Dirac particle in a magnetic field: Symmetries and their breaking by monopole singularities*

Alfred S. Goldhaber

Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794

(Received 18 May 1977)

Some rules governing motion of a charged particle obeying the Dirac equation are assembled, including exact helicity conservation for scattering on an arbitrary finite magnetic field configuration. The singularity at the location of a magnetic monopole invalidates the derivation of the rules mentioned, leaving the Dirac Hamiltonian H undefined for the lowest angular momentum state of the electron in the field of the pole. Specifying the behavior of H under the discrete P, T, and C symmetries determines it almost uniquely. One result is that H may possess a bound state of zero energy, contrary to assertions in early papers on the subject. Zero-energy bound states which violate the superselection rule for electric charge are also studied, including one which is the point limit of a solution for a fermion multiplet interacting with a finite-energy soliton monopole. Implications of such a bound state for second quantization have been considered previously by others and are further analyzed here. The suggestion that monopoles may possess half-integral fermion number is shown to be unwarranted by present evidence.

I. INTRODUCTION

The Dirac equation for a charged spin-$\frac{1}{2}$ particle with gyromagnetic ratio 2 moving in the field of a magnetic monopole was studied thirty years ago,[1,2] but the resultant electron-pole scattering amplitude was found quite recently.[3] The amplitude has some remarkable features:

(a) All except the lowest angular momentum state contribute only to the helicity-nonflip amplitude $f(+\to+)=f(-\to-)$.

(b) The nonflip amplitude depends on electron wave number k only through an overall scale factor $1/k$, and otherwise is determined purely by the scattering angle. The amplitude at a given k is the same as the one obtained at that k by solving the Schrödinger equation for a nonrelativistic particle with gyromagnetic ratio 2.[4]

(c) The lowest angular momentum state contributes only to helicity flip, which is nonvanishing when the incoming electron spin is opposed to the "angular momentum of the electromagnetic field," that is, when the incoming helicity has opposite sign from the product q of electron charge e and monopole strength g. This amplitude obeys the symmetry $f_q(+\to-)=f_{-q}(-\to+)$.

(d) The helicity-flip amplitude has an energy-dependent phase (whose sign is ambiguous) but otherwise is the same function of k as the nonrelativistic amplitude of Ref. 4.

The present article is an attempt to understand, on the basis of symmetry laws governing magnetic interactions of a Dirac electron, the regularities shown by electron-monopole scattering. Section II contains a description of such symmetry laws, including a helicity-conservation principle applicable to a particular type of "g-2" experiment measuring the anomalous moment of a lepton. The symmetries apply not only to ordinary magnetic interactions but also to those of a fermion multiplet minimally coupled to a non-Abelian gauge field.

In Sec. III the symmetries are shown to imply the regularities (a) and (b) in a straightforward manner. However, for the lowest angular momentum state, with $J=|q|-\frac{1}{2}$, the Dirac Hamiltonian H is not well defined, since the radial wave function for this state must not vanish at the origin, and the kinetic momentum operator $\vec{p}-e\vec{A}$ which appears in H is singular there. The discrete PT and PC symmetries may be used to constrain H. These symmetries, together with the requirement that H be self-adjoint, determine the boundary condition on the lowest partial wave at the origin, except for the choice of a sign. This sign ambiguity corresponds precisely to the ambiguity in sign of the phase of the helicity-flip amplitude mentioned in item (d). Thus, an analysis based on general principles of quantum mechanics shows that the scattering amplitudes found in Ref. 3 are the only possible ones consistent with the discrete symmetries.

In the nonrelativistic case the absence of any charge-conjugation symmetry precludes determining the boundary condition at the origin. The result of Ref. 4 for the phase of the helicity-flip amplitude corresponds to one possible choice, equivalent to the condition that the wave function be finite at the origin. The phase disagrees with either Dirac equation phase already at the first order in electron speed.

In Sec. IV a zero-energy bound state is found to be present for one of the two allowed Dirac Hamiltonians. Thus the necessity to revise the boundary condition at the origin vitiates a proof[2] which used a standard method to show that there were no bound states. The techniques used here may also be applied to monopoles arising in spontaneously broken non-Abelian gauge theories, when the dimensions of the monopole interior are small compared to the wavelength or Compton wavelength of the electron. It was shown recently[5] that there is a single bound state for the case of a spin-$\frac{1}{2}$, isospin-$\frac{1}{2}$ fermion interacting with a soliton[6] monopole, provided that the fermion mass is generated by the same scalar isovector field which breaks the SU(2) gauge symmetry. This result is confirmed, but it is also shown that the same theory for fermions of fixed mass yields no bound state, since then helicity is conserved even for the lowest partial wave. The helicity-conserving case corresponds to a particular point in a one-parameter family of self-adjoint extensions of the Hamiltonian, all realizations of the discrete symmetries P, T, C. The extreme cases correspond to a Dirac monopole with two or no bound states. Only these extremes obey the superselection rule for electric charge.

Finally, in Sec. V is explored the nature of a hybrid second-quantization scheme, in which the point monopole is treated as a fixed object, while the fermion degrees of freedom are described by quantum fields. As shown in Ref. 5, the zero-energy bound state has strong implications for the second-quantized formalism: The monopole must be a doublet whose two states are connected by the fermion field. These matters are re-examined here from a somewhat different perspective, with the conclusion that one member of the doublet should be taken as a fermion vacuum, and the other as an occupied fermion or antifermion state, thereby implying the existence of two types of monopoles: fermion and antifermion acceptors.

II. SYMMETRIES OF MAGNETIC INTERACTIONS OF A DIRAC PARTICLE

A. Discrete symmetries

The Dirac equation

$$i\dot{\psi} = H\psi = \{\alpha\cdot[-i\nabla - e\vec{A}(\vec{r},t)] + \beta M + eV(\vec{r},t)\}\psi \tag{2.1}$$

in the absence of external fields exhibits the well-known P, T, and C symmetries

$$H = PHP^{-1} = THT^{-1} = -CHC^{-1} \tag{2.2}$$

with

$$\begin{aligned} P\psi(\vec{r},t) &= \beta\psi(-\vec{r},t),\\ T\psi(\vec{r},t) &= \sigma_2\psi^*(\vec{r},-t),\\ C\psi(\vec{r},t) &= \gamma_2\psi^*(\vec{r},t) \end{aligned} \tag{2.3}$$

using standard conventions for the Dirac matrices $\vec{\alpha}$, β, γ^μ and the Pauli matrices $\vec{\sigma}$. The free-particle angular momentum

$$\vec{J} = \vec{L} + \tfrac{1}{2}\vec{\sigma} = -i\vec{r}\times\nabla + \tfrac{1}{2}\vec{\sigma}$$

obeys the symmetry laws

$$\vec{J} = P\vec{J}P^{-1} = -T\vec{J}T^{-1} = -C\vec{J}C^{-1}. \tag{2.4}$$

In the presence of electromagnetic fields, the symmetries still hold for H under the assumptions

$$\begin{aligned} P\vec{A}(\vec{r},t)P^{-1} &= -\vec{A}(-\vec{r},t),\\ T\vec{A}(\vec{r},t)T^{-1} &= -\vec{A}(\vec{r},-t),\\ C\vec{A}(\vec{r},t)C^{-1} &= -\vec{A}(\vec{r},t),\\ PV(\vec{r},t)P^{-1} &= V(-\vec{r},t),\\ TV(\vec{r},t)T^{-1} &= V(\vec{r},-t),\\ CV(\vec{r},t)C^{-1} &= -V(\vec{r},t). \end{aligned} \tag{2.5}$$

