorous. Nevertheless, there are serious difficulties inherent in it.

We wish to note that, by using a vector potential containing m strings,[4,5]

$$\mathbf{B}' = \frac{g}{mr}\sum_{s=1}^{m}\frac{[\mathbf{n}_s,\mathbf{r}]}{r-\mathbf{n}_s\mathbf{r}}, \quad (30)$$

which also satisfies the primary requirements of Dirac's theory, because a correct Lorentz force is obtained everywhere outside the strings, we can find any "quantization" of magnetic charge. Thus, with m = 1 we have Dirac's theory,[1] while with m = 2 and $\mathbf{n}_1 = -\mathbf{n}_2 = \mathbf{n}$ we have

$$\mathbf{B}^{\mathrm{Schw}}=\frac{g}{2r}\left[\frac{[\mathbf{n},\mathbf{r}]}{r-\mathbf{nr}}-\frac{[\mathbf{n},\mathbf{r}]}{r+\mathbf{nr}}\right], \qquad (31)$$

which is the Schwinger theory.[2]

Unlike Dirac's theory, the magnetic flux in Schwinger's paper[2] is conducted to the origin along two strings, so that the "quantization rule" is as shown in Eq. (2). As a consequence, Schwinger finds the minimum magnetic charge to be given by

$$|g|=137|e|. \qquad (32)$$

Although there are differences between the consequences of the two approaches, as expressed in Eqs. (29) and (32), Schwinger's theory is not essentially different from Dirac's. Since it has all the inherent difficulties present in the Dirac theory, we shall not devote special attention to it.

In the general case Eq. (30) results in a minimum magnetic charge which is m times the Dirac value:

$$|g|=m\frac{137}{2}|e|, \qquad (33)$$

which is easily demonstrated by repeating the derivation of Eq. (1) with the vector potential given in Eq. (30) and running the contour around one of the strings. If two or more strings are included within the contour, the minimum magnetic charge will be altered accordingly.

It seems to us that the possibility in principle of working in the Dirac theory with the vector potential of Eq. (30) emphasizes the weak aspects of theories which use potentials of the type in Eq. (10) and (31). We shall now analyze the compatibility of the second and third original postulates in Dirac's theory, as formulated in Sec. 3.

Dirac's veto ($\psi=0$ on the string) ensures that the observed magnitude of $\psi^*\psi$ is zero on the string. It can be shown that $\psi^*\psi$ goes to zero smoothly on the string.[4] As a consequence, one should observe some effect from this fact; electrons scattered by a monopole will "sense" the string. A change in the direction of the vector **n** in **B(n)** will lead to changes in the scattering pattern for different positions of the vector **n**. In other words, the condition that $\psi^*\psi$ be zero on the string leads to the observability of the string itself. The question of the observability of the string thus has two aspects: On the one hand, the condition that $\psi=0$ on the string removes the δ-function "pulses" in the Lorentz force of Eq. (16), thus rendering them unobservable, but on the other hand, this same condition now makes the string itself observable in the sense described above.

It would seem that a direct answer to the problem of the string's observability could be obtained by solving the problem of electron scattering by a stationary monopole. This problem has in fact been solved by Banderet.[7] However, he started with the assumption that the string is unobservable and that its direction in space makes no difference. Then, in order to simplify the mathematics, he positions the vector **n** parallel to the incident beam of electrons. As a result, even if the electrons did scatter from the string in this case, it would be masked by the cylindrical symmetry of the problem.

The scattering of an electron beam by a stationary monopole could give the solution to the string observability problem if the string is situated at some angle relative to the beam. This problem is now being solved. The appearance of scattering in this case would provide a direct answer to the question of consistency in the original assumptions of Dirac.[1]

What would be the consequences of an observation of the string? It would mean that the Dirac theory describes an infinitely long, infinitely thin solenoid, one end of which is at rest at the origin and the other end of which goes off to infinity. There would not be a theory of magnetic charge as such.

To conclude our discussion of Dirac's theory, we shall make a final comment about the string. We are not obliged to view the string as running from the origin to infinity in a straight line. It can have any shape of curve, and all the above discussion remains valid.

Since Eq. (19) is not satisfied, it seems useless to try other forms for the potential **B** (an example would be potentials which are singular over a plane rather than along some curve).

4. CRITIQUE OF OTHER METHODS OF PROVING EQS. (1) AND (2)

The preceding method of proving Eqs. (1) and (2), as well as a large number of similar proofs, was based on the requirement that the phase of the wave function be single-valued. We will now examine a group-theoretical approach to the derivation of these two equations; this is a second type of proof.

There are a large number of papers devoted to similar types of approaches. However, for illustrative purposes, we shall treat just one of these works, that of Peres,[8] in which Eq. (2) is derived.