In a general field, $\vec{J}$ is not conserved, which makes it of less interest. However, in a monopole field centered at the origin there is a conserved angular momentum which obeys the symmetries (2.4) under the same assumptions (2.5), namely

$$\vec{J} = \vec{r}\times(-i\nabla - e\vec{A}) + \tfrac{1}{2}\vec{\sigma} - eg\hat{r}. \tag{2.6}$$

The last term involves the magnetic charge g, which by (2.5) is reversed by each of the discrete symmetries, since they each reverse the monopole magnetic field $\vec{B} = \nabla\times\vec{A} = g\hat{r}/r^2$. The consequence of (2.6) to be borne in mind is that $\vec{J}^2$ commutes with P, T, and C individually or in combination, while g commutes with any pair of the symmetries, so that the pair can give no more than a gauge transformation to $\vec{A}$. Since the gauge dependence may be chosen in such a way that the radial function ψ_J in a partial-wave decomposition of ψ has no explicit gauge dependence,[3] the implication of (2.5) for the radial Hamiltonian H_J is simply $g\to g$ under PT, PC, or TC, but $g\to -g$ under P, T, or C. The discussion here is simply a formal example of the remark[7] that for monopole magnetoelectrodynamics the discrete symmetries will be invariances of the total Hamiltonian only if they each reverse magnetic charge.

B. Dynamic symmetries

Implicit in Dirac's first paper[8] on the wave equation of the electron is the relation, valid for a pure magnetic field,[1]

$$k^2 \equiv H^2 - M^2 = (-i\nabla - e\vec{A})^2 - e\vec{\sigma}\cdot\vec{B}. \tag{2.7}$$

Since a nonrelativistic particle with gyromagnetic ratio 2 would obey the Schrödinger equation with

$$k^2 \equiv 2MH = (-i\nabla - e\vec{A})^2 - e\vec{\sigma}\cdot\vec{B}, \tag{2.8}$$

it follows that for fixed k^2 the large components (and separately the small components) of a solution of the Dirac equation for a nonsingular magnetic field also solve the Schrödinger equation for the same field. This equivalence, or invariance,[9] is a subtle extension of a classical result into the quantum mechanics of particles with spin: The trajectory of a charged particle in a static magnetic field depends only on its momentum, although the speed of passage along the trajectory depends also on the mass. The quantum version of this statement is that the scattering amplitude of a spin-$\frac{1}{2}$ particle with gyromagnetic ratio 2 (or a spinless particle obeying the Klein-Gordon equation[10]) interacting with a localized magnetic field is a function only of wave number k, and not explicitly dependent on mass.

Even though (2.7) is obtained by setting the electric field and potential to zero in Dirac's formula[8]

$$(i\partial_t - eV)^2 - (-i\nabla - e\vec{A})^2 + e\vec{\sigma}\cdot\vec{B} - ie\vec{\alpha}\cdot\vec{E} = M^2\,, \tag{2.9}$$

it was not utilized in the first solution of the Dirac equation in a uniform field.[11] More than two decades later the invariance relation between (2.7) and (2.8) was exploited to simplify solution of the uniform field problem.[12]

If the Dirac spinor is taken also to be a multiplet in some gauge group, and minimally coupled to the gauge field, then the invariance discussed here still applies whenever the gauge field has a purely static magnetic configuration, that is, $V=0$ and $\vec{A}=\vec{A}(\vec{r})$. In such a case $\vec{A}$ and $\vec{B}$ are matrices in the Lie-algebra representation acting on the spin-$\frac{1}{2}$ gauge-group multiplet.

Another conservation law for a Dirac particle involves the "helicity operator"

$$h \equiv \vec{\sigma}\cdot(-i\nabla - e\vec{A})\,. \tag{2.10}$$

It is easily seen that h obeys

$$dh/dt \equiv e\vec{\sigma}\cdot\vec{E} \tag{2.11}$$

and so is conserved in a static nonsingular magnetic field. Therefore, its square

$$h^2 = (-i\nabla - e\vec{A})^2 - e\vec{\sigma}\cdot\vec{B} \tag{2.12}$$

is also conserved, a fact already implied by the preceding discussion. Far away from a localized magnetic field configuration, $h/(h^2)^{1/2}$ is simply the helicity. This means that the scattering amplitude for a Dirac particle on such a field configuration must be pure helicity nonflip. Once again, the same remark applies to the general case of a multiplet interacting with a gauge field of any group, not just the U(1) of conventional electrodynamics.

The above result is relevant to a somewhat idealized version of the g-2 experiment, in which change of helicity in a magnetic field becomes a measure of the departure of the gyromagnetic ratio of the electron or muon from the Dirac value of $2e\hbar/2mc$. The point is that for vanishing g-2 there could be no change of helicity even if the magnetic field were inhomogeneous on a very short length scale. This is a stronger result than the well-known statement[13] that in a homogeneous field the precession is proportional to g-2. Since in practical experiments the inhomogeneities are negligible on the scale of an electron Compton wavelength, there is no change needed in traditional analyses of the experiments. Nevertheless, it is remarkable that the Dirac equation implies a conservation law which would not follow from the classical equations of motion for a spinning particle in an inhomogeneous magnetic field. Even in the nonrelativistic limit the conservation is special to spin $\frac{1}{2}$, since it depends on the fact that the Hamiltonian is proportional to the square of the helicity operator and so must commute with it. This is not true for any higher spin.

III. ELECTRON-MONOPOLE SCATTERING

A. Discrete symmetries

Since all the symmetry operations of the previous section have definitions which are unambiguous except for possible gauge transformations, while the scattering amplitude f for an electron on some magnetic configuration is invariant under gauge transformations, the symmetries may be applied to f in a straightforward manner. Let us describe the initial and final states by helicity h, which is reversed by P but not T or C, and momentum $\vec{k}$, which is reversed by P or T but not C. Then we get

$$f_q(h\to h', \vec{k}\to\vec{k}') = f_{-q}(-h\to -h',\ -\vec{k}\to -\vec{k}')\,, \tag{3.1P}$$

$$f_q(h\to h', \vec{k}\to\vec{k}') = f_{-q}(h'\to h,\ -\vec{k}'\to -\vec{k})\,, \tag{3.1T}$$

and the product relation

$$f_q(h\to h', \vec{k}\to\vec{k}') = f_q(-h'\to -h, \vec{k}'\to\vec{k})\,. \tag{3.2}$$

The symmetry mentioned in item (a) of the introduction, $f(+\to +) = f(-\to -)$, follows from (3.2), while $f_q(+\to -) = f_{-q}(-\to +)$ [item (c)] follows from either version of (3.1).

B. Dynamic invariances

If the other results in Sec. II held for the singular monopole field, they would imply that the relativistic and nonrelativistic problems give the same scattering amplitudes for equal k. Since the only length scale in the nonrelativistic problem is set by $1/k$, it follows that f would be a function of angle alone, multiplied by $1/k$. These statements are true for the nonflip amplitude, as mentioned in item (b). Further, helicity conservation would imply vanishing of the flip amplitude. Both kinds of dynamic invariance fail only for the lowest-partial-wave amplitude, which is pure helicity flip and has a phase which depends explicity on k/M.

C. Singular point, construction of a self-adjoint extension of the Hamiltonian, and consequent dynamical symmetry violation

Since the magnetic monopole field is singular at its center, it is not obvious that the Dirac Hamiltonian H (2.1) is self-adjoint, as is necessary for the generator of time evolution of the wave function. We may exploit the manifest rotational symmetry of H away from the center to make an angular momentum or partial-wave decomposition of H. Then the issue of self-adjointness may be examined in each partial wave separately. In addition to helicity, the operator $\vec{\sigma}\cdot\hat{r}$ commutes with $\vec{J}$. For the lowest allowed angular momentum state $J=|q|-\frac{1}{2}$, $\vec{\sigma}\cdot\hat{r}$ must take the value $S\equiv q/|q|$, since $|\vec{J}\cdot\hat{r}|=|-q+\frac{1}{2}\vec{\sigma}\cdot\hat{r}|$ must be no greater than J. If $\vec{\sigma}\cdot\hat{r}$ has only one value, then only $f(h=-k\vec{\sigma}\cdot\hat{r}\to h'=+k\vec{\sigma}\cdot\hat{r})$ may be nonvanishing, so that the necessity of helicity flip for the lowest partial wave is obvious. Furthermore, unitarity permits immediate deduction of the magnitude and angular dependence of this helicity-flip amplitude, with only an energy-dependent phase undetermined[3]:

$$kf(h=-kS\to h'=kS)=-q[\sin(\theta/2)]^{2|q|-1}e^{-iS\phi}e^{2i\delta(k)}\ , \quad (3.3)$$

where θ and ϕ are the polar and azimuthal angles with respect to the beam direction. This inevitable failure of helicity conservation shows that the singularity must lead to inconsistencies if H and the helicity operator h are treated as well as well-defined operators whose commutation away from the singularity implies commutation everywhere. Another way to see that such inconsistency may arise is to consider the Jacobi identity for the three Cartesian components of $-i\nabla-e\vec{A}$,

$$0=e\nabla\cdot\vec{B}\,. \quad (3.4)$$

This shows at once that magnetic vector potentials may not be used to describe a distributed magnetic source (just the familiar statement div curl $\vec{A}=0$), but it also suggests possible difficulties even for point sources[14]; the reason it only suggests difficulties is that the differential operations leading to (3.4) are not well defined at the singular point. In any case, the only requirement for a consistent solution of the equations of motion is that H be self-adjoint, and agree with its definition as a differential operator everywhere that the operator is well defined. If these conditions can be satisfied the only possible adjustable parameters in H will appear in boundary conditions on partial-wave functions at the origin. For $J>|q|-\frac{1}{2}$, the partial-wave Hamiltonian is self-adjoint with the boundary condition that the wave function vanish at the origin. This seems obvious from inspection of Eqs. (22) and (25) of Ref. 3, which exhibit the radial Hamiltonian, but it may be shown rigorously with the techniques applied below for the lowest partial wave. Furthermore, the helicity operator is also self-adjoint for $J>|q|-\frac{1}{2}$ with the same boundary condition $\psi(r=0)=0$, and so commutes with H. The self-adjointness of H means that H^2 is defined, so that the equivalence of the relativistic and nonrelativistic Hamiltonians for computing scattering at fixed k is also established.

Let us now focus attention on the lowest partial wave, whose "large" and "small" components may be written as $F(r)\eta_m/r$ and $G(r)\eta_m/r$, respectively, where η_m is an angle-dependent spinor corresponding to $J=|q|-\frac{1}{2}$ and $\vec{\sigma}\cdot\hat{r}=S$. The Hamiltonian acting on the column vector (F,G) is given by Eq. (28) of Ref. 3 and may be written in matrix form as

$$H_0=-iS\rho\,\partial_r+\beta M \quad (3.5)$$

with

$$\rho=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \beta=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\,. \quad (3.6)$$

In the same notation, the helicity operator is evidently

$$h_0=-iS\partial\ . \quad (3.7)$$

Both H_0 and h_0 are Hermitian with the boundary condition at the origin $F=G=0$. However, they are not self-adjoint on this class of wave functions. It is possible to make a self-adjoint extension of H_0, but h_0 will not even be Hermitian with the resultant boundary condition, and consequently need not be conserved.

The Weyl–von Neumann theory of self-adjoint operators[15] requires some definitions: An operator A which is Hermitian ($A_{mn}=A^*_{nm}$) on a dense subset of a Hilbert space has *deficiency indices* $n_\pm$ if the equation

$$A^\dagger\psi=\pm i\psi \quad (3.8)$$

has $n_\pm$ linearly independent (and normalizable) solutions. For simple differential operators $A^\dagger$ means the operator A, but applied to vectors outside the subset on which A is Hermitian. If the deficiency indices are nonzero then A is not self-adjoint. However, if the indices are equal, $n_+=n_-=n$, there is a family of self-adjoint extensions of A, parametrized by a set of $n\times n$ unitary matrices U, corresponding to maps of the eigenvectors with eigenvalue $+i$ onto the $-i$ eigenvectors. For any particular U, given a set of n orthonomormal vectors $v_j^{(+)}$ for $+i$, each one is matched to a particular $v_j^{(-)}$ for $-i$. The domain on which the extension of A is a self-adjoint operator is then the set of vectors lying in the original domain of A, added to arbitrary linear combinations of the n vectors $v_j = v_j^{(+)} + v_j^{(-)}$, on which the action of A as extended is

$$A v_j \equiv i(v_j^{(+)} - v_j^{(-)}) . \tag{3.9}$$

For H_0, the deficiency indices are $n_+ = n_- = 1$, with

$$\begin{aligned} G^{(+)} &= e^{i\alpha}F^{(+)} = \exp[-r(M^2+1)^{1/2} + i\alpha] , \\ G^{(-)} &= -e^{-i\alpha}F^{(-)} = -\exp[-r(M^2+1)^{1/2} - i\alpha + i\chi] \end{aligned} \tag{3.10}$$

and $e^{i\alpha} = (1+iM)/S(M^2+1)^{1/2}$. The single real parameter χ labels the different self-adjoint extensions of H_0. By adjusting χ, one may obtain a $v = v^{(+)} + v^{(-)}$ with arbitrary imaginary ratio G/F. Since the domain of H_0 as extended contains wave functions which vanish at the origin, plus v, we find that H_0 is self-adjoint on the domain of $L^{(2)}$ differentiable functions obeying the condition

$$G(0)/F(0) = iaS \tag{3.11}$$

for any real a. It is trivial that $H_0(a)$ is Hermitian on this set; the profound part of the Weyl–von Neumann construction is the proof that, since $H_0(a)$ has deficiency indices $n_+ = n_- = 0$, it is self-adjoint and has no further self-adjoint extension.

In Sec. IV we shall see a somewhat more complicated self-adjoint extension, but it is perhaps worth noting here that the helicity operator h_0 is a classic example with no such extension, having deficiency indices

$$n_{-S} = 0, \quad n_S = 2 . \tag{3.12}$$

We now come to the question of determining the parameter a. Recall that away from $r = 0$, H obeys the discrete symmetries given in Sec. II, so that in particular one has for H_0

$$CPH_0(CP)^{-1} = \rho H_0^* \rho = -H_0 . \tag{3.13}$$

If this symmetry is to hold for the full H_0, then the boundary condition must be invariant under CP, giving

$$\rho \begin{pmatrix} 1 \\ iaS \end{pmatrix}^* = \lambda \begin{pmatrix} 1 \\ iaS \end{pmatrix} \tag{3.14}$$

or

$$a = \pm 1 . \tag{3.15}$$

It will be relevant later to note that P invariance implies

$$a(S) = a(-S) . \tag{3.16}$$

For either sign of a, we may compute the phase shift resulting from the boundary condition as

$$\tan\delta_a = -ak/(E+M) . \tag{3.17}$$

These are precisely the two possible phase shifts obtained in Ref. 3 by adding a small anomalous magnetic moment κ and taking the limit $\kappa \to 0$. The sign of κe is simply a. For any nonzero κ this procedure plainly gives an H_0 which still obeys the CP symmetry $H_0 \to -H_0$, so that the symmetry should apply in the limit. What the discussion above has shown is that *any* procedure giving a self-adjoint H_0 with the usual CP transformation must yield $a = \pm 1$. Therefore the limits $\kappa \to \pm 0$ exhaust the possibilities for maintaining CP invariance of electron-monopole interactions.