It is shown in ref. 8 that if the angular momentum of an electron moving in the field of a stationary monopole is written in the form

$$\mathbf{J}=\left[\mathbf{r},\left(\frac{\hbar}{i}\nabla-\frac{e}{c}\mathbf{B}\right)\right]+\frac{eg}{c}\cdot\frac{\mathbf{r}}{r}, \qquad (34)$$

where **B** has the form given in Eq. (10), then the following commutation relation is valid:

$$[J_m, J_n]=i\hbar\varepsilon_{mnk}J_k. \qquad (35)$$

Then, according to Peres, Eq. (35) results in, first, the rotational symmetry of the electron–monopole system, and second, the quantization of the angular momentum of motion (eg/c) (r/r) (according to quantum-mechanical rules):

$$eg/\hbar c=k \qquad (k=0, \pm 1, \pm 2\ldots).$$

However, an accurate transformation using the singular function **B** in Eq. (10), taking Eq. (14) into account, leads us to the following expression:

$$[J_m, J_l]=i\hbar\varepsilon_{mlh}J_h+i\hbar e/c\,\{r^2\varepsilon_{mlh}n_h+r_m[\mathbf{r},\mathbf{n}]_l-r_l[\mathbf{r},\mathbf{n}]_m\}\times 8g\theta(\mathbf{nr})\,\delta[r^2-(\mathbf{nr})^2]. \qquad (36)$$

It is clear from Eq. (36) that the conclusion of total rotational symmetry for the system is not correct because the components of J in Eq. (34) do not form the group of Eq. (35). It also follows that the proof of Eq. (2) as presented in ref. 8 is not rigorous, at the very least. We must mention that there are a large number of proofs of Eqs. (1) and (2) which are based on group-theoretic considerations. We will not dwell on these proofs any further here since all of them encounter the same difficulties which face Peres.[8]

Among the various proofs of Eqs. (1) and (2), the papers by Saha and Wilson[3] stand apart. The derivations of these two equations in ref. 3 can only be called proofs in a limited sense, for they are of a more illustrative character rather than proofs. These authors consider a system of stationary electrical and magnetic charges separated by a distance l. By forming the classical Poynting vector for the stationary electric and magnetic fields, they then calculate the angular momentum **M** of this field relative to the line joining the two charges:

$$\mathbf{M} = eg\mathbf{l}/c\,|\mathbf{l}|, \quad (37)$$

where $|\mathbf{l}|$ is the distance between charges. Then, equating $|\mathbf{M}|$ to $k\hbar$ they find that

$$eg/c = k\hbar \qquad (k = 0, \pm 1/2, \pm 1, \pm 3/2, \pm 2 \ldots), \quad (38)$$

which gives either Eq. (1) or (2).

It is clear that this treatment is semiclassical. However, the main difficulty with this derivation is that these authors do not know what quantity they have quantized. The lack of a Lagrangian and canonical formalism prevents them from relating the quantized quantity with any physical quantity obtained from a quantum-mechanical canonical formalism. It is for this reason that their treatment is a quasiclassical illustration of Eqs. (1) and (2).

A similar criticism can be leveled against other work in which Eqs. (1) and (2) are derived without the benefit of a canonical formalism. A canonical formalism is required because quantum mechanics is a Hamiltonian theory.

We conclude by looking at a number of papers which use the Mandelstam formalism to arrive at Eqs. (1) and (2) (see ref. 9, for example). These studies are by their character related to those which use the first type of proof for Eqs. (1) and (2). Without dwelling specifically on the absence of a canonical formalism, we note that the use of paths in line integrals encountered in these papers introduces problems which are similar to those which appear when the string is introduced. Although it is somewhat of an oversimplification, one might say that the line integral in ref. 9 is in some sense analogous to the string, and bears all its attendant difficulties.

5. CONCLUSION

None of the proofs of Eqs. (1) and (2) is free of serious objections. Moreover, it appears that these difficulties are not accidental. Thus, the absence of gradient invariance in the Dirac theory and the simultaneous absence of a group in Eq. (35) are not fortuitous; they seem to be related by the fact that Eqs. (1) and (2) were derived only to an accuracy of δ-function singularity concentrated around a line (the string). This is also characteristic of the other approaches. In other words, one might think that Eqs. (1) and (2) simply do not exist.

The main problem with the theory of magnetic charge is that it is at present impossible to write a Lagrangian for the interaction of an electron's field with the monopole current (the term current is used here in a conditional sense), or for the interaction of a monopole's field with the corresponding electron current. It is even impossible to say whether this difficulty is fundamental or merely technical. The criterion for the validity of any interaction Lagrangian formulation is the a priori postulated form of the Lorentz force.

The Lagrangian suggested by Dirac and Schwinger does not completely satisfy the problem as posed here. In addition, the appearance of the string attests to the major difficulties with these theories.

The fate of Eqs. (1) and (2) of course affects the experimental methods used to detect magnetic charge. It is clear from Eqs. (29) and (30) that the magnetic charge is huge. The appearance of a single magnetic charge of such magnitude could not go undetected by experiment.

However, if one assumes that g can be different, much smaller, for example, this would produce a real revolution in the experimental methods used to find the magnetic charge. Such an assumption would mean that the magnetic charge is now being sought in places where it cannot be found. The absence of a rigorous theory for the magnetic charge speaks in favor of such an assumption.

[1]P. A. M. Dirac Proc. Roy. Soc., A133, 60 (1931); 74, 817 (1948).
[2]J. Schwinger, Phys. Rev., 144, 1087 (1966).
[3]M. N. Saha, Ind. J. Phys., 10, 141 (1936); Phys. Rev., 75, 1968 (1949); H. A. Wilson, Phys. Rev., 75, 309 (1949).
[4]Yu. D. Usachev, Report to the All-Union Inter-University Conference at Uzhgorod (October 1968).
[5]B. M. Bolotovskii and Yu. D. Usachev, in: The Dirac Monopole [in Russian], Mir, Moscow, (1970).
[6]G. Wentzel, Progr. Theor. Phys. Suppl., 37-38, 163 (1966).
[7]P. P. Banderet, Helv. Phys. Acta, 19, 503 (1946).
[8]A. Peres, Phys. Rev., 167, 1449 (1968).
[9]N. Calibbo and E. Ferrari, Nuovo Cimento, 23, 1147 (1962).