For the nonrelativistic problem the radial Hamiltonian analogous to H_0 is given by

$$2MH_{0nr} = -\partial_r^2 . \tag{3.18}$$

Again there is a one-parameter family of self-adjoint operators specified by the boundary condition

$$F(0)/F'(0) = b . \tag{3.19}$$

The conventional choice, followed in Ref. 4, corresponds to $b = 0$, giving a phase shift $\delta_b = 0$. Since there is no C operation defined in the nonrelativistic case, there is no general principle to specify b. To make δ_b agree with δ_a to first order in electron speed one should choose $b = -a/2M$, but this is evidently arbitrary. For no constant b will the nonrelativistic solution equal the relativistic one, so that the necessity of defining H at the singular point has destroyed the equivalence of the Dirac and nonrelativistic Hamiltonians for electrons of equal wave number in a magnetic field. The equivalence could only be restored by the CP-violating assumption $1/a = b = 0$ or ∞.

D. Some consequences for experimental phenomena

The issue of higher-order corrections to the approximation that a Dirac particle is scattering on a static point monopole field remains open, but one would expect appreciable corrections at least for $k \gtrsim 10^3 M$, coming from the anomalous magnetic moment. Within the regime of applicability, the effect of the phase shift δ for an unpolarized initial beam is to rotate the transverse polarization of the outgoing electron by an angle $2S\delta$ clockwise about the outgoing velocity. Note the implications of P for the transverse polarization, that the component perpendicular to the scattering plane is the same for either sign of q, while the component in the plane reverses sign with q.

IV. ELECTRON-MONOPOLE BOUND STATES

A. The Dirac monopole case

From (3.5) and (3.11) it follows that there is a zero-energy bound state for $a = +1$ only, with

$$G = iSF = iS\sqrt{M}\, e^{-Mr}/\sqrt{4\pi}. \quad (4.1)$$

The possibility of such a state was ruled out in Ref. 2 because the conventional nonrelativistic boundary condition $1/a = 0$ was assumed. (This was understandable, as in the Coulomb problem for a Dirac particle the large components of ψ are finite and the small components vanish at the center.) The resulting Hamiltonian is indeed self-adjoint, but of course violates CP symmetry. For general positive a there is a bound state with

$$E = M(1-a^2)/(1+a^2),$$
$$G = iSaF \propto e^{-\mu r}, \quad (4.2)$$
$$\mu = 2aM/(a^2+1).$$

Only for the limiting cases $a \to 0$ or $a \to \infty$ does the bound state disappear into the positive- or negative-energy continuum, respectively.

B. Dirac particle multiplets interacting with soliton monopoles

In the theory of classical SU(2) gauge fields[16] interacting with a scalar isovector field which spontaneously breaks the SU(2) symmetry, "topologically stable" configurations with long-range magnetic monopole fields are known to exist,[6] and interactions of fermion multiplets with such solitons have been considered.[5] In Ref. 5 it was shown that zero-energy bound states may appear for these non-Abelian monopoles. Since, for small ratio of fermion mass to vector-meson mass, the bound-state wave function has negligible overlap with the interior of the monopole, one may analyze these bound states quite accurately by representing the complexities of the monopole interior with a boundary condition consistent with the requirement of self-adjointness. To do this we shall specialize to the case of an isospin-$\frac{1}{2}$ fermion multiplet interacting with a minimum-strength monopole, so that q for the two types of fermions takes on the values $\pm\frac{1}{2}$. Once again, only the lowest partial wave (which has $J=0$ here) will be sensitive to the boundary conditions. In that partial wave, we must solve the problem of self-adjoint extension for a Hamiltonian acting on a four-component wave function, with an F and G for either sign of q. Our previous discussion was restricted by implicit use of the charge superselection rule,[17] so that the boundary condition could not mix the two q sectors. However, if a soliton monopole is treated as a static object, it can mix charges. We shall see later that this charge mixing is likely to be eliminated already by classic radiative corrections, but for comparison with the earlier work[5] we may entertain temporarily the possibility of such mixing. Evidently now the adjoint of the Hamiltonian with $F(0) = G(0) = 0$ boundary conditions will have two linearly independent eigenfunctions corresponding to either eigenvalue $\pm i$, and hence the deficiency indices will be $n_+ = n_- = 2$. The self-adjoint extensions will be parameterized by the set of unitary 2×2 matrices, or 4 real parameters. Once again, further conditions are needed to constrain the parameters, and we resort to the discrete symmetries to help.

We may simplify the analysis by observing that the antiunitary operations T and PT each commute with the adjoint of H_0. Consequently they interchange eigenvectors with eigenvalues $\pm i$. The effect of the Weyl–von Neumann construction described in the previous section is to extend the domain of H_0 by including functions whose boundary values at the origin are multiples of a fixed vector, which is simply the chosen combination $v^{(+)} + v^{(-)}$ evaluated at the origin. Evidently this time the allowed boundary values for the extended domain will be linear combinations of two such fixed vectors. Since the four functions corresponding to $\pm i$ eigenvalues, when evaluated at the origin, yield four linearly independent four-component vectors, any two made equally of $+i$ and $-i$ eigenfunction contributions would be candidates to give a self-adjoint extension. Since T and PT interchange eigenvalues, any two linearly independent vectors which form the basis for a representation of the T or PT symmetry would be such candidates. Of course, the boundary condition vectors v_i must obey an additional constraint to assure Hermiticity of H,

$$v_i^\dagger \rho S v_j = 0. \quad (4.3)$$

As explained in the Appendix, if the vectors v_i also form the basis for a representation of C and P, then the resulting boundary condition obeys (4.3) and guarantees a self-adjoint extension of H_0 with all the desired symmetries.

To illustrate the simplified technique let us redo the case discussed in Sec. III. We desire a single two-component vector $(F(0), G(0))$ which is an eigenfunction of PT,

$$(F(0), G(0)) = (1, ia), \quad (4.4)$$

precisely the result (3.14) of the more cumbersome direct procedure used before. This form automatically obeys (4.3), which therefore gives no further constraint.

For the problem of interest now, we may represent TPv_j by βv_j^*, making use of arbitrariness in the choice of phases for the $S=+1$ and $S=-1$ sectors. Insisting that each vector be an eigenvector of PT, with eigenvalue +1, gives

$$v_j = (a_j, ib_j;\, c_j, id_j), \quad (4.5)$$

where the letters each stand for a real quantity.

Orthonormalization of the vectors reduces eight real parameters to five, and the further imposition of (4.3) gives the constraint

$$a_1b_2 - a_2b_1 - c_1d_2 + c_2d_1 = 0\,, \tag{4.6}$$

which brings us to four real parameters in a convenient form. See the Appendix for some amplifying remarks.

In principle there is a degree of arbitrariness in the specification of the discrete symmetries, each of which anticommutes with S. If we introduce isospin Pauli matrices τ_1, τ_2, and $\tau_3 \equiv S$, then any real linear combination of τ_1 and τ_2 might appear in each of P, T, and C. However, there are strong constraints: Given the choice already made for PT, we may insist without loss of generality that v_j be an eigenvector of CT. Since CT commutes with S, it might contain a factor $e^{i\epsilon S}$, but only $\epsilon = N\pi/2$ would allow v_j to be an eigenvector. For even N, we get

$$v_j = (a_j,\ +ia_j;\ b_j,\ \pm ib_j)\,, \tag{4.7}$$

while for odd N, the last entry is reversed in sign. We analyze the two cases separately:

(i) Even N means CT may be represented by $CTv_j = -i\beta\rho v_j$, leading to

$$\begin{aligned} v_+ &= (a,\ ia;\ b,\ ib)\,, \\ v_- &= (c,\ -ic;\ d,\ -id)\,. \end{aligned} \tag{4.8}$$

Since P, T, C each carry $S=+1$ to $S=-1$, and since each takes a pair $(1, i)$ into a pair $(1, -i)$ with some overall phase, we must choose $Pv_j = \beta\tau_1 v_j$ (the alternative $\tau_1 \to i\tau_2$ is equivalent), giving

$$\begin{aligned} v_+ &= (a,\ ia;\ b,\ ib)\,, \\ v_- &= (b,\ -ib;\ a,\ -ia)\,. \end{aligned} \tag{4.9}$$

The Dirac monopole, charge-conserving case corresponds to setting either a or b to zero. For $b=0$ there are two zero-energy bound states, one of which migrates up in energy as b departs from zero, the other down. Both states disappear into the continuum for $a=\pm b$, and there are none for $|b| > |a|$. The case $a=\pm b$ is interesting, because this boundary condition is easily seen to give a self-adjoint extension of the helicity operator $h_0 = -i\tau_3\partial_r$, as well as of the Hamiltonian. Therefore h_0 is conserved in this case. This immediately shows no bound state could be present, since a function exponentially decreasing with r could not be an eigenvector of h_0. The classical interpretation of helicity conservation is that an electron plunges into the pole and reverses charge in order to conserve the "field angular momentum" $-q\hat{r}$, while its helicity is unchanged. In the charge-conserving Dirac monopole case, helicity is flipped to balance the change in field angular momentum.

In terms of the SU(2) gauge theory with a nonsingular vector potential, helicity conservation is guaranteed by the theorem in Sec. II, so that $a=\pm b$ corresponds to the limit of infinite charged vector-meson mass, or vanishing soliton radius, for a doublet of fixed mass interacting with an SU(2) soliton monopole.

(ii) Odd N means CT may be represented by $CTv_j = -i\tau_3\beta\rho v_j$, leading to

$$\begin{aligned} v_+ &= (a,\ ia;\ b,\ -ib)\,, \\ v_- &= (c,\ -ic;\ d,\ id)\,. \end{aligned} \tag{4.10}$$

Again, P must be chosen as $\beta\tau_1$, giving

$$\begin{aligned} v_+ &= (1,\ i;\ 1,\ -i)\,, \\ v_- &= (1,\ -i;\ -1,\ -i)\,, \end{aligned} \tag{4.11}$$

where the signs in v_- are constrained by the condition $Pv_\pm = \pm v_\pm$. It is at once evident that there is one bound state of zero energy, corresponding to v_+, while no other linear combination of the v_j gives a bound state.

This is the zero-soliton-radius limit of the case considered in Ref. 5, in which the mass of the fermion doublet is generated by coupling to the scalar isovector field, so that outside the monopole the mass becomes $M\tau_3$ (which corresponds to a trivial interchange of the third and fourth entries in our vectors v_j). Inside, the covariant derivative of the scalar field is nonzero, so that the helicty operator $\vec{\sigma}\cdot(-i\nabla - e\vec{A})$ no longer commutes with H. However, the failure of commutation is proportional to M, and so helicity conservation should be valid asymptotically as k/M diverges. By some rearrangement of choice of phases, (4.11) may be replaced by

$$\begin{aligned} v_1 &= (1,\ 1;\ 1,\ 1)\,, \\ v_2 &= (1,\ -1{:}\ -1,\ 1)\,. \end{aligned} \tag{4.12}$$

These indeed give eigenstates of $h_0 = \pm k$ except for corrections of order M/k.

In the charge-conserving limit there is an exact U(1) symmetry, so that the phase $\epsilon\tau_3$ is arbitrary. In particular, the even N solutions with $a=0$ or $b=0$ also form a basis for representation of the discrete symmetries with odd N. However, there is no continuous family of such representations. The form (4.11) is an isolated solution.

To summarize, allowing charge-mixing while still imposing the discrete symmetries gives a continuous family of solutions, connecting the case of a Dirac monopole interacting with two kinds of fermion with opposite charge, each with a zero-energy bound state, to the case of a Dirac pole interacting with two kinds of fermion, each with no bound state. Just at the point where the bound states disappear, helicity conservation is exact, at the cost of maximal violation of charge conservation. A second isolated boundary condition possibility gives one zero-energy bound state,[5] charge conservation up to $(k/M)^2$ corrections (which become large for relativistic incident particles) and helicity conservation up to $(M/k)^2$ corrections (which become small in the relativistic limit).

C. Inhibition of charge-exchange processes by classical radiative effects

In the fully quantized theory, or even in the classical theory with the gauge fields free to respond to disturbances, electric charge must be conserved, as discussed already in Ref. 16. For the following, let us assume that the monopole possesses a definite, quantized electric charge. Further, let us take the radius of the pole as $R \approx 1/M_V$, where the vector-meson mass M_V is large compared to M/e^2, with e the electron charge. If M and M_V were both generated by spontaneous symmetry breaking, they should be of the same order of magnitude, but our whole discussion is only relevant for the case $M \ll M_V$. Since low-mass fermions but not low-mass charged vector mesons are known, this restriction may have some interest.

Consider now the collision of a relativistic electron with a monopole supposed capable of changing charge. Since the portion of the electron's field which does not touch the pole must travel freely forward, giving bremsstrahlung of expected energy $\approx \gamma e^2/R \approx \gamma e^2 M_V$, the probability of charge exchange must be exponentially damped for $M \ll e^2 M_V$. At the other extreme, if the electron were quite nonrelativistic it would face a potential barrier against charge exchange of height $e^2 M_V$, resulting in exponential damping by a barrier penetration factor $\approx Me^2/k$. Thus, from the static to the relativistic regime, penetration of the pole by the electron with consequent charge exchange is overwhelmingly suppressed. This has been deduced without any assumptions about the energy required for the charge lost by the fermion to be taken up in other kinds of charge-carrying excitation.

The implication is that a zero-energy, mixed-charge bound state might still be present, but the size of its coupling to the exterior, i.e., the coefficient of its e^{-Mr} "tail," must be exponentially small. Under the assumptions on M/M_V made here, only the Dirac monopole solutions of Secs. III and IV A could be relevant to experimental observations.

The phenomenon of decoupled zero-energy bound states is present for a Dirac particle with a small anomalous magnetic moment of either sign, interacting with a monopole.[18] A bound state is found for each partial wave, but all except the one discussed here in part A are decoupled by the centrifugal barrier.

One may ask whether classical radiative corrections might decouple zero Euclidean-action fermion trajectories in the presence of pseudoparticles. Such trajectories have been discussed recently,[19] but the question raised here seems open.

V. SECOND QUANTIZATION FOR FERMIONS IN THE PRESENCE OF MONOPOLES: ARGUMENT FOR INTEGRAL FERMION NUMBER

The discoverers of zero-energy fermion-monopole bound states observed an interesting consequence for the action of the fermion field if the standard anticommutation relation

$$\delta(x_0 - y_0)\{\psi^\dagger_\alpha(x), \psi_\beta(y)\} = \delta_{\alpha\beta}\delta^{(4)}(x-y) \qquad (5.1)$$

were to hold[5]: The field ψ must connect two degenerate spinless states of the pole, which therefore differ in fermion number n by one unit. In other words, if the bound state were unoccupied, an electron could be "dropped" into it, radiating at least one electron mass in photon energy, while if the state were occupied, a positron could be dropped into it. Avoiding the asymmetry suggested by use of the term "occupied," the authors of Ref. 5 proposed to assign to the two condition values $n = \pm\frac{1}{2}$. Since monopole strength g is absolutely conserved in such theories, any n may be replaced by

$$n' = n + xg \qquad (5.2)$$

(where x is an aribitrary real number) with no observable consequences resulting from the redefinition. Both n and n' are odd under charge conjugation. By the same token, if baryon and lepton number were exactly conserved, each baryon could be said to carry lepton number π, without possibility of contradiction. Thus the intriguing notion of "half-fermions" suggested by $n = \pm\frac{1}{2}$ has no meaning for a single isolated pole. However, it is a convenient choice because it makes n manifestly even under P and T. It is the automatic result of defining the fermion density as

$$\rho(x) = \tfrac{1}{2}[\psi^\dagger(x), \psi(x)]. \qquad (5.3)$$

To gain insight into these matters, let us consider some observable properties and reactions involving poles and electrons. It was shown recently that if particles bearing electric and/or magnetic charges obey the usual connection between spin and statistics, then composites formed from these particles will also, even though the spin of such a composite may differ by a half integer from the sum of the spins of its components.[20] This means that if the $n=-\frac{1}{2}$ pole is a boson, so also is the $n=+\frac{1}{2}$ pole, so that fermion number here has nothing to do with statistics. Furthermore, two such monopoles of opposite magnetic charge, and possessing a total electric charge and n both of one unit, in combination would form a spin-$\frac{1}{2}$ fermion which could turn into an electron plus photons.

While the n values have no meaning by themselves, the electric charge and charge density of a monopole are observable. The Dirac quantization condition[21] could be obeyed if all monopoles had integer multiples of an electron charge, or if they all had half-integer electric charges. For a theory with only point monopoles and electrons, any given pole could be in a state with either of two charges (depending on $n=\pm\frac{1}{2}$) differing by e. However, if the charges were 0 and $-e$, it would make more sense to consider the first as unoccupied, with $n=0$, and the second (perhaps different in mass) as occupied, with $n=1$. In that case there is a distinct possibility with observable consequences, which bears a relation to the case discussed so far analogous to that of integrally charged triplets to quarks. Monopoles could be labeled by a new dichotomic variable, so that a pole of given strength g and type (+) could have electron number $n=0$ or $n=1$, while a pole of type (−) could have electron number $n=0$ or $n=-1$. This two-type classification would arise naturally if the fermion density were defined by normal ordering

$$\rho(x)=N[\psi(x)\psi^\dagger(x)]\equiv\psi^\dagger\psi-\langle 0|\psi^\dagger\psi|0\rangle \quad . \tag{5.4}$$

where $|0\rangle$ is the fundamental or vacuum state of the theory. The normal-order definition, as the name suggests, is the normal starting point for perturbative quantum electrodynamics. For free electrons, it is equivalent to the commutator definition, but in the presence of an external potential, especially one which produces an electron bound state, normal ordering is used, not the commutator definition.[22] Clearly, to use this definition here one must identify one of the monopole states as the vacuum and the other as an electron state of charge $-e$ if it is higher in n, or as a positron state of charge $+e$ if it is lower in n. This choice involving two types of poles seems unavoidable if a consistent QED is to be built in the presence of massive monopoles. Two poles of opposite strength but same type (+) could annihilate to form 0, 1, or 2 electrons, while two of opposite type could form −1, 0, or 1 electron (where "−1" means a positron). Also a type (+) and a type (−) pole of the same g would be distinguishable and suffer no symmetry constraint on their wave function.

It is worthwhile to examine in detail the most concrete argument for half-integer fermion number, which depends on a subtle thought experiment.[5] Consider the Dirac equation for an electron in the presence of two poles of equal and opposite strength $g_1=g$ and $g_2=-g$. When the poles are widely separated, there will be two zero-energy bound states, one localized around g_1, the other around g_2. If the poles approach, the two states will mix with each other, resulting in a splitting of the energy levels. Because the Hamiltonian anticommutes with CT, the levels will always have equal and opposite energies. For a separation of order $1/M$, the bound states will disappear, one into the positive-energy continuum, the other into the negative-energy continuum. Now suppose the monopoles are adiabatically separated. When they are again far apart, the fermion field might be expected to take the form

$$\begin{aligned}\psi=&\sum_{E>M} a(E,\xi)\phi(E,\xi,\vec{r})e^{-iEt}\\ &+\sum_{E<M} b^\dagger(E,\xi)\phi(-E,\xi,\vec{r})e^{iEt}\\ &+\frac{1}{\sqrt{2}}[\phi_1(\vec{r})+\phi_2(\vec{r})]a\\ &+\frac{1}{\sqrt{2}}[\phi_1(\vec{r})-\phi_2(\vec{r})]b^\dagger\,,\end{aligned}\tag{5.5}$$

where ϕ_1, ϕ_2 are normalized zero-energy bound-state wave functions localized around g_1, g_2, respectively. At first sight, the above expression seems to lead to violations of the cluster decomposition property one expects for a quantum field. The problem is that a and b are both associated with wave functions localized at two widely separated points. However, observe that the vacuum is a coherent superposition

$$\begin{aligned}|0\rangle &=(|0_+\rangle+|0_-\rangle)/\sqrt{2}\,,\\ |0_\pm\rangle &=(1\mp a^\dagger b^\dagger)|0\rangle/\sqrt{2}\,.\end{aligned}\tag{5.6}$$

The operators $c_\pm=(a\pm b^\dagger)/\sqrt{2}$ and $d_\pm=(a^\dagger\mp b)/\sqrt{2}$ annihilate $|0_\pm\rangle$. Thus, for the vacuum $|0_+\rangle$, we write the zero-energy state contribution to ψ as

$$c_+\phi_1(\vec{r})+d_+^\dagger\phi_2(\vec{r})\,,\tag{5.7a}$$

while for $|0_-\rangle$ we write

$$d_-^\dagger\phi_1(\vec{r})+c_-\phi_2(\vec{r})\,.\tag{5.7b}$$

In either case, the number operator receives a contribution

$$n=c^\dagger c-d^\dagger d\tag{5.8}$$

while the number density operator (using either the commutator or normal-ordering definition for ρ in terms of a and b, giving the commutator definition *only* in terms of c and d) becomes, for

$|0_\pm\rangle$,

$$\rho(\vec{r}) = (c^\dagger c - \tfrac{1}{2})\phi^\dagger_{1,2}\phi_{1,2} - (d^\dagger d - \tfrac{1}{2})\phi^\dagger_{2,1}\phi_{2,1}. \tag{5.9}$$

For the state $|0_+\rangle$ there is a fermion number $-\frac{1}{2}$ and charge $+e/2$ on g_1 and fermion number $+\frac{1}{2}$ and charge $-e/2$ on g_2; for $|0_-\rangle$ the fermion numbers and charges are exchanged.

So far the analysis seems to confirm the half-integer quantum numbers, but in fact higher-order electromagnetic effects cannot be accommodated consistently. In the state $a^\dagger b^\dagger|0\rangle$ there is a Coulomb interaction energy

$$\begin{aligned} V_{\text{Coulomb}} &= -e^2 \int d^3r\, d^3r' \phi^\dagger_+(\vec{r})\phi_+(\vec{r})\phi^\dagger_-(\vec{r}\,')\phi_-(\vec{r})/|\vec{r}-\vec{r}\,'| \\ &\approx -\frac{e^2}{2}\int d^3r\, d^3r' \phi^\dagger_1(\vec{r})\phi_1(\vec{r})\phi^\dagger_1(\vec{r}\,')\phi_1(\vec{r}\,')/|\vec{r}-\vec{r}\,'| \\ &= -2M\alpha \ln 2\,. \end{aligned} \tag{5.10}$$

Consequently there is an oscillation at this frequency between the states $|0_+\rangle$ and $|0_-\rangle$, implying exchange of charge and fermion number between the two poles, no matter how great their separation. Of course this would violate current conservation, and so should only be taken as a sign of inconsistency of the assumptions.

A little thought shows that the inconsistency will arise unless the field in the limit of large pole separation contains either the combination

$$a\phi_1 + b^\dagger\phi_2$$

or

$$a\phi_2 + b^\dagger\phi_1\,.$$

In other words the only consistent hypothesis is that the monopoles are of opposite types. This means that the Hilbert space on which the electron field acts is a superposition of two different Fock spaces. Since the field does not mix the two spaces, it is immaterial whether the superposition is coherent, or incoherent (which would mean that different photon radiation patterns were associated with the different Fock spaces).

It was argued[5] that, since there are four distinct configurations for the two poles close together ($0, e^-, e^+, e^+e^-$), there should be only four configurations when they were separated ($\pm e/2$ on g_1, $\pm e/2$ on g_2). However, this counting is ambiguous, since it neglects the radiative coupling which makes (e^+e^-) unstable against decay to photons. The result of the present analysis is that for large separation there are two possible sets of four configurations, corresponding to the two different Fock spaces. The CT symmetry assures that the two sets will occur with equal frequency, but does not require half-integer fermion numbers.

An amusing consequence of the existence of two types of pole is the possibility that an electron-containing pole could pass close to an electron-containing antipole. The collision would be like that of two uranium nuclei, in the sense that a single-particle electron state would be driven into the negative-energy continuum. Presumably the consequences should be similar also.[23]

Addendum

Questions raised by R. Jackiw have led me to clarify and extend the analysis of this section, strengthening the main conclusion.

(i) It has been assumed here that the fermion and electromagnetic currents may be taken as proportional. This is justified by the arguments in Sec. IV C which show that the only realistic case in which single-particle wave functions would lead to a reasonable approximation to an exact theory would be an electrically charged fermion interacting with a Dirac monopole. Furthermore, even in the unrealistic charge-mixing approximation of a soliton monopole arising in a gauge theory, only the Fermi field carries the minimum unit of charge; the gauge and scalar fields carry twice that charge. Therefore, it is hard to imagine how these other fields could cancel a half-integral charge associated with half-integral fermion number.

(ii) The one-dimensional problem of a fermion bound to a kink configuration of a scalar field, reviewed in Ref. 5, needs to be discussed further. To obtain analytic results let us make a point approximation here also, so that the problem is that of a fermion of mass M for $x > x_1$, $-M$ for $x < x_1$. It is easily seen that there is one bound state, and it has $E=0$. If an antikink were located at $x_2 < x_1$, two bound states would be found. As the kinks approached, the bound states would split, one going up in energy, the other down, but they would only merge into the positive- and negative-energy continua for exactly zero separation of the kinks. Since the adiabatic approximation must break down for a finite separation determined by the dimensions of the kink, it follows that adiabatic motion does not lead to a unique identification of the high-energy state with a fermion and the low-energy state with an antifermion. Consequently, already for the one-dimensional case one is entitled with no fear of inconsistency to postulate two types of kinks, fermion acceptors and antifermion acceptors.

(iii) The results for one dimension are suggestive, though not conclusive, for the three-dimensional monopole case. They suggest that the assumption in Sec. V, that the electron bound states for a pole-antipole pair would merge into the continua at finite pole separation, is probably wrong: It would only happen for zero separation. Even if this were not the case, the merging could not be an adiabatic process occurring in finite time, since the bound-state wave functions must become infinitely spread out as they approach the continua. In the analog problem of U-U collisions, the negative-energy electron wave function remains confined by the Coulomb barrier even as it enters the negative-energy sea,[23] but monopoles provide no such barrier. Therefore, once again it is impossible to connect the pole-antipole vacuum adiabatically with the ordinary vacuum. This means the two-type classification is again permissible and consistent.

(iv) In the main part of Sec. V it was found that it is very hard to decide the question of half-integral fermion number from general and abstract considerations, but that a straightforward effort to implement $n=\pm\frac{1}{2}$ led to inconsistencies. Conversely, in this addendum it has been found that no inconsistencies arise from the two-type hypothesis with integral fermion number. Taken together, these arguments show that introduction of the radical and mysterious concept of solitons with fermion number $\pm\frac{1}{2}$ would be premature at best.

VI. SUMMARY

The discrete and the dynamical symmetries of motion for a Dirac electron in a magnetic field impose powerful constraints, including a helicity-conservation law which is not widely known despite nearly fifty years of research on the Dirac equation.

The field singularity at the location of a magnetic monopole leads to violation of all the dynamical symmetries, but only for the lowest partial wave. In nonrelativistic terms, for this partial wave a quasicentrifugal potential, strong enough to shield the heart of the monopole from a spinless particle, is exactly canceled by the attractive magnetic dipole interaction of the electron. Consequently, to define a self-adjoint Hamiltonian it is necessary to supplement the Dirac differential operator with a boundary condition at the pole, different from the usual requirement that the wave function be everywhere finite. Instead, there is a $1/r$ divergence at the pole, proportional to a particular spinor. For a Dirac monopole, the boundary-condition spinor is determined by a single real parameter, which therefore labels the different self-adjoint extensions of the Hamiltonian. However, only for two values of that parameter may the discrete CP symmetry be realized by the extended H.

Thus, the Weyl–von Neumann theory of self-adjoint extension, combined with the discrete symmetries, leads to an almost unique (actually bivalued) choice of H. The two possibilities are precisely those resulting when H is taken as a limiting form of the well-defined Hamiltonian for an electron endowed with a small anomalous magnetic moment of either sign.[3] If the anomalous moment enhances the Dirac moment (as for a real electron) there is a zero-energy bound state for the limiting H.[18]

The same techniques used for the Dirac monopole may be applied to the unphysical case in which the pole is a point limit of a classical finite-energy field configuration arising in a non-Abelian gauge theory. This is unphysical because such a monopole violates the charge superselection rule by mixing different fermion charge states for the lowest partial wave. While such classical field configurations might well be related to monopoles, one would expect classical and quantum radiative corrections to decouple different charge states for fermions with energy much less than the mass of a charged vector meson, as well as to ensure that charge lost by a fermion is absorbed by the gauge field.[24]

Unphysical or not, at the level of single-particle wave functions the problem of a fermion multiplet interacting with a charge mixing pole is well defined. The example of a fermion doublet has been analyzed in detail, reproducing a result[5] which yielded a single, charge-mixing, zero-energy bound state. In addition, a different realization of the discrete symmetries gives a continuous one-parameter family of boundary conditions, with limiting extremes of a Dirac monopole with either two or no bound states. In the middle of the range is the amusing case of exact helicity conservation, entailing maximal-charge flip and of course no bound states.[25]

Perhaps the most subtle issue yet investigated is the physical significance of a zero-energy bound state for second quantization of the fermions. A previous argument[5] that the result would be two equivalent vacuum states of the monopole, with fermion number $\pm\frac{1}{2}$, has been rebutted. The general counterargument is that a perturbative expansion like quantum electrodynamics depends on the existence of a unique vacuum state. Two different vacua connected by the Fermi field would require a wholly new approach, if it could be done consistently at all. Detailed consideration of consistency requirements on the Fermi field in the presence of a pole-antipole system supports the view that fermion numbers (0, 1) or (0, −1) are the only allowed choices. The resulting picture, involving two types of monopoles, realizes in second quantization the CT invariance found in the single-particle, first-quantized theory.

The mathematical beauty and richness of monopole physics continues to unfold. From the theoretical point of view, the next step may be the development of a generalized QED of charged fermions in the field of a point monopole. If that can be achieved, it will leave still open the challenge of second quantization for the degrees of freedom associated with a monopole, perhaps treated as a nonperturbative excitation of meson fields. Of course, the biggest challenge is still to find a monopole in nature.

ACKNOWLEDGMENTS

I thank Yoichi Kazama and Chen Ning Yang for introducing me to the issues which led to this paper. Roman Jackiw and Claudio Rebbi sent numerous papers and made useful private remarks. Ronald Douglas, Joel Pincus, Barry Simon, and Craig Tracy provided invaluable training in the Weyl–von Neumann theory of self-adjoint extensions of Hermitian operators. William Weisberger gave much time to careful and critical discussions essential to understanding the physics of zero-energy bound states in second quantization.

APPENDIX A: DISCRETE SYMMETRIES AND SELF-ADJOINT EXTENSIONS OF THE HAMILTONIAN

It is suggested in Sec. IV B that imposing discrete-symmetry requirements on a pair of vectors giving the possible boundary values of the fermion isodoublet wave function at the origin would be an efficient way to obtain self-adjoint extensions of H obeying these symmetries. Clearly all such vector pairs will give the symmetries, but will they give self adjoint extensions? To answer this, let us establish a standard basis

$$v_1^{(\pm)} = (1, \pm e^{\pm i\alpha}; 0, 0)\,, \qquad v_2^{(\pm)} = (0, 0; 1, \mp e^{\pm i\alpha})\,. \tag{A1}$$

The Weyl–von Neumann procedure corresponds to choosing vector pairs

$$v_j = v_j^{(+)} + U_{jk} v_k^{(-)}\,. \tag{A2}$$

Since $PT = \beta K$ (where K stands for complex conjugation) exchanges $v_j^{(+)}$ with $v_j^{(-)}$, PT will be represented on the pair v_j only if the unitary matrix U is symmetric. For a general U there will be representation of $PT(\theta) \equiv \beta K e^{i\theta\tau_3}$ with some choice of θ. The phase θ clearly will not influence the eigenvalue spectrum of the extended Hamiltonian, and so we may set it to zero, restricting attention to a three-parameter set of U matrices. The parity operation $P = \beta\tau_1$ gives $v_2^{(\pm)} \leftrightarrow v_1^{(\pm)}$. Therefore, if P is also to be represented on v_j, the matrix U must be symmetric with respect to both its diagonals, and so has only two adjustable parameters. Conversely, two eigenvectors obeying $PTv_j = v_j$ and $Pv_j = (-1)^{j+1} v_j$ must correspond to a possible self-adjoint extension of H, since these conditions imply

$$v_j = x_j v_1^{(+)} + x_j^* v_1^{(-)} + y_j v_2^{(+)} + y_j^* v_2^{(-)} \tag{A3PT}$$

and

$$x_j = (-1)^{j+1} y_j\,, \tag{A3P}$$

which is equivalent to specifying a two-parameter U in (A2). Therefore, any pair of vectors giving a nondegenerate representation of P, T, and C will give boundary conditions which guarantee that the extended H is self-adjoint. The proviso "nondegenerate" refers to the need for two different eigenvalues of P, which is satisfied by the solutions in Sec. IV B.

---

*Research partially supported by the National Science Foundation under Grant No. PHY 76-15328.

[1]P. P. Banderet, Helv. Phys. Acta 19, 503 (1946). This paper is the earliest I know to give the invariance relation (2.7).

[2]Harish-Chandra, Phys. Rev. 74, 883 (1948).

[3]Y. Kazama, C. N. Yang, and A. S. Goldhaber, Phys. Rev. D 15, 2287 (1977).

[4]J. Schwinger, K. A. Milton, W.-Y. Tsai, L. L. DeRaad, Jr., and D. C. Clark, Ann. Phys. (N.Y.) 101, 451 (1976).

[5]R. Jackiw and C. Rebbi, Phys. Rev. D 13, 3398 (1976).

[6]G. 't Hooft, Nucl. Phys. B79, 276 (1974); A. M. Polyakov, Zh. Eksp. Teor. Fiz. Pis'ma, Red. 20, 430 (1974) [JETP Lett. 20, 194 (1974)].

[7]N. F. Ramsey, Phys. Rev. 109, 225 (1958).

[8]P. A. M. Dirac, Proc. R. Soc. London A117, 610 (1928).

[9]See T. Kato, *Perturbation Theory of Linear Operators*, 2nd edition (Springer, New York, 1976), p. 545 ff. and M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, to be published), Vol. III.

[10]M. S. Plesset, Phys. Rev. 36, 1728 (1930).

[11]I. I. Rabi, Z. Phys. 49, 507 (1928).

[12]M. H. Johnson and B. A. Lippmann, Phys. Rev. 76, 828 (1949).
[13]V. Bargmann, L. Michel, and V. L. Telegdi, Phys. Rev. Lett. 2, 435 (1959).
[14]H. J. Lipkin, W. I. Weisberger, and M. Peshkin, Ann. Phys. (N.Y.) 53, 203 (1969).
[15]H. Weyl, Math. Ann. 68, 220 (1910); J. von Neumann, Math. Ann. 102, 49 (1930). F. Reisz and B. Sz-Nagy, *Functional Analysis* (Ungar, New York, 1955). K. M. Case [Phys. Rev. 80, 797 (1950)] applied the technique of self-adjoint extension to the problem of an electron in a strong Coulomb field. As in the discussion here, there is a continuous one-parameter family of self-adjoint extensions, but there is no symmetry (such as CP here) to restrict the choice of parameter to discrete values. Therefore it becomes necessary to carry out calculations for realistic charge distributions to deduce the actual effects of large nuclear charge. See Ref. 23. Self-adjoint Dirac Hamiltonians are also discussed in B. W. Roos and W. C. Sangren, J. Math. Phys. 3, 882 (1962); K. Jörgens, Lecture Notes in Math. 280, 87 (1972); J. Weidmann, *ibid.* 448, 71 (1975).
[16]C. N. Yang and R. L. Mills, Phys. Rev. 96, 191 (1954).
[17]G. C. Wick, A. S. Wightman, and E. P. Wigner, Phys. Rev. 88, 101 (1952).
[18]Y. Kazama and C. N. Yang, Phys. Rev. D 15, 2300 (1977).
[19]R. Jackiw and C. Rebbi, Phys. Rev. D 16, 1052 (1977).
[20]A. S. Goldhaber, Phys. Rev. Lett. 36, 1122 (1976).
[21]P. A. M. Dirac, Proc. R. Soc. London A133, 60 (1931); Phys. Rev. 74, 817 (1948).
[22]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).
[23]Ya. B. Zel'dovich and V. S. Popov, Usp. Fiz. Nauk 105, 403 (1971) [Sov. Phys.—Usp. 14, 673 (1972)]; J. Rafelski, B. Müller, and W. Greiner, Nucl. Phys. B68, 585 (1974); L. P. Fulcher and A. Klein, Phys. Rev. D 12, 1194 (1975).
[24]The discussion in Sec. IV on radiative decoupling of charge-mixing configurations was quite conservative, since one would expect a change of 2 units of electron charge to require an energy of order M_V. This would be true even if the charge were deposited on the soliton monopole, making it into a dyon. B. Julia and A. Zee, Phys. Rev. D 11, 2227 (1975).
[25]C. J. Callias [Phys. Rev. D (to be published)] considers the interaction of a fermion doublet with a point monopole. He imposes the superselection rule for electric charge, but does not exploit systematically the discrete symmetries. As a result, he is able to find the limit corresponding to a Dirac monopole with 2 zero-energy bound states, but not any of the other cases analyzed here, including the point-monopole limit of the fermion-soliton bound state in Ref. 5. I thank the author for sending me a copy of his paper.